UNIVERSITAT DE BARCELONA

Computer-aided rational enzyme design for industrial and technological applications

Gerard Santiago Morcillo

Departament de Química Física
Facultat de Química
Universitat de Barcelona
Programa de doctorat en Química Teòrica i Modelització Computacional

Computer-aided rational enzyme design for industrial and technological applications

Memòria presentada per:

Gerard Santiago Morcillo

Per optar al grau de Doctor per l'Universitat de Barcelona

Dirigida per:

Prof. Victor Guallar Tasies

Dra. Maria Fàtima Lucas

Tutor:

Prof. Jaime Rubio Martínez

UNIVERSITAT DE BARCELONA

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Barcelona, 2019

Sic parvis magna

Sir Francis Drake (1540-1596)

List of Contents

Agraïments

Abans de començar els capítols de la tesis m'agradaria dedicar unes línies a totes les persones que han recorregut aquest camí amb mi.

En primer lloc donar les gràcies al Prof. Victor Guallar, qui em va donar l'oportunitat d'unir-me al seu equip com a estudiant de pràctiques ara fa cinc anys. A la Dra. Fátima Lucas que em va guiar a l'inici d'aquesta etapa. A tots el meus companys del grup Electronic and Atomic Protein Modelling, tant els que ja han deixat el grup, Emanuele, Marina, Israel, Sandra, Pedro, Ignacio, Jorge, Dani, Isaac, Oliver a tots els que encara hi son, Ferran, Pep, Joan, Ruben, Carles, Martí, Oriol, Marc, Sergi i Jelisa, i als companys de Nostrum Biodiscovery. I també a tots el companys del departament de Life Science i al BSC en general.

Un agraïment a la xarxa de col·laboradors amb els que hem treballat durant aquests anys, al Prof. Ángel Martínez i a la Dra. Susana Camarero del CIB-CSIC, al Dr. Manuel Ferrer del ICP-CSIC i al Prof. Roland Ludwig de la BOKU. I a tots els companys fets gràcies aquests projectes.

Finalment, unes línies a la meva família i amics que han aguantat les meves "batalletes" de cièntific els últims anys. I com no, i per acabar, a la Núria, t'estimo.

List of Abbreviations

ABTS	2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)
ANL	aniline
DM	double mutant
DMPD	n,n-dimethyl-p-phenylenediamine
EH	ester hydrolase
EHO	ethyl hexanoate
HAD	haloacid dehalogenase
kcat	catalytic constant
Km	michaelis constant
kcat/Km	catalytic efficiency
MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
P450	cytocrome P450
PANI	polyaniline
PELE	protein energy landscape exploration
PDB	protein data bank
pNPP	p-nitrophenyl propionate
PPD	p-phenylenediamine
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
SASA	solvent accessible surface area
SP	spin density

Chapter 1: Introduction

1. Enzymes

In a general point of view, enzymes are biological macromolecules capable to perform catalytic reactions and return to the original state once the products are released. Almost all metabolic processes in cells require a biological catalyst and they play a crucial role in the life cycle.

Regarding the composition, enzymes are chains of amino acids, which are organic compounds containing amine and carboxyl functional groups, alongside a ramification chain specific for each amino acid. From around 500 naturally occurring amino acids currently known, only 20 are present in the genetic code. There are different classification levels such as core structural-functional groups, polarity or pH, but classification by side chain group type is the most common (**Fig. 1.1**).

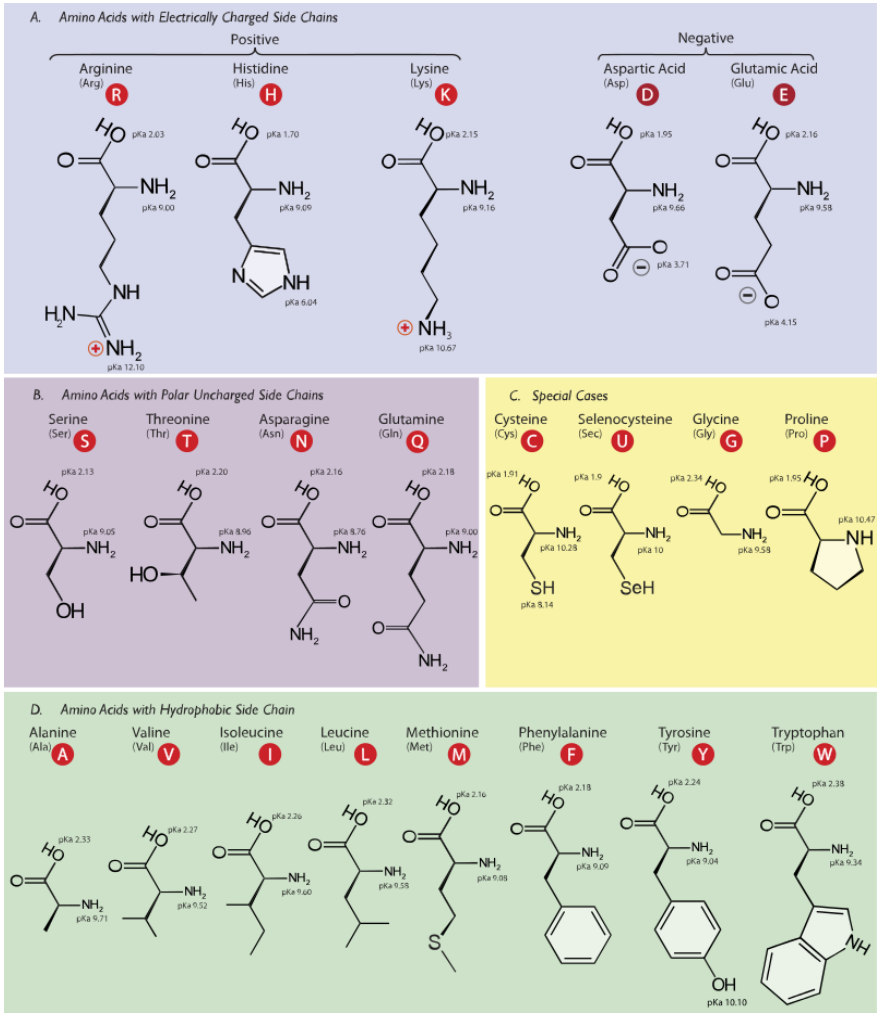


Figure 1.1: The 21 proteinogenic α-amino acids found in eukaryotes (including the selenocysteine), grouped according to their side chains' pKa values and charges carried at physiological pH (7.4).

From a structural point of view, most enzymes are globular proteins. It is worthy to notice the ribozyme exception, RNA molecules that are capable of catalyzing specific biochemical reactions, similar to the action of protein enzymes. Previous 21 natural amino acids (and the unnatural ones as well) can fold together achieving four different complexity levels. Amino acids can form

“polymers” through the peptide bond, this level is known as protein primary structure. Protein “polymers” can adopt higher organizations levels. This folding level is the protein secondary structure, regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the alpha-helix, beta -sheet and turns. Because secondary structures are local, many regions of different secondary structure can be present in the same protein molecule. When different secondary structures from the same amino acids “polymer” pack together, the protein is assuming a new organization level, the tertiary structure. Tertiary structure is generally stabilized by nonlocal interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen bonds, disulfide bonds, and even posttranslational modifications. The term "tertiary structure" is often used as synonymous with the term fold. The tertiary structure is what controls the basic function of a protein. When several protein units pack together forming bigger complexes, this level is called quaternary structure (**Fig. 1.2**).

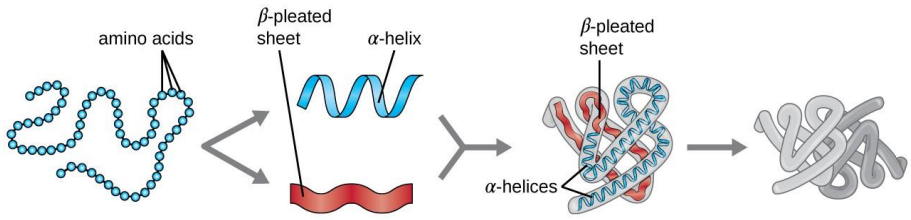


Figure 1.2: Protein different levels of structural organization (modification of work by the National Human Genome Research Institute).

Many enzymatic families have ions, like copper, zinc, magnesium or iron, and other cofactors, such as heme groups, flavins or nicotinamides, in their structure. These elements are usually playing a key role in the catalysis and sometimes in the enzyme folding too.

Most biochemical reactions take place inside a specific place called binding or active site (Singh, Kumar, Mittal, & Mehta, 2016), usually shared through enzymatic families and with common elements and geometrical disposition. From an energetic point of view, enzymes reduce activation energies, speeding up the reaction to happen (**Fig. 1.3**).

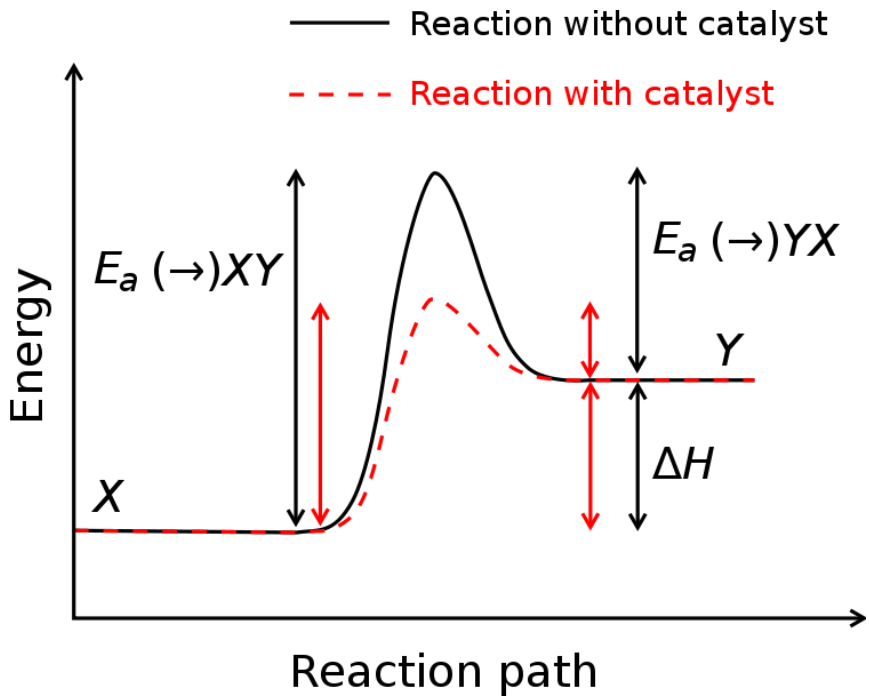


Figure 1.3: Reaction activation energy differences with and without a catalyst.

Enzymatic presence in nature is really large, it is possible to find them everywhere, from the human body to submarine volcanoes. The extensive distribution range has favored enzymatic evolution, allowing the appearance of many functions for enzymes.

1.1. Enzymes for industrial and technological purposes: applications and properties

Enzymes have been used by humans since the beginning of civilization, three thousand years ago. It is reported beer, bread, and winemaking by ancient Sumerians and Egyptians (Sewell, 2014), all these products required of yeast fermentation to be done from cereals or fruits. These firsts (and rudimentary) fermentation attempts were, through the centuries, slowly improved and standardized, starting the original "biotechnological companies". Despite the fact that they didn't know it, they were enzymology pioneers and put the seeds for future investigations and research areas.

But (and moving extremely fast through centuries of human history) (**Fig. 1.4**), molecular biology's last decade's expansion pushed forward the use of enzymes for several and more complex industrial and technological purposes (Singh et al., 2016). Currently, industrial enzymes are a growing market of more than 5 billion dollars (Chapman, Ismail, & Dinu, 2018), and it keeps growing every year.

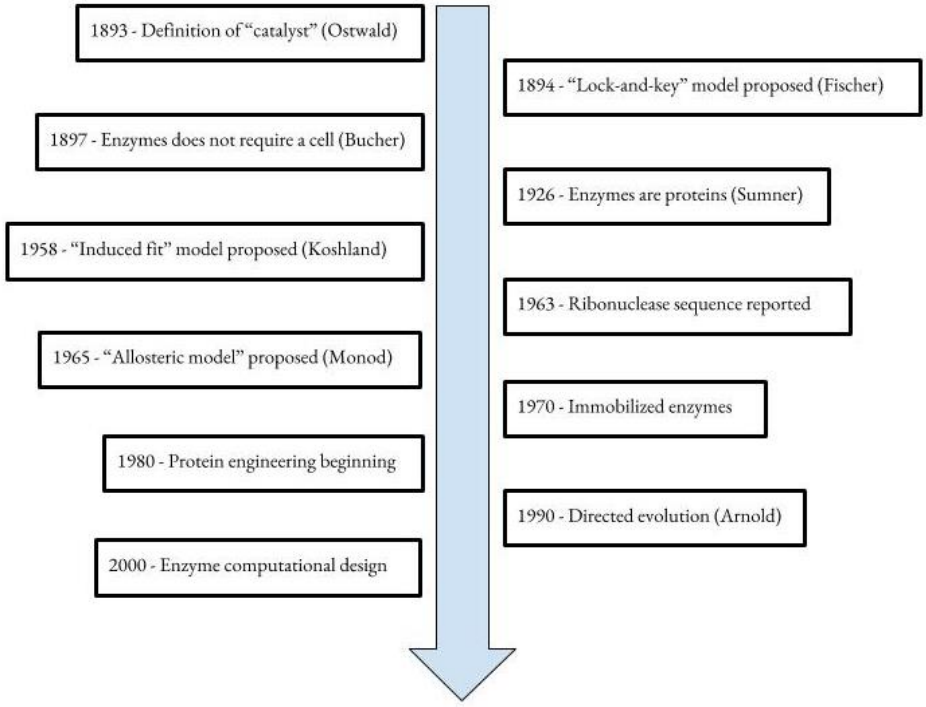


Figure 1.4: Enzymology last two centuries milestones (Robinson, 2015).

During the 20th century, we have learned that enzymes can be cultured, modified and produced largely by gene manipulation and these allowing the extensive use in an enormous variety of applications, from waste treatment to pharmaceutical applications, enzymes are present in almost all industries (**Table 1.1**).

Table 1.1: Industrial applications of enzymes (Choi, Han, & Kim, 2015).

Application field	Enzyme examples	Properties
Fine and bulk chemistry industry	Dehydratases	Remove water molecules
	Nitrilases	Remove nitrile groups
	Transferases	Transfer molecules
	Isomerases	Convert one isomer to another
Pharmaceutical industry	Dehalogenases	Catalyzes the removal of a halogen atom
	Hydrolases	Use water to break a chemical bond
	Keto-reductases	Transfer of electrons from one molecule, the reductant to another, the oxidant
Food industry and Cosmetic industry	Glycoside hydrolase	Catalyze the hydrolysis of glycosidic bonds in complex sugars
	Glycosyltransferases	Establish natural glycosidic linkages
	Hydrolases	Use water to break a chemical bond
Textile industry and Pulp and paper industry	Cellulases	Decomposition of cellulose and of some related polysaccharides
	Proteases	Helps proteolysis by hydrolysis of peptide bonds
	Oxidoreductases	Transfer of electrons from one molecule, the reductant to another, the oxidant

Enzymatic presence in different industrial areas is really relevant, reducing waste generation and creating new possible production paths, being a key player in the biotechnology world.

Of many enzymatic properties, one of the most interesting for industrial purposes is substrate promiscuity or the capability to react with several ligands. Substrate promiscuity implies that enzymes are able to catalyze the same reaction over different substrates from natural activity performance. This aspect is remarkably important in hydrolases for which many practical promiscuous applications have been studied (Beloqui, de María, Golyshin, & Ferrer, 2008; Carboni-Oerlemans et al., 2006; Hult & Berglund, 2007).

Serine-hydrolases, one of the main focus of this thesis, accept a wide range of esters and are extensively used in the resolution of racemic acids and alcohols; they are also active towards fatty acids (of different lengths and types). Lipases, for example, can accommodate a large variety of substrates, other than just fatty acids or esters, like alicyclic, bicyclic, and aromatic esters and even esters based on organometallic sandwich compounds (C.-S. Chen & Sih, 1989; Ghanem, 2007; Kapoor & Gupta, 2012; Schmid & Verger, 1998). Laccases can also accepted a wide range of substrates, such as phenols, polyphenols or aromatic amines (Alcalde, 2015).

Nevertheless, the extensive industrial use of enzymes is limited. Classical chemistry protocols are in many aspects the selected choice, due to the associated implementation expenses and the catalytic versatility of the heterogeneous reactivity (Tanimu, Jaenicke, & Alhooshani, 2017; Wang et al., 2017). Alongside, one of the biggest biocatalysis challenges is the competitiveness against functionalized surfaces to perform heterogeneous catalysis (Wittstock, Biener, & Bäumer, 2010), where small microchips (or reactors) can host thousands of immobilized reactive sites (**Fig. 1.5**). The

overcoming of these limitations, as we will discuss in the following chapters, is the enzymatic improvement and development.

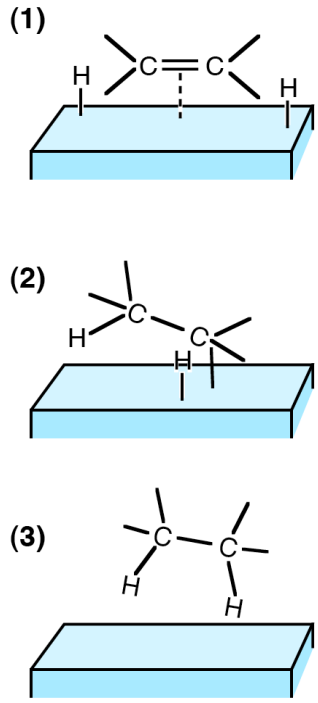


Figure 1.5: Hydrogenation of ethene on a catalytic solid surface, (1) adsorption, (2) Reaction, and (3) desorption.

1.2. Laccases

1.2.1. General aspects

Oxidoreductases (EC 1) are enzymes capable to catalyze redox reactions by using cofactors like metal ions, flavins or hemes (Martínez et al., 2017). Due to the enormous amount of reactions catalyzed by oxidoreductases the industrial interest for this superfamily has increased exponentially.

It is a big enzymatic family with many biologically relevant groups inside, heme-dependent proteins (like peroxidases), flavoproteins, oxygenases (like lytic polysaccharide monooxygenases - LPMOs) or laccases (**Fig. 1.6**) (Lomize, Hage, & Pogozheva, 2018; Lomize, Lomize, Krolicki, & Pogozheva, 2017).

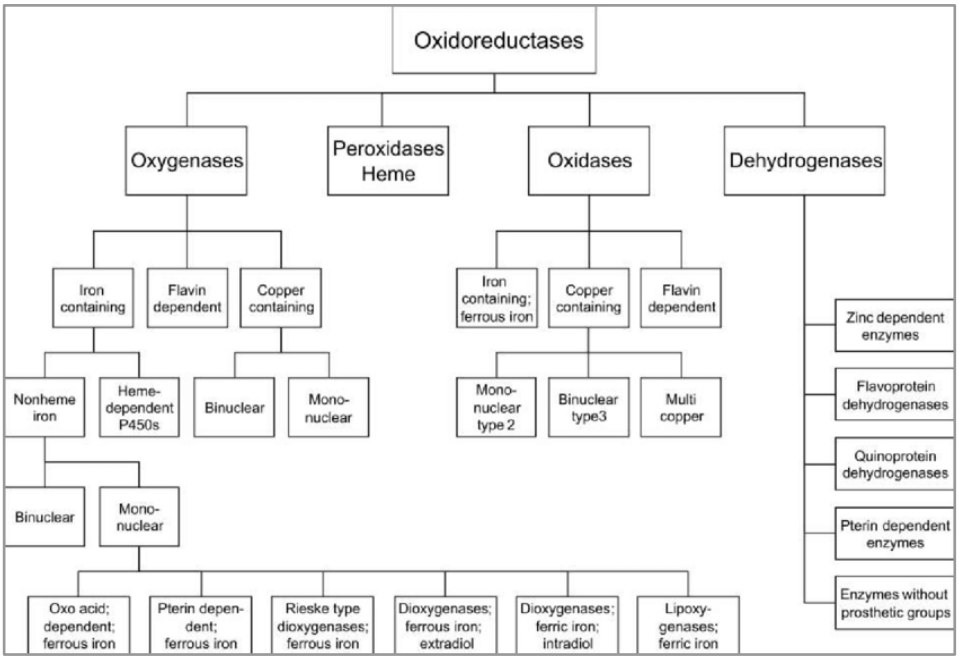


Figure 1.6: Oxidoreductases classification (Blank, Ebert, Buehler, & Bühler, 2010).

Inside oxidoreductase family, laccases are one of the most promising members from a biotechnological point of view. Laccases (EC 1.10.3.2) are oxidases with four coppers: a trinuclear cluster buried in the structure, where molecular oxygen will be reduced to water, and a single copper molecule close to the surface, where the substrate oxidation will take place (Claus, 2004). After the oxidation happens, the enzyme re-generates in the trinuclear cluster reducing molecular oxygen to water (**Fig 1.7**).

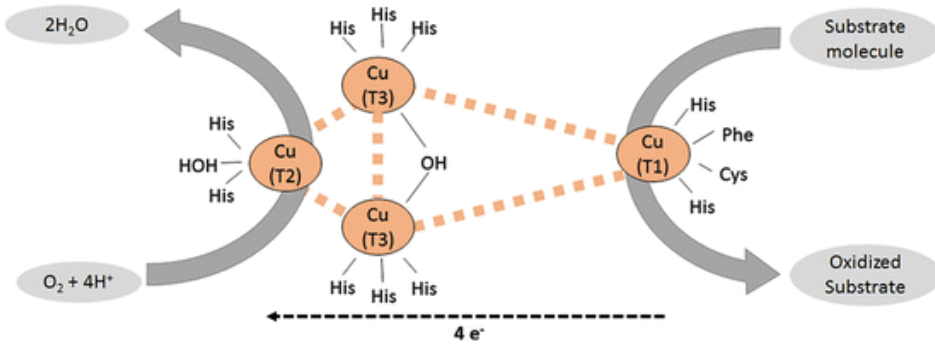


Figure 1.7: Laccases reaction mechanism simplified scheme (Rodríguez-Delgado & Ornelas-Soto, 2017).

The broad ranges of possible reactions and the absence of external cofactors, with water as the only by-product, increase the industrial potential of laccases. Originally, laccases were located in eukaryotic organisms, but more recently they have been described in bacteria as well (Claus, 2003).

1.2.2. Structure

Laccases often are expressed as isozymes, forming large and stable multimeric complexes. The high stability is probably due to the glicolisation rate (10-45%). Buried inside the structure, as stated before, laccases have four different coppers, distributed in three different types (**Fig. 1.8**).

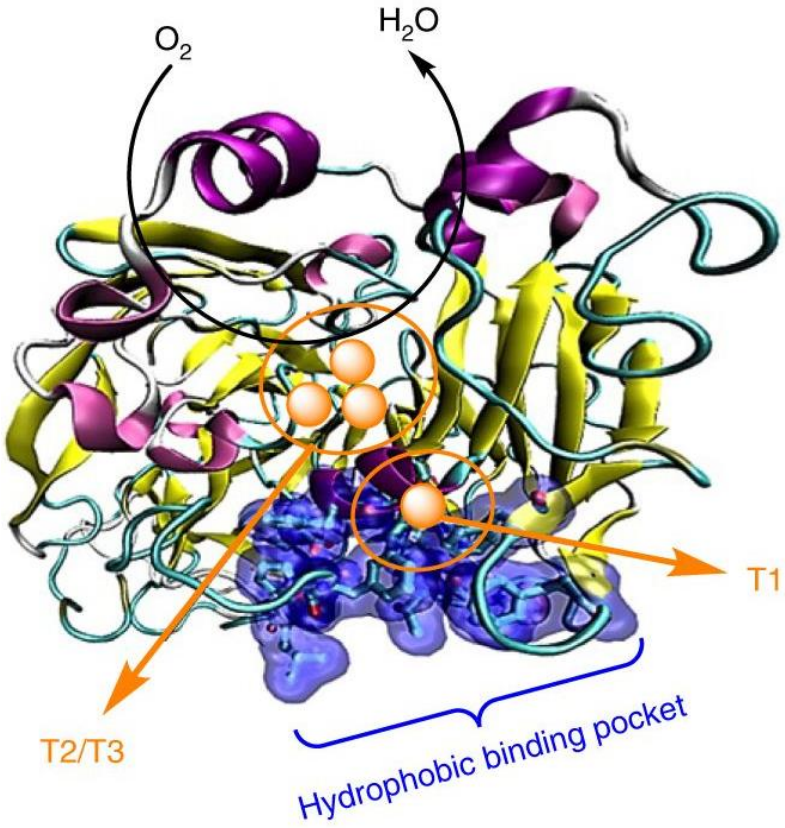


Figure 1.8: Laccase three-dimensional overview (Elouarzaki, Cheng, Fisher, & Lee, 2018).

Type 1 copper has trigonal coordination, with a cysteine (methionine in bacteria) and two histidines. The cysteine-copper covalent bond is also responsible for the typical laccase blue color in solution. Type 2 and 3 are forming the trinuclear cluster, where reduction of molecular oxygen to water takes place. Type 2 holds two histidine coordinations and type 3 holds four (**Fig. 1.9**). These multiple copper centers drive electrons without releasing toxic peroxide intermediates, accomplished by four mono-electronic oxidations of the substrate catalyzed by the type 1 copper.

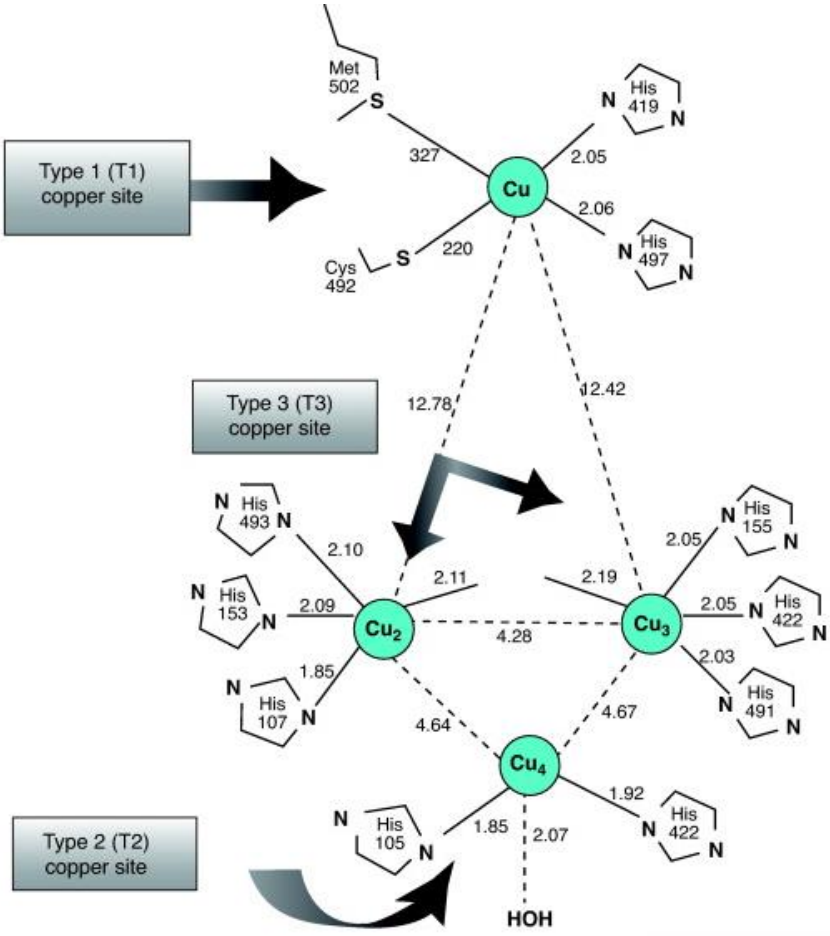


Figure 1.9: Laccase copper first coordination sphere (Santhanam, Vivanco, Decker, & Reardon, 2011).

There are many examples of laccases catalytic potential, polymer formation (Fodor et al., 2016), aromatic compounds oxidations (Zeng, Zhu, Wu, & Lin, 2016), etc.

1.2.3. Applicability and engineering strategies

Laccases have a broad range of applications in several fields, food industry, pulp and paper industry, textile industry, soil remediation, synthetic chemistry or cosmetic industry. Their effectiveness in persistent organic pollutants degradation has received considerable attention in bioremediation related fields (Catherine, Penninckx, & Frédéric, 2016; Gasser, Ammann, Shahgaldian, & -X. Corvini, 2014; Majeau, Brar, & Tyagi, 2010; Strong & Claus, 2011; Viswanath, Rajesh, Janardhan, Kumar, & Narasimha, 2014), for example, biosensors for environmental pollution detection and monitoring (Rao, Scelza, Acevedo, Diez, & Gianfreda, 2014). Laccases are useful for tracking down pollutants such as dyes (Sen, Raut, Bandyopadhyay, & Raut, 2016), polycyclic aromatic hydrocarbons (Librando & Pappalardo, 2013), endocrine disruptors (Husain & Qayyum, 2013) or even antibiotics. Antibiotics are among the most used classes of drugs by our society, and the non-metabolized one end up in the environment (Larsson & Joakim Larsson, 2014). For this purpose, common water treatment is totally useless and cannot remove antibiotics (Oulton, Kohn, & Cwiertny, 2010), while other potential treatments came out with undesirable secondary effects like high implementation costs or extra pollutants formation (Y. Chen, Stemple, Kumar, & Wei, 2016).

Laccases can be successfully applied for bioremediation of hard chemicals or organic compounds, such as antibiotics or polycyclic aromatic hydrocarbons (PAH). Importantly, their catalytic performance and applicability can be expanded by raising T1-copper redox potential by using the laccase-mediator system (Ding et al., 2016; Shi, Ma, Han, Zhang, & Yu, 2014; Suda, Hata, Kawai, Okamura, & Nishida, 2012). By using intermediate molecules - with high redox potential-, laccases can react over difficult substrates, such as the already mentioned antibiotics and PAHs or complex polymers lignin or cellulose fibers (Bourbonnais & Paice, 1990) (**Fig. 1.10**); a laccase-mediator system is

then efficiently expanding enzyme substrate range (Morozova, Shumakovich, Shleev, & Iaropolov, 2007).

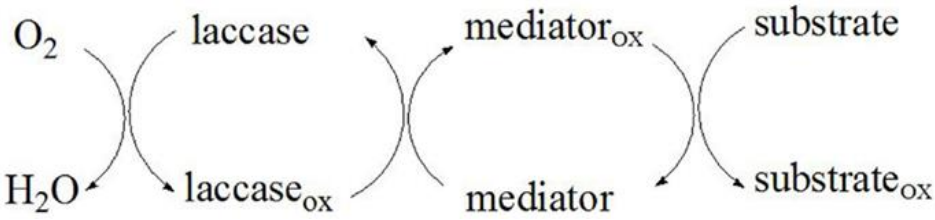


Figure 1.10: Laccase mediator system scheme (Christopher, Yao, & Ji, 2014).

Pulp and paper industry has also been taking advantage of laccase catalysis. Chlorine and oxygen-based chemical oxidants are used for the preparation of paper at an industrial level, specifically for the separation and degradation of lignin fibers. This classical chemical approach came out with several problems such as recycling, cost, and toxicity, which remain unsolved. However, in the existing bleaching process, the previously mentioned laccase-mediator system could be easily implemented, leading to a partial replacement of chlorine and oxygen-based chemical on this industry (Rodríguez-Couto, 2019).

Laccases have also been used in food processing for the elimination of undesirable phenolic compound in baking, juice processing or wine stabilization (Rodríguez Couto, Couto, & Herrera, 2006). The application of laccases over these products modulates flavors and aromas giving taste and color to them. Notice that color and aroma can change due to polymerization of phenolic molecules (Ribeiro, Henrique, Oliveira, Macedo, & Fleuri, 2010).

As mentioned before, laccases have many interesting applications and, due to potential industrial uses, they have been targeted by many enzyme engineering efforts. The first engineering bottleneck was a crystallographic structure determination and since 1998, with Ducros et al. work (Ducros et al., 1998), a

few active laccase structures have been obtained (Hakulinen & Rouvinen, 2015). Using this structural information, mutations around the copper could be done, studying the importance of different residues for laccase activity (Durão et al., 2008)(Z. Chen et al., 2010)(Gupta et al., 2012). Once initial key residues were identified, two main engineering approaches have been applied on laccases:

1) Raising T1 copper redox potential
2) Improving substrate-enzyme interaction

In 1999, Solomon and collaborators developed a laccase variant with an increased T1 redox potential (Xu et al., 1999), these enhanced variants were obtained through site-directed mutagenesis. Similar results were obtained through directed evolution protocols (Maté et al., 2010) and T1 copper redox potential was raised as well. But increasing redox potential has shown to be significantly difficult. This bottleneck partially shifted the research to improve substrate-enzyme interaction. In literature, there are many examples of this type of research with many different methodologies (even a database (Sirim, Wagner, Wang, Schmid, & Pleiss, 2011)), from site-directed mutagenesis to directed evolution and the recent addition of molecular modeling and bioinformatics (Mate & Alcalde, 2015; Mateljak et al., 2019). This addition has proved to be a successful sinergy. Mutations in the binding event can shift the ligand into a more buried position and provide a more favorable electrostatic environment for the oxidation to happen (Monza et al., 2015).

1.3. Serine Esterases

1.3.1. General aspects

Carboxylesterases or serine hydrolases (EC 3.1.1.1) constitute part of the enzymes capable to cleave the carboxyl ester bond and release the corresponding carboxylic acid and alcohol. This reaction happens via a proton transfer lead by a serine part of a catalytic triad, which includes an additional histidine and aspartic acid, with the formation of two tetrahedral intermediate states and an acyl-enzyme phase (J. Aranda et al., 2014) (**Fig. 1.11**). The reaction only requires the catalytic triad and water.

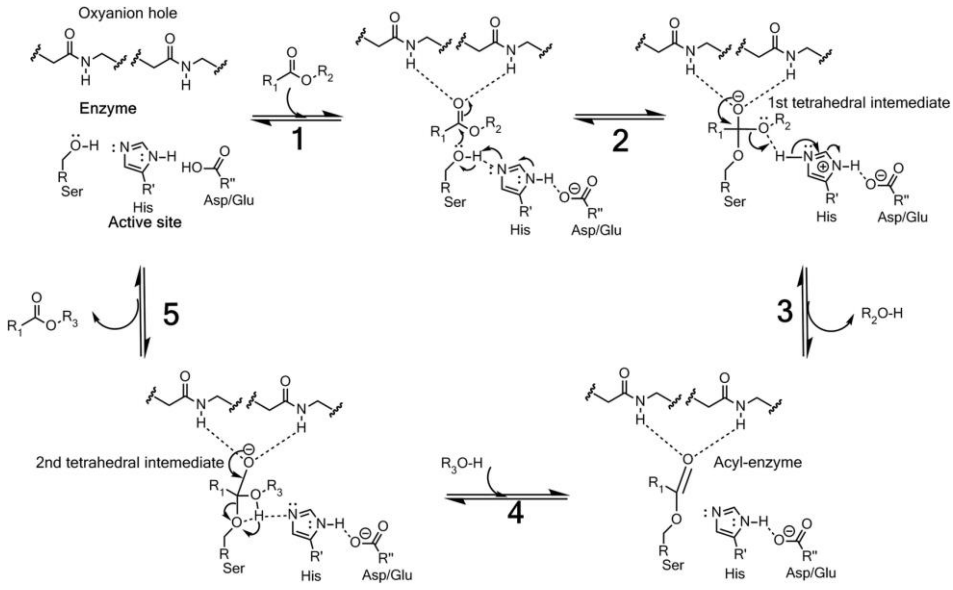


Figure 1.11: A catalytic triad general system as commonly found in serine hydrolases and hydrolysis steps. The acid residue (commonly glutamate or aspartate) aligns and polarises the base (usually histidine) which activates the nucleophile (often serine or cysteine, occasionally threonine) (Armendáriz-Ruiz, Rodríguez-González, Camacho-Ruíz, & Mateos-Díaz, 2018).

Serine esterases are widely distributed in the environment, even in extreme locations, and take part in important cellular physiological functions. Besides, they are among one of the most important industrial enzymes. Closely related to esterases we find widely used and important enzyme families, such as the lipases or proteases, which share catalysis and structural features.

1.3.2. Structure

Several high-resolution three-dimensional structures have been solved by X-ray crystallography and other methods (Akoh, Lee, Liaw, Huang, & Shaw, 2004). Based on this research some common structural features of these enzymes can be described (**Fig. 1.12**):

1) Serine esterases have the classical alpha/beta folding (Claus, 2004; Lenfant et al., 2013). This type of tertiary structure is composed of a main core formed of predominantly parallel beta-strands surrounded by alpha-helices (Nardini & Dijkstra, 1999; Ollis et al., 1992; Schrag & Cygler, 1997).
2) The nucleophilic serine of the catalytic triad rests at a hairpin turn between a beta-strand and an alpha-helix, in an extremely conserved peptide sequence: Gly-X-Ser-X-Gly, forming a characteristic sequence motif commonly named the 'nucleophilic elbow'. There is one exception, Lipase B from *Candida Antarctica* (usually known as CALB) does not present this sequence motif (Uppenberg, Hansen, Patkar, & Alwyn Jones, 1994).
3) Serine hydrolases' active site is formed by three residues called the catalytic triad consisting of serine, histidine and aspartic acid or glutamic acid (Brady et al., 1990; Winkler, D'Arcy, & Hunziker, 1990). Despite identical chemical composition, lipases and proteases (two serine hydrolases) structurally differ by having different orientations,

resulting in inverted stereochemistry of the catalytic triad (Dodson, Guy Dodson, Lawson, & Winkler, 1992).

4) Substrate binding is aided by two amino groups C terminal neighbors of the catalytic serine forming the oxyanion hole (Lang & Dijkstra, 1998; Lang, Mannesse, De Haas, Verheij, & Dijkstra, 1998).

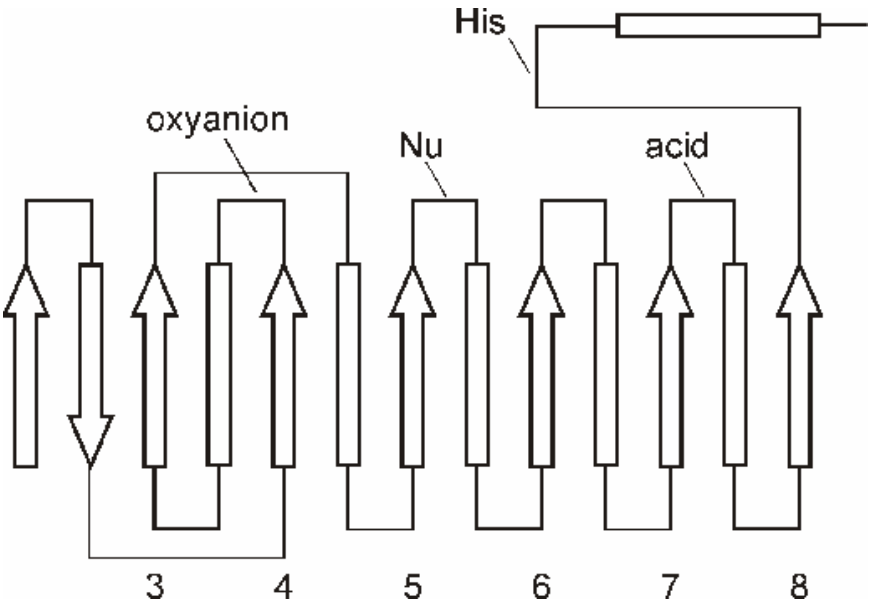


Figure 1.12: Schematic diagram of α/β-hydrolase folding and serine hydrolases main elements. Oxyanion: residues that stabilize oxyanion, Nu: nucleophilic residue; for lipase esterases, and proteases this is a serine; α-helices are shown as rectangles, β-sheets as arrows (Uwe T. Bornscheuer & Kazlauskas, 2005).

1.3.3. Applicability and engineering strategies

Lipases, esterases or proteases are among the most used industrial enzymes, with several industrial applications in many different fields (**Fig. 1.13**), even though they, especially lipases, do not express easily in classical bacterial expression systems (Valero, 2012). For example in food processing, beverages or perfume industries molecules like ferulic, sinapic, caffeic, and coumaric acids are widely

used (Karra-Chaabouni, Pulvin, Touraud, & Thomas, 1996). The esters of these acids are easily found in cereals, agro-industrial residues, and biopulp (Asther et al., 2002). Using feruloyl and cinnamoyl esterases of *Aspergillus niger* (among other possible candidates) the hydroxycinnamic acids can be released from those scratch materials. Other esterases, such as the one obtained from *Fusarium oxysporum*, play an important role producing flavoring and fragrance compounds from geraniol and fatty acids (Asther et al., 2002; Christakopoulos et al., 1998).

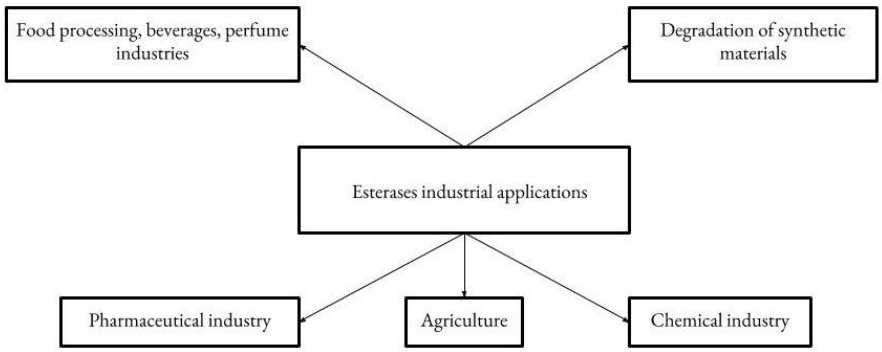


Figure 1.13: Esterases most relevant industrial applications (Panda & Gowrishankar, 2005).

Esterases and lipases are extensively employed in the dairy industry, and for production of wine, beer or juice fruits. They are used for transesterification purposes, where low-value oils and fats are transformed into more valuable ones. As an example, *Lactobacillus casei* CL96 esterases and lipases are significantly used to hydrolyzed milk fat for flavor improvement in cheese-related products (Y.-J., -J., & B., 2001). *Pseudomonas fragi* serine hydrolases were submitted to different physical and chemical treatments to stimulate strawberry flavor

production (Kermasha, Bisakowski, Ismail, & Morin, 2000). In general, esterases play an important role in the flavor industry (Fukuda et al., 2000).

More recently, serine hydrolases are being used for plastic polymer degradation. To degrade such problematic components, cholesterol esterase and polyurethanes (and others) are widely used (Howard, Crother, & Vicknair, 2001; Jahangir et al., 2003). Recently, a serine hydrolase with polyethylene terephthalate (PET) as natural substrate was isolated from *Ideonella sakaiensis*, crystallized (Yoshida et al., 2016) and further rationally designed to obtain improved binding affinities (Austin et al., 2018). Since PET is one of the most abundant plastics in the world, this finding presents a new opportunity for serine hydrolases.

Esterases and other serine hydrolases have also applications in the pharmaceutical industry. They play a major role in, for example, chiral synthesis, an important process in some drug refinement step (Uwe T. Bornscheuer, 2002b). An example of this relevance, is an esterase from *Trichosporon brassicae*. This particular enzyme can produce optically pure (S)- and/or (R)-ketoprofen [2-(3-benzoylphenyl) propionic acid], which is very effective in the reduction of inflammation and relief of pain resulting from arthritis, sunburn, menstruation, and fever (Shen, Xu, Wu, & Liu, 2002). As well, commercially available anti-inflammatory drug (NSAID), ibuprofen (R, S)-2-(4-isobutylphenyl) propionic acid is produced with an esterase from *Pseudomonas sp.* S34 (G.-J. Kim et al., 2002; Y.-H. Kim, Cha, & Cerniglia, 2002).

Agriculture is also an application field for esterases. Phosphodiesters are synthetic organophosphorus compounds, with broad applications as insecticides and nematicides. But after use, remaining residues of these compounds are extremely toxic, causing crop contamination. To fix this problematic,

phosphodiesterases from *Brevundimonas diminuta* and *Alteromonas sp.* have been extensively used in land detoxification by degrading these organophosphorus compounds (Horne, Harcourt, Sutherland, Russell, & Oakeshott, 2002). Further research has been carried out in this particular field and, currently, many different esterases have been identified as potential candidates to be used in land detoxification (Manco, Porzio, & Suzumoto, 2018).

Esterases are extensively applied in other industrial fields, such as the pulp and paper, textile, leather, and baking industries. Steryl esterase and cholesteryl esterase from *Pseudomonas sp.*, *Chromobacterium viscosum* and *Candida rugosa* are playing a significant role in reducing pitch problems during paper production (Kontkanen, Tenkanen, Fagerström, & Reinikainen, 2004).

Due to all the previous applications, esterases have been an important target for enzyme engineering. Improvement of the catalytic performances using directed evolution and/or rational design has been the main goal in the development of esterases as enantioselective biocatalysts (U. T. Bornscheuer, Bottcher, & Schmidt, 2009; Uwe T. Bornscheuer, 2002a). Furthermore, the development of complementary techniques (like bioinformatics, protein crystallography or molecular modeling) for deeply analyzing the mechanisms of enzymatic activity, makes possible today an extensive use of rational, semi-rational and random protein mutation of esterases (Jochens et al., 2011).

An example of enzyme engineering in esterases, involved 3-buten-2-yl acetate racemic resolution. From the numerous wild-type ester hydrolases tested, the best performance was obtained with an esterase from *Pseudomonas fluorescens (Baumann, Hauer, & Bornscheuer, 2000)*. When directed evolution was carried over this enzyme, a double mutant was found, improving enantioselectivity and obtaining higher reaction rates (Schmidt et al., 2006). In this same esterase, four positions were recognized using the 3DM database (Kuipers et al., 2010), and

subjected to saturation mutagenesis for improving the enantioselectivity on 3-phenyl butyric acid esters and a tailored variant was obtained in terms of improved reaction rates and enantioselectivity (Cheeseman, Tocilj, Park, Schrag, & Kazlauskas, 2004; Jochens & Bornscheuer, 2010).

1.4. Enzyme Engineering

As it has been noticed in previous subchapters, enzymes have multiple applications; this industrial interest has been key for driving improving efforts. Enzyme engineering has grown alongside the technology improvement and business increasing. Novel enzymes have appeared in the last years through the new methodologies, like directed evolution (Arnold & Volkov, 1999) or "*de novo*" designs (Jiang et al., 2008). This new generation has increased by thousands the industrial enzymatic relevance.

Following an "historical" path, enzyme engineering started by developing point mutations on the primary structure of a protein using the first crystal structures. The previous irruption of genetic editing molecular techniques, like site-directed mutagenesis (a polymerase chain reaction (PCR) variation), opened such a possibility. This rational design, introduced through site-directed mutagenesis, mandates a detailed knowledge of the protein structure, its function and mechanism (Ridong Chen, 1999) (**Fig. 1.14**). The power of this type of designs has been shown on several occasions by the generation of an enhanced superoxide dismutase (an already fast enzyme) (Getzoff et al., 1992) or the complete inversion of coenzyme specificities for both isocitrate and isopropylmalate dehydrogenases (R. Chen, Greer, & Dean, 1996; Ridong Chen, 1999; Hurley, Chen, & Dean, 1996), as some relevant examples, where individual substitutions generated new and improved enzyme variants.

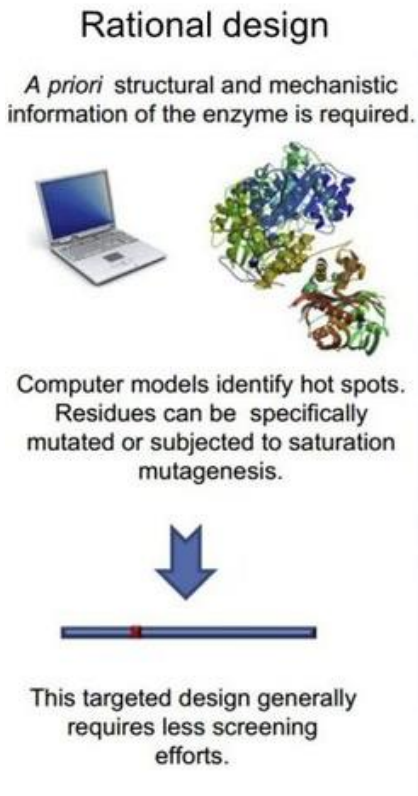


Figure 1.14: Rational design protocol summary (Eriksen, Lian, & Zhao, 2014).

Directed evolution is an opposite concept to rational design, it does not require any information about enzyme structure or how it relates to catalytic function (Kushner & Arnold, 1997; Stemmer, 1994). Directed evolution techniques mimic natural evolution but with a "human touch": it is laboratory accelerated by focusing on individual genes expressed in fast-growing microorganisms (**Fig 1.15**). Already existing proteins (from natural or engineered sources) are selected as the starting point, then mutations are randomly introduced and the progeny screening reveals (or not) enhanced variants. These new variants can then act as parent types in further evolution protocols (Arnold, 2018).

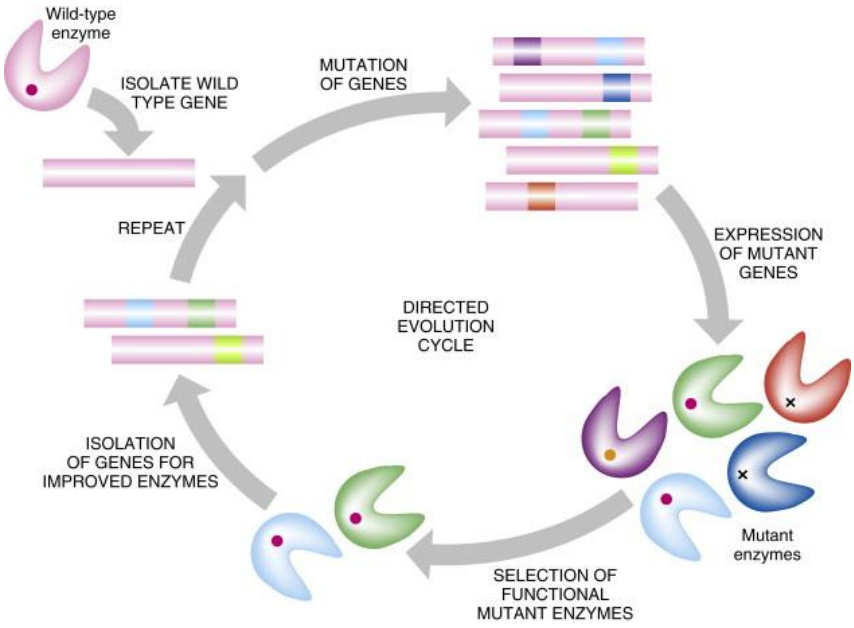


Figure 1.15: General directed evolution pipeline (Clark & Pazdernik, 2016).

There are many different well-settled protocols to mimic natural evolution, one of the firsts was “error-prone” PCR, where through different media conditions, polymerase are forced to introduce mutations in the expanding sequence (McCullum, Williams, Zhang, & Chaput, 2010). An alternative initial procedure involved using chemical and physical agents to randomly damage DNA (Bridges & Woodgate, 1985; Freese, 1959; Lai, Huang, Wang, Li, & Wu, 2004). Following these approaches, new techniques appeared that require genetic recombination, like DNA shuffling, where two genes are combined using endonucleases to generate gene fragments and then combining them through a DNA ligase (Zhao, 1997). Also genetic recombination dependent, Mutagenic Organized Recombination Process by Homologous IN vivo Grouping (MORPHING) is a random mutagenic method for short protein regions, where libraries can be prepared with different mutational loads in DNA segments of less than 30 amino acids so that they can be reassembled into the remaining

unaltered DNA regions (Gonzalez-Perez, Molina-Espeja, Garcia-Ruiz, & Alcalde, 2014).

Since directed evolution is based on random mutagenesis or genetic recombination, there is limited control over resulting outputs. This is connected with directed evolution high throughput screening needs, aiming to explore as much mutagenesis space as possible. It is precisely this limitation (one of) the bridgehead for "in silico" protein engineering.

The addition of computational power and algorithm developments has added a new dimension to enzyme engineering and, these recent years have witnessed new advances in silico aided protein design. From one site, they help us in the understanding of mutagenesis/directed evolution outputs that randomly appears due to the protocols behavior (C. Aranda et al., 2019; Carro et al., 2018; Pardo et al., 2016). The study of enzymatic reactivity has required from first understanding how reactants reach the active site and second how they become products. The description of each step has different theory levels (Marti et al., 2008).

The understanding of how reactants reach the active site, or ligand diffusion, can be investigated at the molecular mechanics (MM) theory level. MM treats atoms as rigid spheres and the bonds between them as strings of variable strength and distance, building all-atoms systems. Using this approach it is possible to simulate complex molecular systems like enzymes. There are many different computational methods using MM, such as molecular dynamics (MD) (Karplus & Petsko, 1990), docking or Monte-Carlo (MC) based algorithms.

MD allows the study of the protein dynamics by the iterative resolution of Newton's equation (Karplus, 2002). At each iteration, the calculation propagates positions and velocities from the potential energy gradients.

Docking simulations involve the prediction of the best bound orientation between two biomolecules. These techniques has been mostly used in the prediction protein-ligand and protein-protein interactions. First, exploration methods provide with several bounding conformations that are then ranked by scoring functions (from MM, empirical data or knowledge-based functions) (Kitchen, Decornez, Furr, & Bajorath, 2004).

MC based methods use the random motion of the system to explore the conformational space, resulting in a non-time dependent trajectory. However, the limitation on proposing collective moves with proper acceptance rates has delay the use of MC techniques in biological systems. Protein Energy Landscape Exploration (PELE) software address this gap through the combination of MC with protein structure prediction, being capable of modeling, as we will see in this thesis, protein-ligand dynamics in enzymology studies (Borrelli, Vitalis, Alcantara, & Guallar, 2005; Sancho, Santiago, Amengual-Rigo, & Guallar, n.d.).

With the application of all these methods its possible to study complex events like protein dynamics or ligand-receptor interactions (Huang, Kalyanaraman, Bernacki, & Jacobson, 2006; Laurent, Breslmayr, Tunega, Ludwig, & Oostenbrink, 2019; Spyrakis et al., 2013).

Reactivity description can be deeply studied with quantum mechanics (QM). The use of this theory level is now capable of describing the reactivity of systems involving several hundreds of atoms with significant accuracy. The application of this theory level in large biological systems like proteins, however, is not yet computationally feasible. For this reason, the modelling of enzymatic reaction is often studied using limited parts of the active site (Lind & Himo, 2013; Siegbahn & Himo, 2009, 2011) or by hybrid methods like QM/MM. The latter method treats with QM the active site region and with MM

the rest of the system, defining a border between both theory levels in the protein. There are many examples of QM/MM approaches to study enzymatic reactivity in for example P450 or chorismate mutase (Senn & Thiel, 2007; Shaik et al., 2010), but it can be applied to understand all reactivities (M. A. Martí et al., 2008). For more complex reaction systems, ranging from reactions of large molecules to chemical processes in solutions, like enzymes, one might use a method like the empirical valence bond, which can be understood as an extension of force-field approaches to chemical reactivity and it is related here to other QM/MM methods (Warshel & Florián, 2004).

Furthermore, all these methodologies can be used to boost enzyme engineering by attempting to change/enhance the activity of existing enzymes and to create new enzymes (Kiss, Çelebi-Ölçüm, Moretti, Baker, & Houk, 2013; S. Martí et al., 2008).

The prediction of mutational effects on the overall of the protein's behavior is a difficult topic. To overcome current and future challenges, computational tools have been developed to predict the consequences of mutations on protein stability, binding affinity or ligand diffusion (Potapov, Cohen, & Schreiber, 2009). It is possible to argue then that the future of enzyme engineering should be based on the ability to better predict all these (not always related) factors.

At this point a new idea has to be introduced, *de novo* enzyme engineering. Then, where there are not natural enzymes, new ones have to be created (Vaissier Welborn & Head-Gordon, 2018). Computational tools allow the design of catalytic active sites based on a transition states generating the theozyme and then dock it in a protein scaffold (Hellinga & Richards, 1991; Malisi, Kohlbacher, & Höcker, 2009; Tantillo, Jiangang, & Houk, 1998; Zanghellini et al., 2006). The following steps will be improving the binding mode of the active site and further characterization and optimization (**Fig. 1.16**). These protocols

introduce new problems that have to be faced and solved to successfully answer this challenge, being the most important the protein folding or the need to find amino acid sequences that fold into the desired three-dimensional structures. The first attempts on this field reduced the complexity of the problem by focusing their efforts on α-helical barrels, with a tertiary structure of helices organized around a central channel with a metal center (Thomson et al., 2014).

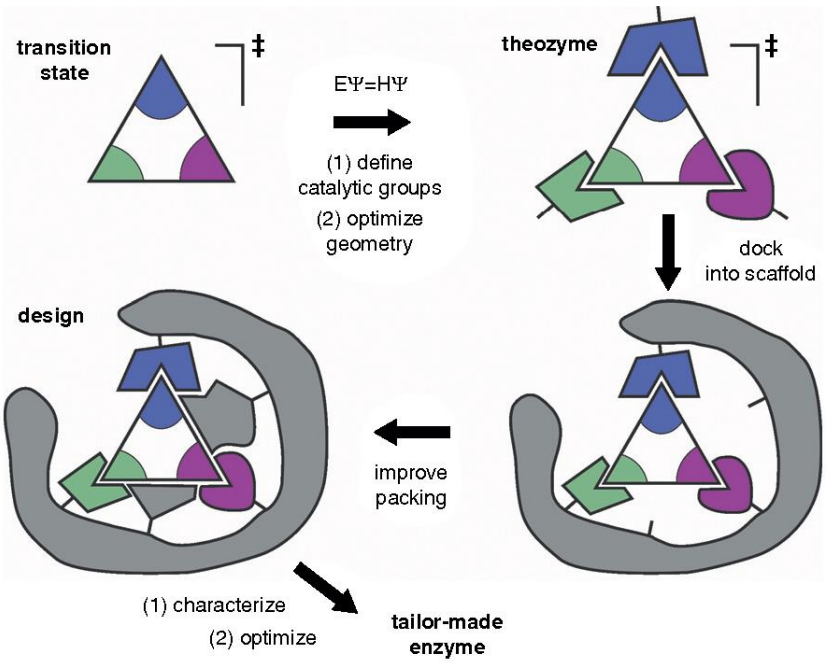


Figure 1.16: Computational de novo design simplified steps (Kries, Blomberg, & Hilvert, 2013).

The initial success encourages more challenging research, increasing design complexity. Mimicking hydrolytic enzymes active sites (based on amino acids) is one example (Bolon & Mayo, 2001). More complex catalytic active sites consisting of a nucleophilic cysteine–histidine dyad and amides groups (backbones preferably) for oxyanion stabilization was de novo designed. The

resulting proteins were found to cleave activated ester substrates by a two-step acylation/deacylation mechanism as reported (Richter et al., 2012).

The main computational *de novo* design limitation is the development from scratch of the required protein folding which presents several problems in the following in vitro validations.

Chapter 2: Objectives

Along **Chapter 1**, we have discussed the enzyme engineering potential over laccases and esterases. Motivated by recent advances in computer-aided enzyme design and the scientific collaboration network in our lab, the main aim of this thesis is: *the development of enhanced enzyme variants for industrial and technological applications.* Our aim is that such biocatalysts will be achieved through the application of state-of-the-art "in silico" techniques, with further experimental validation. To achieve such a goal, we performed several characterizations and engineering studies. The thesis is presented as a "compendium" of research articles that define particular (specific) objectives:

1) **Elucidate the basic principles of enzymatic promiscuity**: Why an enzyme can accept more substrates than another? Can we predict it? The first objective aims to answer these questions. To achieve this goal a large serine-esterases promiscuity dataset has been used to extract enzymatic structural parameters. These parameters have to be rationalized to reach the expected answer. Furthermore, this rationalization will be expanded to another enzyme family, aiming for a general rule for enzymatic promiscuity.

2) **Computer-aided rational enzyme design**: Despite natural (or wild-type) enzymes can catalyze a biotechnology-relevant reaction, they often are not fitted for its direct industrial use. On chosen reactions, and through "in silico" methods, we aim to design new mutations whose effects can improve the overall enzyme performance, and help in their industrial implementation.

3) **Designing of novel enzymes**: This particular objective aims for the development of a "new class" of enzymes, seeking the introduction of additional artificial active sites in currently functional enzymes. Our goal is to boost the overall performance of the enzyme by increasing the

substrate range, activity and, as a long term goal, complementary active sites reactivities (towards cascade reactions).

Chapter 3: Results

In this chapter, we will present the results as article compendia. Each article present genuine investigations carried out during the thesis, where molecular modelling has played a major role in the study of enzymatic properties and in the design of tailored biocatalyst.

Article I: Gerard Santiago et al. Determinants and prediction of esterase substrate promiscuity patterns. ACS Chemical Biology (November 2017).
Article II: Gerard Santiago et al. Computer-Aided Laccase Engineering: Toward Biological Oxidation of Arylamines. ACS Catalysis (July 2016).
Article III: Gerard Santiago et al. Rational engineering of multiple active sites in an ester hydrolase. Biochemistry (March 2018).
Article IV: Gerard Santiago et al. PluriZyme: catalytic and structural advantages of having two active sites in an enzyme (work in revision).

Publications not included in the thesis:

Article V: Isabel Pardo et al. Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. Catalysis Science and Technology (December 2015)
Article VI: Cristina Coscolín et al. Controlled manipulation of enzyme selectivity through immobilization-induced flexibility constraints. Applied Catalysis A: General (August 2018).
Article VII: Felipe de Salas et al. Structural and biochemical characterization of high-redox potential laccase engineered in the lab and comparison with the fungal wild-type (work in revision).
Book chapter I: O2-dependent heme enzyme. Modeling O2-dependent heme enzymes: a quick guide for non-experts (October 2018).

3.1. Determinants and Prediction of Esterase Substrate Promiscuity Patterns

Gerard Santiago‡ζ, Mónica Martínez-Martínez†ζ, Cristina Coscolín†ζ, , Jennifer Chow§, Peter J. Stogios‖ , Rafael Bargiela†γ, Christoph Gertler⊥δ, José Navarro-Fernández†, Alexander Bollinger#, Stephan Thies#, Celia Méndez-García∇ϵ, Ana Popovic‖ , Greg Brown‖ , Tatyana N. Chernikova⊥, Antonio García-Moyano○, Gro E. K. Bjerga○, Pablo Pérez-García§, Tran Hai⊥, Mercedes V. Del Pozo†, Runar Stokke◆, Ida H. Steen◆, Hong Cui‖ , Xiaohui Xu‖ , Boguslaw P. Nocek¶, María Alcaide†, Marco Distaso⊥, Victoria Mesa∇, Ana I. Peláez∇, Jesús Sánchez∇, Patrick C. F. Buchholz%, Jürgen Pleiss% , Antonio Fernández-Guerra$●■, Frank O. Glöckner$●, Olga V. Golyshina⊥, Michail M. Yakimov⌀★, Alexei Savchenko‖ , Karl-Erich Jaeger#α, Alexander F. Yakunin‖ ψ , Wolfgang R. Streit§ψ, Peter N. Golyshin⊥ψ, Víctor Guallar*‡βψ , Manuel Ferrer*†ψ , and The INMARE Consortium

† Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain
‡ Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
§ Biozentrum Klein Flottbek, Mikrobiologie & Biotechnologie, Universität Hamburg, 22609 Hamburg, Germany
‖ Department of Chemical Engineering and Applied Chemistry, University of Toronto, M5S 3E5 Toronto, Ontario, Canada
⊥ School of Biological Sciences, Bangor University, LL57 2UW Bangor, United Kingdom
Institut für Molekulare Enzymtechnologie, Heinrich-Heine-Universität Düsseldorf, 52425 Jülich, Germany
∇ Department of Functional Biology-IUBA, Universidad de Oviedo, 33006 Oviedo, Spain
○ Uni Research AS, Center for Applied Biotechnology, 5006 Bergen, Norway
◆ Department of Biology and KG Jebsen Centre for Deep Sea Research, University of Bergen, 5020 Bergen, Norway
¶ Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, 60439 Illinois, United States
% Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, 70569 Stuttgart, Germany
$ Jacobs University Bremen gGmbH, Bremen, Germany
● Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany
■ University of Oxford, Oxford e-Research Centre, Oxford, United Kingdom
⌀ Institute for Coastal Marine Environment, Consiglio Nazionale delle Ricerche, 98122 Messina, Italy
★ Immanuel Kant Baltic Federal University, 236041 Kaliningrad, Russia
α Institute for Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
β Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

ACS chemical biology

Articles

Cite This: ACS Chem. Biol. 2018, 13, 225−234

Determinants and Prediction of Esterase Substrate Promiscuity Patterns

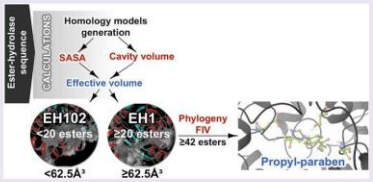
Mónica Martínez-Martínez,[†,⊥] Cristina Coscolín,[†,⊥] Gerard Santiago,[‡,¶] Jennifer Chow,[§] Peter J. Stogios,[∥] Rafael Bargiela,[†,▽] Christoph Gertler,[⊥,δ] José Navarro-Fernández,[†] Alexander Bollinger,[#] Stephan Thies,[#] Celia Méndez-García,[▽,ε] Ana Popovic,[∥] Greg Brown,[∥] Tatyana N. Chernikova,[⊥] Antonio García-Moyano,[○] Gro E. K. Bjerga,[○] Pablo Pérez-García,[§] Tran Hai,[⊥] Mercedes V. Del Pozo,[†] Runar Stokke,[◆] Ida H. Steen,[◆] Hong Cui,[∥] Xiaohui Xu,[∥] Boguslaw P. Nocek,[¶] María Alcaide,[†] Marco Distaso,[⊥] Victoria Mesa,[▽] Ana I. Peláez,[▽] Jesús Sánchez,[▽] Patrick C. F. Buchholz,[%] Jürgen Pleiss,[%] Antonio Fernández-Guerra,[$,●,■] Frank O. Glöckner,[$,●] Olga V. Golyshina,[⊥] Michail M. Yakimov,[ø,★] Alexei Savchenko,[∥] Karl-Erich Jaeger,[#,α] Alexander F. Yakunin,[∥,ψ] Wolfgang R. Streit,[§,ψ] Peter N. Golyshin,[⊥,ψ] Víctor Guallar,[*,‡,β,ψ] Manuel Ferrer,[*,†,ψ] and The INMARE Consortium

[†]Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain
[‡]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
[§]Biozentrum Klein Flottbek, Mikrobiologie & Biotechnologie, Universität Hamburg, 22609 Hamburg, Germany
[∥]Department of Chemical Engineering and Applied Chemistry, University of Toronto, M5S 3E5 Toronto, Ontario, Canada
[⊥]School of Biological Sciences, Bangor University, LL57 2UW Bangor, United Kingdom
[#]Institut für Molekulare Enzymtechnologie, Heinrich-Heine-Universität Düsseldorf, 52425 Jülich, Germany
[▽]Department of Functional Biology-IUBA, Universidad de Oviedo, 33006 Oviedo, Spain
[○]Uni Research AS, Center for Applied Biotechnology, 5006 Bergen, Norway
[◆]Department of Biology and KG Jebsen Centre for Deep Sea Research, University of Bergen, 5020 Bergen, Norway
[¶]Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, 60439 Illinois, United States
[%]Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, 70569 Stuttgart, Germany
[$]Jacobs University Bremen gGmbH, Bremen, Germany
[●]Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany
[■]University of Oxford, Oxford e-Research Centre, Oxford, United Kingdom
[ø]Institute for Coastal Marine Environment, Consiglio Nazionale delle Ricerche, 98122 Messina, Italy
[★]Immanuel Kant Baltic Federal University, 236041 Kaliningrad, Russia
[α]Institute for Bio- and Geosciences IBG-1: Biotechnology, Forschunsgzentrum Jülich GmbH, 52425 Jülich, Germany
[β]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Esterases receive special attention because of their wide distribution in biological systems and environments and their importance for physiology and chemical synthesis. The prediction of esterases' substrate promiscuity level from sequence data and the molecular reasons why certain such enzymes are more promiscuous than others remain to be elucidated. This limits the surveillance of the sequence space for esterases potentially leading to new versatile biocatalysts and new insights into their role in cellular function. Here, we performed an extensive analysis of the substrate spectra of

continued...



Received: November 21, 2017
Accepted: November 28, 2017
Published: November 28, 2017

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

145 phylogenetically and environmentally diverse microbial esterases, when tested with 96 diverse esters. We determined the primary factors shaping their substrate range by analyzing substrate range patterns in combination with structural analysis and protein−ligand simulations. We found a structural parameter that helps rank (classify) the promiscuity level of esterases from sequence data at 94% accuracy. This parameter, the active site effective volume, exemplifies the topology of the catalytic environment by measuring the active site cavity volume corrected by the relative solvent accessible surface area (SASA) of the catalytic triad. Sequences encoding esterases with active site effective volumes (cavity volume/SASA) above a threshold show greater substrate spectra, which can be further extended in combination with phylogenetic data. This measure provides also a valuable tool for interrogating substrates capable of being converted. This measure, found to be transferred to phosphatases of the haloalkanoic acid dehalogenase superfamily and possibly other enzymatic systems, represents a powerful tool for low-cost bioprospecting for esterases with broad substrate ranges, in large scale sequence data sets.

Enzymes with outstanding properties in biological systems and the conditions favoring their positive selection are difficult to predict. One of these properties is substrate promiscuity, which typically refers to a broad substrate spectrum and acceptance of larger substrates. This phenomenon is important from environmental,[1] evolutionary,[2−5] structural,[6−8] and biotechnological[9,10] points of view. The relevance of substrate promiscuity is indisputable as the operating basis for biological processes and cell function. As an example, the evolutionary progress of enzymes from lower to higher substrate specificity allows the recruitment of alternate pathways for carbon cycling and innovations across metabolic subsystems and the tree of life by maximizing the growth rate and growth efficiency.[11] Promiscuous enzymes are energetically more favorable than specialized enzymes,[4] and therefore, the cell does not require many different enzymes to take up substrates, favoring genome minimization and streamlining.[12] In addition, the acquisition of new specificities without compromising primary or ancestral ones is a major driver of microbial adaptation to extreme habitats.[13] From a more practical standpoint, along with requirements of a technical nature such as selectivity, scalability and robustness, a narrow substrate spectrum is one of the most frequent problems for industrial enzyme applications.[14] A consensus exists that "the more substrates an enzyme converts the better," opening application ranges with consequent reduction of the production cost of multiple enzymes.[10,14,15]

Enzymes with wide substrate ranges occur naturally, as systematically investigated for halo-alkane dehalogenases,[16] phosphatases,[1] beta-lactamases,[2,17] and hydroxyl-nitrile lyases.[5] Some enzymes are more promiscuous than others simply due to their fold or degree of plasticity or the presence of structural elements or mutations occurring under selection in the proximity of the active-site cavity and access tunnels favoring promiscuity. However, the general explanation, if any, by which an enzyme binds and converts multiple substrates is unknown, although molecular insights have been reported for single enzymes.[18] A tool that can clearly distinguish promiscuous versus non-promiscuous enzymes and suggest substrates potentially being converted or not by them might therefore be valuable in applying low-cost sequencing in discovery platforms in any biological context.

In an ideal scenario, functional characterization of enzymes with genomics[19] and metagenomics[10,20] techniques using a large library of substrates would guide the analysis of sequence-to-promiscuity relationships and explore the mechanistic basis of promiscuity. In addition, such studies may help identify a new generation of highly promiscuous microbial biocatalysts. However, extensive bioprospecting and biochemical studies are rare,[10] despite the growing number of sequences available through low-cost sequencing efforts[21] and the growing number of enzymes that are typically characterized with limited substrate sets.[14] To address this knowledge gap, we functionally assessed the substrate specificity of a set of 145 phylogenetically, environmentally, and structurally diverse microbial esterases (herein referred to as "EH," which means Ester Hydrolase) against a customized library of 96 different substrates to find predictive markers of substrate promiscuity rather than discrete determinants of substrate specificity that may differ from protein to protein. EHs were selected for an analysis of substrate promiscuity because they typically have specific definitions of molecular function, can be easily screened in genomes and metagenomes compared with many other classes of proteins, are among the most important groups of biocatalysts for chemical synthesis, and are widely distributed in nature, with at least one EH per genome.[14]

Our work adds important insights and empirical, structural, and computational data to facilitate the elucidation of the molecular basis of substrate promiscuity in EHs, which was further extended to phosphatases from the haloalkanoic acid dehalogenase (HAD) superfamily. This was achieved by deciphering what we consider a predictive structural marker of substrate promiscuity and by establishing the reasons why certain such enzymes are more promiscuous than others and can convert substrates that others cannot. This study does not pretend to generate a quantitative measure to predict the number of compounds that an enzyme will hydrolyze but a tool and a parameter that will help in ranking (classifying) promiscuity level. Following on from that, we propose in this work the first molecular classification method of this kind derived from first principle molecular simulations and with clear physical/structural interpretation. This work also provides an example of the utility of this parameter to screen the sequence space for highly promiscuous EHs that may compete with best commercial EH preparations. We also provide first preliminary evidence of a number of underexplored microbial phylogenetic lineages containing EHs with a prominent substrate range.

■ RESULTS AND DISCUSSION

The Substrate Range of 145 Diverse EHs. A total of 145 EHs were investigated. Extensive details of the sources and screen methods are provided in the Supporting Information Methods and Table S1. In an environmental context, the source of enzymes was highly diverse because they were isolated from bacteria from 28 geographically distinct sites (125 EHs in total) and from six marine bacterial genomes (20 EHs; Supporting Information Figure S1). A phylogenetic analysis also indicated that sequences belong to bacteria distributed across the entire phylogenetic tree (Supporting Information Results and Figure S2).

The 145 putative proteins exhibited maximum amino acid sequence identities (Supporting Information Table S1) ranging from 29.1 to 99.9% to uncharacterized homologous proteins in public databases, with an average value (reported as %, with the

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

interquartile range (IQR) in parentheses) of 74.3% (40.3%). The pairwise amino acid sequence identity for all EHs ranged from 0.2 to 99.7% (Supporting Information Table S2), with an average value of 13.7% (7.6%). BLAST searches were performed for all query sequences by running NCBI BLASTP against the current version of the Lipase Engineering Database[22] using an E-value threshold of 10^{-10} and were successful for all but nine candidates. A total of 120 EH sequences were unambiguously assigned to some of the 14 existing families (F) of the Arpigny and Jaeger classification, which are defined based on amino acid sequence similarity and the presence of specific sequence motifs.[14,23] These EHs included sequences with a typical α/β hydrolase fold and conserved G-X-S-X-G (FI, 20; FIV, 36; FV, 33; FVI, 5; and FVII, 6) or G-X-S-(L) (FII, 9) motifs and sequences with a serine beta-lactamase-like modular (non-α/β hydrolase fold) architecture and a conserved S-X-X-K motif (FVIII, 11). An additional set of nine sequences were assigned to the *meta*-cleavage product (MCP) hydrolase family[24] and six to the so-called carbohydrate esterase family,[25] both with typical α/β hydrolase folds. Finally, one was a cyclase-like protein from the amido-hydrolase superfamily.[26] Sequences-to-family assignments are summarized in the Supporting Information Table S1. Taken together, the primary sequence analysis suggests that the diversity of polypeptides is not dominated by a particular type of protein or highly similar protein clusters but consists of diverse nonredundant sequences assigned to multiple folds and subfamilies, which are distantly related to known homologues in many cases.

The substrate profiles of all EHs were examined using a set of 96 chemically and structurally distinct esters (Supporting Information Table S3). We are aware that the number of compounds hydrolyzed may be an ambiguous indicator of promiscuity, because the size and composition of the library may influence the results. For this reason, the composition of the library was not random but based on including esters with variation in size of acyl and alcohol groups and with growing residues (aromatic, aliphatic, branched, and unbranched) at both sides, leading to more challenging substrates because a larger group adjacent to the ester bond increases the difficulty of conversion. Halogenated, chiral, and sugar esters, lactones, and an alkyl diester were also included. Esters with nitro substituents were not included. We used the partitioning coefficient (log P value) to indicate the chemical variability of the esters because this parameter reflects electronic and steric effects and hydrophobic and hydrophilic characteristics. Log P was determined with the software ACD/ChemSketch 2015.2.5. Log P values (Supporting Information, Table S3) ranged from −1.07 (for methyl glycolate) to 23.71 (for triolein), with an average value (IQR in parentheses) of 3.13 (2.86), which indicates that the ester library used in this study had broad chemical and structural variability. Nevertheless, adding new substrates could surely help (and even change) the ranking of the EHs herein analyzed. The dynamic range of the assay may also influence the results. For this reason, to detect enzyme−substrate pairs for a given EH, the ester library was screened with each of the 145 EHs in a kinetic pH indicator assay in 384-well plates,[24,27,28] which unambiguously allow quantifying specific activities at pH 8.0 and 30 °C, using a substrate concentration above 0.5 mM (see Supporting Information, Results). Two commercial lipases, CalA and CalB from *Pseudozyma aphidis* (formerly *Candida antarctica*), were included in the assays for comparison. Using this data set, we linked the biocatalytic data to the sequence information for the respective enzyme. In this study, sequence information meant any sequence that encoded an EH of interest. Biocatalytic data meant experimental data on substrate conversion (i.e., units g^{-1} or U g^{-1}) followed for 24 h.

We determined the probability of finding an EH with a broad substrate profile by plotting the number of esters that were hydrolyzed by all preparations. Figure 1 shows that the number

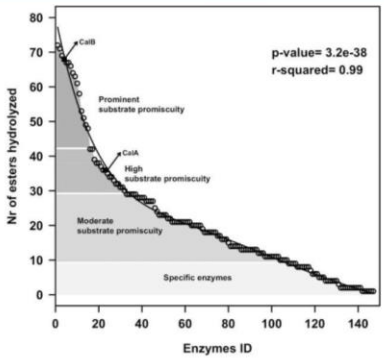


Figure 1. Number of ester substrates hydrolyzed by each of the 145 EHs investigated in this study. The commercial preparations CalA and CalB (marked with filled square) are also included. This figure is created from data in the Supporting Information Table S1. The activity protocol established and used to identify the esters hydrolyzed by each EH was based on a 550 nm follow-up pH indicator assay described in the Supporting Information Methods. The list of the 96 structurally different esters tested is shown in Figure 2. Full details of the activity protocol are provided in the Supporting Information Methods. The trend line shows a not-single exponential fit of the experimental data. The fit was obtained using R script and the "lm" function, to extract a polynomial regression of degree 6 with the following line "model ← lm(MM[,1] ~ poly(MM[,2],6,raw = TRUE))", where MM[,1] corresponds to the number of esters hydrolyzed, and MM[,2] the position in the *x* axis (from 1 to 147).

of esters hydrolyzed by all 147 EHs (including CalA/B) fits to an exponential distribution (r^2 = 0.99; p value 3.2e^{-38}; Pearson's correlation coefficient) with a median of 18 substrates per enzyme, nine hits at the 25th percentile, and 29 hits at the 75th percentile. On the basis of this distribution and a previously established criterion,[1] we considered an enzyme specific if it used nine esters or fewer (27% of the total), as showing moderate substrate promiscuity if it used between 10 and 29 esters (51% of the total), and as showing high-to-prominent promiscuity if it used 30 or more esters (22% of the total). This criterion indicated a percentage of EHs with a prominent substrate range similar to that found for HAD phosphatases (24%).[1]

Phylogeny Is a Predictive Marker of Substrate Promiscuity. Hierarchical clustering was performed to evaluate the differences in substrate range patterns (Figure 2). For the sake of simplicity, clustering was performed for those EHs that hydrolyzed 10 or more esters (i.e., 107 total EHs). We first observed a large percentage of enzymes with presumptive broad active site environments that accommodated large aromatic and sterically hindered esters such as benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate (49% of the total), benzoic acid-4-formyl-phenylmethyl ester (27%), 2,4-dichlorophenyl 2,4-dichlorobenzoate (~8%), 2,4-dichlorophenyl 2,4-dichlorobenzoate (~5%), and diethyl-2,6-dimethyl 4-phenyl-1,4-dihydro

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

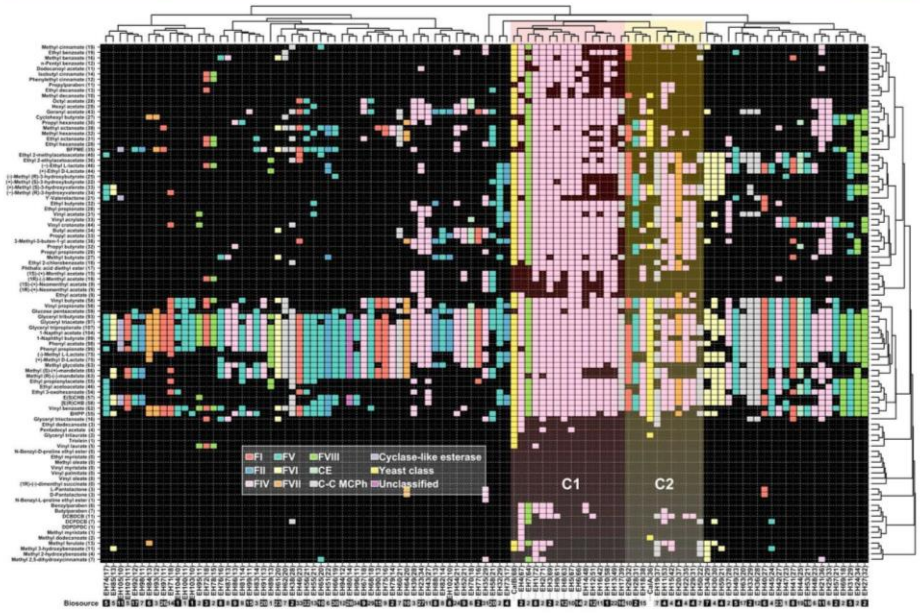


Figure 2. Hierarchical clustering of the substrate ranges of the EHs. Only EHs that hydrolyzed 10 or more esters were considered (107 in total, including CalA/B). This figure is created from data in the Supporting Information Table S3. The specific activities of the EHs for each of the 96 esters were determined as described in Figure 1. The list of the 96 esters tested and the frequency of each ester considered as a hit (in brackets) are shown on the left side. The ID code representing each EH is given at the bottom. Each hydrolase is named based on the code "EH," which means Ester Hydrolase, followed by an arbitrary number from 1 to 145 for the most to least promiscuous enzyme. The number in brackets indicates the number of esters hydrolyzed by each enzyme. The biosource of each EH is indicated at the bottom with a number in white or black squares that follows the nomenclature in the Supporting Information Figure S1. The figure was created with the R language console using a binomial table with information about the activity/inactivity (1/0) of the analyzed enzymes against the 96 substrates as a starting point. For the central graphic, which shows the data in Supporting Information Table S3, we used the drawing tools provided by the basic core packages of R. The hierarchical clusters of the enzymes (shown at the top) and substrates (shown on the right side) were generated by calculating a distance matrix using a "binomial" method and the hclust function to generate the tree. Using the functions as .phylo and plot.phylo from the ape package, the clusters were added to the top and right of the figure. A combination of the Set1 palette from the R package RColorBrewer and colors from the basic palette from R were used as the color palette for sequences assigned to each family (F; see inset), including FI to FVII, carbohydrate esterase (CE), and carbon−carbon *meta*-cleavage product hydrolase (C−C MCP) families, all with a typical α/β hydrolase fold, FVIII serine beta-lactamase with non α/β hydrolase fold, and cyclase-like protein from the amido-hydrolase superfamily. Sequences that were not unambiguously ascribed to existing families were referred to as "Unclassified," and those of yeast origin were assigned to "yeast class." The two "clusters" C1 and C2 that contained the most substrate-promiscuous EHs are color-coded under a shadowed background.

pyridine-3,5-dicarboxylate (~1%). Therefore, even though the EHs in this study were identified by a selection process based on the utilization of short esters (see Supporting Information Methods), the isolation of EHs with ample substrate spectra and the ability to hydrolyze very large substrates was not compromised.

We detected drastic shifts in substrate specificity (Figure 2), with glyceryl tripropionate as the only substrate hydrolyzed by all EHs. This is consistent with the high sequence variability within EHs, with an average pairwise identity of 13.74%. We then sought to determine the primary factors shaping the substrate range and thus defined different functional clusters. First, we observed that global sequence identity was of limited relevance for inferring the substrate range because no correlation was found (r^2 = 0.25) between the differences in identity and the number of esters that were hydrolyzed (Supporting Information Tables S1 and S2). Second, comparisons of the substrate range and the hydrolysis rate (U g^{-1} for the best substrates) were performed (Supporting Information Table S1). No correlation existed (r^2 = 0.073), suggesting that our assay conditions allow evaluating the promiscuity level whatever the hydrolytic rate of the EH is. In addition to the low correlation values, no threshold above or below which one could qualitatively classify the substrate range was observed in both cases, so that sequence identity and hydrolytic rate are neither predictive nor classification parameters of promiscuity. Additionally, no link between substrate range and habitat was found because EHs from the same biosource fell into separate clusters (Figure 2). Phylogeny-substrate spectrum relationships were further examined. Figure 2 indicates that the broad substrate-spectrum EHs did not cluster in a single phylogenetic branch, yet substrate promiscuity was mostly found for members of one of 10 subfamilies covered. Indeed, 67% of the EHs that could hydrolyze 30 or more esters

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

(mostly located in clusters C1 and C2 in Figure 2) were assigned to FIV,[14,23] and this percentage increased to 84% when considering only those EHs that could hydrolyze 42 to 72 esters (Figure 2; cluster C1). In addition to FIV members, a FVIII serine beta-lactamase showed prominent substrate spectra (see cluster 1). Members of both families (FIV, 8; FVIII, 1; see cluster C1) hydrolyzed as many esters (from 61 to 72) as the yeast family member CalB (68 esters), the most promiscuous commercially available lipase preparation used for the production of fine chemicals.[29]

Phylogeny was thus indicated as a predictive marker of the substrate range of EHs, as although a broad substrate scope was assigned to several sequence clusters, this feature was prevalent in members of FIV. A query sequence that matched FIV could be easily identified by means of the consensus motif GDSAGG around the catalytic serine; this family is also called the hormone-sensitive lipase (HSL) family because a number of FIV EHs display a striking similarity to the mammalian HSL.[14,23] Noticeably, the location of some FIV members in functional clusters with narrow substrate spectra (Figure 2) suggests that factors other than phylogeny contribute to the substrate spectra of EHs.

The Active Site Effective Volume Is a Prominent Marker of EH Promiscuity. Structural-to-substrate spectrum relationships were further examined by protein−ligand simulations to find additional markers of promiscuity. Crystals from recombinant EH1,[28] the protein with the broadest substrate range under our assay conditions, were obtained as described in the Supporting Information Methods. The enzyme with the widest substrate range was considered the best candidate for understanding the nature of promiscuity. This enzyme seems to have a wide active site environment as, under our assay conditions, it accepted 72 esters ranging from short (*e.g.*, vinyl acetate) to large (*e.g.*, 2,4-dichlorobenzyl-2,4-dichlorobenzoate; Figure 2). We also obtained crystals of recombinant EH102, which was isolated from the same habitat[28] but had a restricted substrate range, hydrolyzing only 10 of the 96 esters tested (Figure 2). Crystallographic data and refinement statistics for the two structures are given in Supporting Information Table S4.

To rationalize the substrate range shown by EH1 and EH102, we performed substrate migration studies using the software Protein Energy Landscape Exploration (PELE), which is an excellent tool to map ligand migration and binding, as shown in studies with diverse applications.[30−32] To map the tendency of a substrate to remain close to the catalytic triad, the substrate was placed in a catalytic position, within a proton abstraction distance from the catalytic serine, and allowed to freely explore the exit from the active site. The PELE results for both proteins and glyceryl triacetate are shown in Figure 3a. Clearly, EH1 has a significantly better binding profile, with an overall lower binding energy and a better funnel shape, whereas EH102 had a qualitatively unproductive binding-energy profile. This difference in the binding mechanism can be explained by the catalytic triad environment. EH1 has a somewhat wide but buried active site, whereas EH102 has a surface-exposed catalytic triad (Figure 4a). These structural differences translate into significant changes in the active site volume, as defined using Fpocket; the active site cavity of EH1 is 3-fold larger than that of EH102. Moreover, important changes are observed when inspecting the solvent exposure of the cavity. Figure 3b shows the relative solvent accessible surface area (SASA) for the substrate along the exploration of PELE, computed as a (dimensionless) percentage (0−1) of the ligand SASA in solution. Even at catalytic positions (distance Ser(O)−substrate(C) ∼ 3−4 Å), in EH102 we observe that ∼40% of the surface of the substrate is accessible to the solvent, which greatly destabilizes the substrate and facilitates escape to the bulk solvent. By contrast, EH1 has a larger but almost fully occluded site, with relative SASA values of approximately 0−10%, which can better stabilize the substrate.

After defining key points underlying the promiscuity of EH1, i.e., a larger active site volume and a lower SASA (Figure 4a), we extended the analysis to other EHs. First, we collected all 11 available crystal structures (Supporting Information Table S1) and computed the active site volume and relative SASA of the catalytic triad (Figure 5, square symbols). We next extended the analysis to the rest of the EHs using homology modeling (using the 11 crystals available) and produced a structural model for 84 additional enzymes. The missing ones were those with sequence identities of less than 25% (to an existing crystal) or those for which the catalytic triad could not be unambiguously identified (*i.e.*, not suitable alignments). Figure 5 (circle symbols) shows the active site effective volume data for all structural models. The analysis indicated a ratio threshold of 62.5 Å^3 for qualitatively classifying substrate promiscuity. Note that the relative SASA of the catalytic triad (derived from the GetArea server, see Supporting Information Methods) adopts values of 0−100; the actual value of the effective volume threshold will depend on the chosen range. We observed that values equal to or higher

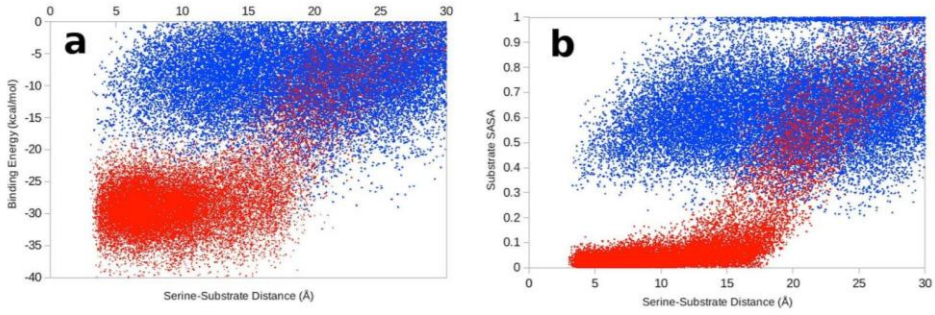


Figure 3. Protein Energy Landscape Exploration (PELE) analysis. Panel a shows the protein−substrate interaction plots for EH1 (red) and EH102 (blue). Panel b shows the relative SASA for glyceryl triacetate in EH1 (red) and EH102 (blue) computed as a dimensionless ratio (0−1) using PELE.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

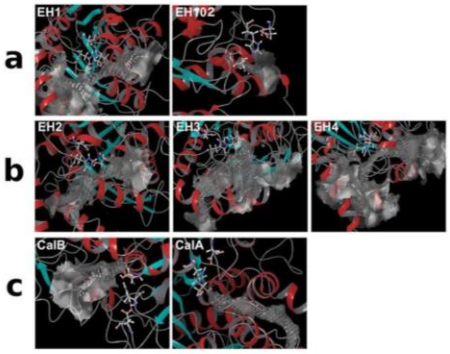


Figure 4. Catalytic triad exposure of selected EHs with the broadest and lowest substrate ranges. (a) The catalytic triad (ball-and-sticks) and the main adjacent cavity (gray clouds) as detected by SiteMap are underlined to demonstrate the differences between a promiscuous (EH1) and nonpromiscuous (EH102) EHs. EH1 can hydrolyze 72 esters and has a defined hidden binding cavity (effective volume: 166.7 Å^3). EH102, by contrast, can hydrolyze only 10 esters and has a surface-exposed triad (high SASA) and an almost negligible binding cavity (38.5 Å^3). The three top EHs with the broadest substrate ranges (b), positioned in the ranking after EH1, and the commercial CalB and CalA lipases (c), are also represented. On each panel, we highlight the catalytic triad and the main adjacent cavity as detected by SiteMap, demonstrating the differences in active site topology. EH2, EH3, and EH4, all assigned to FIV (as EH1), hydrolyzed 71, 69, and 67 esters and have defined but distinct hidden binding cavities (500, 200, and 200 Å^3, in the same order), as EH1. CalB, which was capable of hydrolyzing 68 esters, has a binding cavity (200 Å^3) that is also hidden but highly different from those of the other EHs. CalA, by contrast, hydrolyzed only 36 esters and has a low surface-exposed triad (SASA), with restrictive access to the catalytic triad (1000 Å^3).

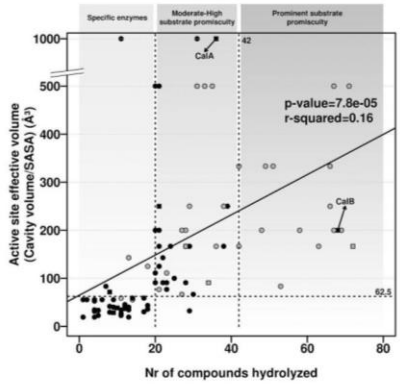


Figure 5. Defining of the substrate range of the EH by topology of the catalytic environment. The figure shows the relationships between the active site effective volume (in Å^3) and enzyme promiscuity (number of substrates hydrolyzed). Note that the presented data were obtained using the active site cavity volume computed in Å^3 and SASA as a dimensionless ratio from 0 to 100 using the GetArea server (http://curie.utmb.edu/getarea.html). The panel contains information for EHs for which crystal structures (square) and homology models (circles) could be unambiguously established (sequence identity ≥25%) and the catalytic triad identified. Gray circles and squares indicate the EHs assigned to FIV. The analysis indicated a threshold ratio (indicated by a horizontal dashed gray line) at which it is possible to qualitatively classify substrate promiscuity based on hydrolysis of at least 20 substrates. Phylogenetic analysis further extended the substrate spectra to ≥42 esters, as only enzymes assigned to FIV and conforming to the 62.5 Å^3 threshold, together with CalB, were capable of converting such a high number of esters. The positioning for the commercial CalA and CalB lipases is indicated.

than 62.5 Å^3 corresponded to EHs with activity for 20 or more of the 96 substrates tested and the opposite. There were only six outliers out of 95 EHs that did not follow this rule. Thus, the performance is of excellent (with 94%) accuracy if used as a classifier. The effective volume, however, does not have quantitative predictions for the exact number of esters hydrolyzed (r^2 = 0.16 for data in Figure 5), most likely because above the 62.5 Å^3 threshold, the capability to hydrolyze more or less substrates may specifically depend on the topology of the catalytic environment (Figure 4a−c), which may differ within families. Particularly, none of the different family members that conformed to the ≥62.5 Å^3 threshold, except those from FIV (*i.e.*, at least 50% of its members as shown in Figure 5, gray circle symbols) and CalB, could hydrolyze 42 or more esters. Therefore, the classification potential of the effective volume measure increased when combined with phylogenetic data. Noticeably, we observed that the predictive capacity of cavity volume/SASA is not influenced by the presence of flexible elements in the structure (Supporting Information Results).

The Active Site Effective Volume Is Also Indicative of Molecules Being Accepted As Substrates. We further used the active site cavity volume/SASA to also dissect its role in substrate specificity. We restricted the analysis to the 96 EHs for which this value could be unambiguously calculated (see above). The analysis indicated that the conversion of 34 esters was only observed for EHs conforming to the ≥62.5 Å^3 threshold (Supporting Information Figure S3). All but two (vinyl crotonate and ethyl acetate) could be considered large alkyl or hindered aromatic esters and included important molecules in synthetic organic chemistry such as paraben esters. This suggests that active sites with larger volume and a lower SASA (i.e., cavity less exposed to the surface) will most likely support hydrolysis of these esters. Therefore, the effective volume measure could be used to some extent as an indicator of substrates that may or may not be hydrolyzed by EHs. However, not all EHs fitting the ≥62.5 Å^3 threshold could convert all 34 of these esters, implying that this measure does not allow deepening into substrate specificity, which may depend on the topology of the catalytic environments as mentioned previously (Figure 4a−c). However, we found that the probability that benzyl-, butyl-, and propyl-paraben esters, major intermediates in chemical synthesis, are converted by members of the FIV with an effective volume ≥ 62.5 Å^3 is significantly higher (~35%) than that of EHs from FIV with a volume < 62.5 Å^3 and EHs from other families, whatever the value of the effective volume (approaching zero percent in our study); for those EHs for which effective volume could not be calculated, this probability is as low as 1.9% (Supporting Information Figure S4). This again exemplifies that the effective volume measure, when combined with phylogenetic information, is not only indicative of a promiscuity level but also can be used to predict the capacity to hydrolyze esters such as paraben esters. Screen programs to find EHs capable of converting

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

paraben esters should most likely be directed to find those assigned to FIV and with cavity volume/SASA $\geq$ 62.5 Å^3.

The Effective Volume Is Also a Marker of Substrate Promiscuity in Proteins Other than EHs. In order to evaluate the possibility that the active site effective volume may be a marker of substrate promiscuity in other enzymes, substrate spectra-effective volume relationships should be investigated in other protein families. In this line, Huang *et al.*[1] recently performed a systematic analysis of the substrate spectra of 200 phosphatases of the HAD superfamily, when tested against a set of 167 substrates. We collected the available crystal structures of each of the HAD phosphatases (Supporting Information Table S5) and computed the active site effective volume. We restricted the analysis to C2 cap members as they were reported to have a broader substrate spectrum,[1] and crystal structures with low to high effective volume are available. Interestingly, we observed that the effective volume (using the two conserved aspartic catalytic residues as the corrective SASA factor) was highly correlated (r^2 = 0.92) with the substrate range (Figure 6). Thus, the effective volume can be used as a molecular

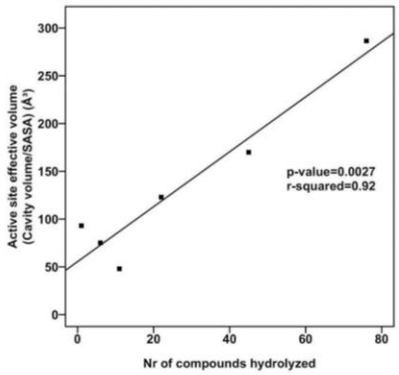


Figure 6. Relationships between the active site effective volume (in Å^3) and enzyme promiscuity (number of substrates hydrolyzed) of C2 members of HAD phosphatases. The number of substrates converted by each HAD phosphatase was obtained from Huang *et al.*[1] and is summarized in Supporting Information Table S5. The panels contain information for HAD phosphatases for which crystal structures were available and the catalytic residues identified. The active site effective volume (in Å^3) was calculated as described in Figure 5.

classification parameter of substrate promiscuity of phosphatases of the HAD superfamily when crystal structures are available. When this analysis was extended to the rest of the enzymes using homology modeling, we observed a similar trend to that of EHs (Supporting Information Figure S5). That is, no correlation existed (r^2 = 0.043), but still the effective volume can be used as a classifier of the substrate range as for EHs. Indeed, although a threshold could not be unambiguously established, sequences with the top 10 effective volumes belong to moderate-high to high promiscuity enzymes.

In conclusion, we found that the topology around the catalytic position, by means of an active site effective volume (cavity volume/SASA) threshold, is a dominant criterion of substrate promiscuity in EHs, which can be further extended by adding phylogenetic analysis. The rationale behind this parameter is as follows. Large volumes increase promiscuity until a certain value at which the cavity becomes too exposed and is not capable of properly accommodating and, importantly, retaining the substrate in specific catalytic binding interactions. This point is well captured by the SASA percentage of the catalytic triad, a dimensionless ratio that corrects for large volume measures in exposed sites. Importantly, the parameters of active site volume and relative SASA can be easily transferred to other systems. Indeed, the fact that the EHs investigated herein have different folds and that this parameter was also a marker of substrate spectra for phosphatases of the HAD superfamily opens the possibility of applying the effective volume measure to other enzymes requiring substrate anchoring. In all cases, the effective volume threshold-to-substrate relationships must be established. We would like to make note that the active site volume is not a static property, as the active site will breathe, depending on how flexible the protein is. In addition to that, the 62.5 Å^3 threshold for qualitatively classifying substrate promiscuity is based on the analysis of 147 EHs when tested against 96 esters. Although increasing the number of EHs and esters may influence this threshold and increase accuracy, it will not affect the fact that the measurement of the effective volume (cavity volume/SASA) can be used as the first molecular classification method of substrate promiscuity in EHs.

Our measurement is not a quantitative one, but rather a qualitative ranking (classification) procedure that will allow, for example, selecting sequences in databases for expression, particularly those encoding promiscuous enzymes capable of converting multiple substrates. This will substantially reduce reagent and labor costs compared to methods requiring the extensive cloning of all genes, and the expression and characterization of all enzymes in databases to later find those being promiscuous.[33] This possibility was herein examined by successfully mapping the open reading frames from the TARA Oceans project assemblies[34] and by identifying a high number of sequences encoding EHs with presumptive prominent substrate promiscuity (Supporting Information Results, Figures S6 and S7). Application of the effective volume measure to examine the sequences daily generated or deposited in databases requires having some crystals or X-ray structures for the model production. This limitation prevents predicting promiscuity from sequences lacking any structural information. Indeed, 36% of the EHs in this study (52 of the 147, including CalA/B) could not be included in the correlation because no calculation was possible. Accumulation of structural information and design and application of better modeling algorithms in the future will help solving this limitation.[35] Future studies might also explore molecular dynamics (MD) simulations to measure also the flexibility of the active site and not just the size of the cavity. By using this strategy, it was recently reported that the broad promiscuity of the members of the alkaline phosphatase superfamily arises from cooperative electrostatic interactions in the active site, allowing each enzyme to adapt to the electrostatic needs of different substrates.[36] In the particular case of EH phylogeny, a marker which does not require a three-dimensional structure was also suggested as a predictive classification marker of the substrate range. Indeed, this study suggests that in case of an unknown EH for which a crystal structure is not available or a homology model could not be established, then its assignation to family IV[14,23] increases the likelihood that this EH is promiscuous.

The present study not only provides clear evidence that substrate promiscuity in EHs has evolved from different core structural domains fitting an effective volume around the active

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

site, albeit with a bias toward that occurring in FIV members, but also from different phylogenetic lineages, many of which remain underexplored to date (Supporting Information Results and Figure S2). These are new findings as it was previously thought that the substrate range in a superfamily increased from a single ancestral core domain,[1] and because the identities of some microbial groups containing promiscuous enzymes, herein EHs, were previously unknown. Finally, this study also enabled the selection of a set of EH candidates that can compete with best commercial EHs such as CalB, as they show a broader substrate profile and specific activities up to 3-fold higher (Supporting Information Table S6). Their sequences can be used to search databases for similar promiscuous EHs. Further investigations should also determine the occurrence of other types of promiscuous EH phenotypes with broader substrate ranges than those identified in this study. For example, at least the stability of substrate-promiscuous EHs at different temperatures and with various solvents, along with the occurrence and evolution of secondary reactions, should be investigated in terms of condition and catalytic promiscuity.

METHODS

Protein Samples. Two main sources of EHs were used in the present study, all of them isolated *via* naïve and sequence-based screens in genomes and metagenomes. A first set of samples was EHs previously reported, as in the bibliography (69 in total), and that were herein substrate-profiled for first time. A second set was EHs (77) that are herein reported for first time. The extensive details of the source, cloning, expression, and purification of each of the active and soluble EHs are provided in the Supporting Information Methods and Table S1.

Ester Bond Hydrolysis Activity Assessment: Substrate Profiling Tests with 96 Esters. Hydrolytic activity was assayed at 550 nm using 96 structurally diverse esters in 384-well plates as previously described.[24,27,28] Before the assay, a concentrated stock solution of the esters was prepared at a concentration of 100 mg mL^{-1} in acetonitrile and dimethyl sulfoxide (DMSO). The assays were conducted according to the following steps. First, a 384-well plate (Molecular Devices, LLC, CA, USA) was filled with 20 μL of 5 mM *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid (EPPS) buffer, at pH 8.0, using a QFill3 microplate filler (Molecular Devices, LLC, CA, USA). Second, 2 μL of each ester stock solution was added to each well using a PRIMADIAG liquid-handling robot (EYOWN TECHNOLOGIES SL, Madrid, Spain). The ester was dispensed in replicates. After adding the esters, the 384-well plate was filled with 20 μL of 5 mM EPPS buffer, at pH 8.0, containing 0.912 mM Phenol Red (used as a pH indicator) using a QFill3 microplate filler. The final ester concentration of the ester in each well was 1.14 mg mL^{-1}, and the final concentration of Phenol Red was 0.45 mM. A total of 2 μL of protein extract (containing 1−5 mg mL^{-1} pure protein or 200 mg mL^{-1} wet cells expressing proteins) was immediately added to each well using an Eppendorf Repeater M4 pipet (Eppendorf, Hamburg, Germany) or a PRIMADIAG liquid-handling robot. Accordingly, the total reaction volume was 44 μL, with 4.5% (v/v) acetonitrile or DMSO in the reaction mixture. After incubation at 30 °C in a Synergy HT Multi-Mode Microplate Reader, ester hydrolysis was measured spectrophotometrically in continuous mode at 550 nm for a total time of 24 h. Commercially available CALA L and CALB L (Novozymes A/S, Bagsvaerd, Denmark) were diluted 10-fold with 5 mM EPPS buffer, at pH 8.0, and 2 μL of this solution was used immediately for reaction tests under the conditions described before. In all cases, specific activities (in U g^{-1} protein) were determined. One unit (U) of enzyme activity was defined as the amount of wet cells expressing EHs or pure EHs required to transform 1 μmol of substrate in 1 min under the assay conditions using the reported extinction coefficient ($\varepsilon_{\text{Phenol-red}}$ at 550 nm = 8450 M^{-1} cm^{-1}). All values were corrected for nonenzymatic transformation (i.e., the background rate) and for the background signal using *E. coli* cells that did not express any target protein (control cells included empty vectors). Note that a positive reaction was indicated by the restrictive criterion of a change greater than 6-fold above the background signal. Specific activity determinations (in U g^{-1}) for wet cells expressing each of the selected EHs or pure or commercial proteins are available in Supporting Information Tables S3 and S6, respectively.

Structural Determinations and Homology Modeling. The proteins EH1 and EH102 were expressed, purified, and crystallized using the sitting-drop method in Intelliplate 96-well plates and a Mosquito liquid-handling robot (TTP LabTech) according to previously described procedures.[37] For EHs for which crystal structures were not available, homology models were developed using Prime software from Schrödinger. Prime uses BLAST (with BLOSUM62 matrix) for homology search and alignment and refines the results using the Pfam database and pairwise alignment with ClustalW.

Protein Energy Landscape Exploration (PELE) Simulations. We used Protein Energy Landscape Exploration (PELE) software to sample the binding modes of glyceryl triacetate with EH1 and EH102.[38,39] The initial structures were taken from the coordinates of the EH1 and EH102 crystal structures (PDB codes: 5JD4 and 5JD3, respectively). The protonation state of titratable residues was estimated with the Protein Preparation Wizard (PROPKA)[40] and the H++ server (http://biophysics.cs.vt.edu/H++) followed by visible inspection. At pH 8 (the pH at which the activity assays were performed), the catalytic triad histidine residues were δ-protonated, and the catalytic triad aspartic acid residues were deprotonated, resulting in the formation of a histidine-serine and histidine-aspartic hydrogen-bonding network. The glyceryl triacetate structure was fully optimized with Jaguar[41] in an implicit solvent, and the electrostatic potential charges were computed with the density functional M06 at the 6-31G* level of theory. The ligand parameters were extracted from these for the classic simulations.

Cavity Volume and Solvent Accessible Surface Area (SASA) calculation. The relative Solvent Accessible Surface Area (SASA) for a residue was obtained using the GetArea Web server.[42] Cavity volumes were computed with Fpocket,[43] a very fast open-source protein pocket (cavity) detection algorithm based on Voronoi tessellation. Fpocket includes two other programs (dpocket and tpocket) that allow the extraction of pocket descriptors and the testing of owned scoring functions, respectively.

For extensive details of the methods, see Supporting Information Methods.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.7b00996.

Supporting Results, Methods, Figures S1−S7, and Table S4 (PDF)

Tables S1−S3, S5, and S6 (XLS)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: victor.guallar@bsc.es.

*E-mail: mferrer@icp.csic.es.

ORCID

Gerard Santiago: 0000-0002-0506-3049

Jürgen Pleiss: 0000-0003-1045-8202

Alexander F. Yakunin: 0000-0003-0813-6490

Víctor Guallar: 0000-0002-4580-1114

Manuel Ferrer: 0000-0003-4962-4714

Present Addresses

[γ]Current address: School of Chemistry, Bangor University, LL57 2UW Bangor, UK.

[δ]Current address: Lehrstuhl für Biotechnologie, RWTH Aachen University, Aachen, Germany.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

[c]Current address: Carl R. Woese Institute for Genomic Biology, Urbana, USA.

Author Contributions

[§]These authors contributed equally to this work.

Author Contributions

[ψ]These authors contributed equally in coordinating activities.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.C. thanks the Spanish Ministry of Economy, Industry and Competitiveness for a Ph.D. fellowship (Grant BES-2015-073829). V.M. thanks the Francisco José de Caldas Scholarship Program (Administrative Department of Science, Technology and Innovation, COLCIENCIAS). The authors acknowledge the members of the MAMBA, MAGICPAH, ULIXES, KILLSPILL and INMARE Consortia for their support in sample collection. David Rojo is also acknowledged for his valuable help with log P calculations. This project received funding from the European Union's Horizon 2020 research and innovation program [Blue Growth: Unlocking the potential of Seas and Oceans] under grant agreement no. 634486 (project acronym INMARE). This research was also supported by the European Community Projects MAGICPAH (FP7-KBBE-2009-245226), ULIXES (FP7-KBBE-2010-266473), and KILLSPILL (FP7-KBBE-2012-312139) and grants BIO2011-25012, PCIN-2014-107, BIO2014-54494-R, and CTQ2016-79138-R from the Spanish Ministry of Economy, Industry and Competitiveness. The present investigation was also funded by the Spanish Ministry of Economy, Industry and Competitiveness within the ERA NET IB2, grant no. ERA-IB-14-030 (MetaCat), the UK Biotechnology and Biological Sciences Research Council (BBSRC), grant no. BB/M029085/1, and the German Research Foundation (FOR1296). R.B. and P.N.G. acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) *via* the Welsh Government. O.V.G. and P.N.G. acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. A.Y. and A.S. gratefully acknowledge funding from Genome Canada (2009-OGI-ABC-1405) and the NSERC Strategic Network grant IBN. A.I.P. was supported by the Counseling of Economy and Employment of the Principality of Asturias, Spain (Grant FC-15-GRUPIN14-107). V.G. acknowledges the joint BSC-CRG-IRB Research Program in Computational Biology. The authors gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF).

■ REFERENCES

(1) Huang, H., Pandya, C., Liu, C., Al-Obaidi, N. F., Wang, M., Zheng, L., Toews Keating, S., Aono, M., Love, J. D., Evans, B., Seidel, R. D., Hillerich, B. S., Garforth, S. J., Almo, S. C., Mariano, P. S., Dunaway-Mariano, D., Allen, K. N., and Farelli, J. D. (2015) Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. U. S. A. 112*, E1974−1983.

(2) Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B., and Barkman, T. J. (2012) Enzyme functional evolution through improved catalysis of ancestrally non-preferred substrates. *Proc. Natl. Acad. Sci. U. S. A. 109*, 2966−2971.

(3) Yip, S. H., and Matsumura, I. (2013) Substrate ambiguous enzymes within the *Escherichia coli* proteome offer different evolutionary solutions to the same problem. *Mol. Biol. Evol. 30*, 2001−2012.

(4) Price, D. R., and Wilson, A. C. (2014) Substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol. 12*, 110.

(5) Devamani, T., Rauwerdink, A. M., Lunzer, M., Jones, B. J., Mooney, J. L., Tan, M. A., Zhang, Z. J., Xu, J. H., Dean, A. M., and Kazlauskas, R. J. (2016) Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *J. Am. Chem. Soc. 138*, 1046−1056.

(6) Hult, K., and Berglund, P. (2007) Enzyme promiscuity: mechanism and applications. *Trends Biotechnol. 25*, 231−238.

(7) Copley, S. D. (2015) An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci. 40*, 72−78.

(8) London, N., Farelli, J. D., Brown, S. D., Liu, C., Huang, H., Korczynska, M., Al-Obaidi, N. F., Babbitt, P. C., Almo, S. C., Allen, K. N., and Shoichet, B. K. (2015) Covalent docking predicts substrates for haloalkanoate dehalogenase superfamily phosphatases. *Biochemistry 54*, 528−537.

(9) Nobeli, I., Favia, A. D., and Thornton, J. M. (2009) Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol. 27*, 157−167.

(10) Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol. 9*, 22−34.

(11) Braakman, R., and Smith, E. (2014) Metabolic evolution of a deep-branching hyperthermophilic chemoautotrophic bacterium. *PLoS One 9*, e87950.

(12) Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014) Implications of streamlining theory for microbial ecology. *ISME J. 8*, 1553−1565.

(13) Lan, T., Wang, X. R., and Zeng, Q. Y. (2013) Structural and functional evolution of positively selected sites in pine glutathione S-transferase enzyme family. *J. Biol. Chem. 288*, 24441−24451.

(14) Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., and Golyshin, P. N. (2015) Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransform. 33*, 235−249.

(15) Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature 409*, 258−268.

(16) Koudelakova, T., Chovancova, E., Brezovsky, J., Monincova, M., Fortova, A., Jarkovsky, J., and Damborsky, J. (2011) Substrate specificity of haloalkane dehalogenases. *Biochem. J. 435*, 345−354.

(17) Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A., and Sanchez-Ruiz, J. M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J. Am. Chem. Soc. 135*, 2899−2902.

(18) Amin, S. R., Erdin, S., Ward, R. M., Lua, R. C., and Lichtarge, O. (2013) Prediction and experimental validation of enzyme substrate specificity in protein structures. *Proc. Natl. Acad. Sci. U. S. A. 110*, E4195−4202.

(19) Anton, B. P., Chang, Y. C., Brown, P., Choi, H. P., Faller, L. L., Guleria, J., Hu, Z., Klitgord, N., Levy-Moonshine, A., Maksad, A., Mazumdar, V., McGettrick, M., Osmani, L., Pokrzywa, R., Rachlin, J., Swaminathan, R., Allen, B., Housman, G., Monahan, C., Rochussen, K., Tao, K., Bhagwat, A. S., Brenner, S. E., Columbus, L., de Crécy-Lagard, V., Ferguson, D., Fomenkov, A., Gadda, G., Morgan, R. D., Osterman, A. L., Rodionov, D. A., Rodionova, I. A., Rudd, K. E., Söll, D., Spain, J., Xu, S. Y., Bateman, A., Blumenthal, R. M., Bollinger, J. M., Chang, W. S., Ferrer, M., Friedberg, I., Galperin, M. Y., et al. (2013) The COMBREX project: design, methodology, and initial results. *PLoS Biol. 11*, e1001638.

(20) Colin, P. Y., Kintses, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., Hyvönen, M., Morgavi, D. P., Janssen, D. B., and Hollfelder, F. (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun. 6*, 10008.

(21) Chen, C., Huang, H., and Wu, C. H. (2017) Protein bioinformatics databases and resources. *Methods Mol. Biol. 1558*, 3−39.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

(22) Fischer, M., and Pleiss, J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res. 31*, 319−321.
(23) Arpigny, J. L., and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J. 343*, 177−183.
(24) Alcaide, M., Tornés, J., Stogios, P. J., Xu, X., Gertler, C., Di Leo, R., Bargiela, R., Lafraya, A., Guazzaroni, M. E., López-Cortés, N., Chernikova, T. N., Golyshina, O. V., Nechitaylo, T. Y., Plumeier, I., Pieper, D. H., Yakimov, M. M., Savchenko, A., Golyshin, P. N., and Ferrer, M. (2013) Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochem. J. 454*, 157−166.
(25) Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., and Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J. 432*, 437−444.
(26) Popovic, A., Hai, T., Tchigvintsev, A., Hajighasemi, M., Nocek, B., Khusnutdinova, A. N., Brown, G., Glinos, J., Flick, R., Skarina, T., Chernikova, T. N., Yim, V., Brüls, T., Paslier, D. L., Yakimov, M. M., Joachimiak, A., Ferrer, M., Golyshina, O. V., Savchenko, A., Golyshin, P. N., and Yakunin, A. F. (2017) Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. *Sci. Rep. 7*, 44103.
(27) Janes, L. E., Löwendahl, C., and Kazlauskas, R. J. (1998) Rapid quantitative screening of hydrolases using pH indicators. Finding enantioselective hydrolases. *Chem. Eur. J. 4*, 2317−2324.
(28) Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Rico Eguizabal, E., Guerrero, Mdel C., Yakunin, A. F., and Ferrer, M. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol. 79*, 3553−3562.
(29) Daiha, K. d. G., Angeli, R., de Oliveira, S. D., and Almeida, R. V. (2015) Are lipases still important biocatalysts? A study of scientific publications and patents for technological forecasting. *PLoS One 10*, e0131624.
(30) Borrelli, K. W., Cossins, B., and Guallar, V. (2009) Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem. 31*, 1224−1235.
(31) Hernández-Ortega, A., Borrelli, K., Ferreira, P., Medina, M., Martínez, A. T., and Guallar, V. (2011) Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J. 436*, 341−350.
(32) Santiago, G., de Salas, F., Lucas, F., Monza, E., Acebes, S., Martínez, A., Camarero, S., and Guallar, V. (2016) Computer-aided laccase engineering: toward biological oxidation of arylamines. *ACS Catal. 6*, 5415−5423.
(33) Barak, Y., Nov, Y., Ackerley, D. F., and Matin, A. (2008) Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J. 2*, 171−179.
(34) Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B. T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M. B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., Bork, P., et al. (2015) Structure and function of the global ocean microbiome. *Science 348*, 1261359.
(35) Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. (2016) Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins: Struct., Funct., Genet. 84* (Suppl 1), 4−14.
(36) Barrozo, A., Duarte, F., Bauer, P., Carvalho, A. T. P., and Kamerlin, S. C. L. (2015) Cooperative electrostatic interactions drive functional evolution in the alkaline phosphatase superfamily. *J. Am. Chem. Soc. 137*, 9061−9076.
(37) Alcaide, M., Stogios, P. J., Lafraya, Á., Tchigvintsev, A., Flick, R., Bargiela, R., Chernikova, T. N., Reva, O. N., Hai, T., Leggewie, C. C., Katzke, N., La Cono, V., Matesanz, R., Jebbar, M., Jaeger, K. E., Yakimov, M. M., Yakunin, A. F., Golyshin, P. N., Golyshina, O. V., Savchenko, A., and Ferrer, M. (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol. 17*, 332−345.
(38) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B 105*, 6474−6487.
(39) Borrelli, K. W., Vitalis, A., Alcantara, R., and Guallar, V. (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput. 1*, 1304−1311.
(40) Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des. 27*, 221−234.
(41) Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J., and Friesner, R. A. (2013) Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem. 113*, 2110−2142.
(42) Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem. 19*, 319.
(43) Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinf. 10*, 168.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

SUPPORTING INFORMATION

Determinants and prediction of esterase substrate promiscuity patterns

Mónica Martínez-Martínez, Cristina Coscolín, Gerard Santiago, Jennifer Chow, Peter J. Stogios, Rafael Bargiela, Christoph Gertler, José Navarro-Fernández, Alexander Bollinger, Stephan Thies, Celia Méndez-García, Ana Popovic, Greg Brown, Tatyana N. Chernikova, Antonio García-Moyano, Gro E.K. Bjerga, Pablo Pérez-García, Tran Hai, Mercedes V. Del Pozo, Runar Stokke, Ida H. Steen, Hong Cui, Xiaohui Xu, Boguslaw P. Nocek, María Alcaide, Marco Distaso, Victoria Mesa, Ana I. Peláez, Jesús Sánchez, Patrick C. F. Buchholz, Jürgen Pleiss, Antonio Fernández-Guerra, Frank O. Glöckner, Olga V. Golyshina, Michail M. Yakimov, Alexei Savchenko, Karl-Erich Jaeger, Alexander F. Yakunin, Wolfgang R. Streit, Peter N. Golyshin, Víctor Guallar, Manuel Ferrer. The INMARE Consortium

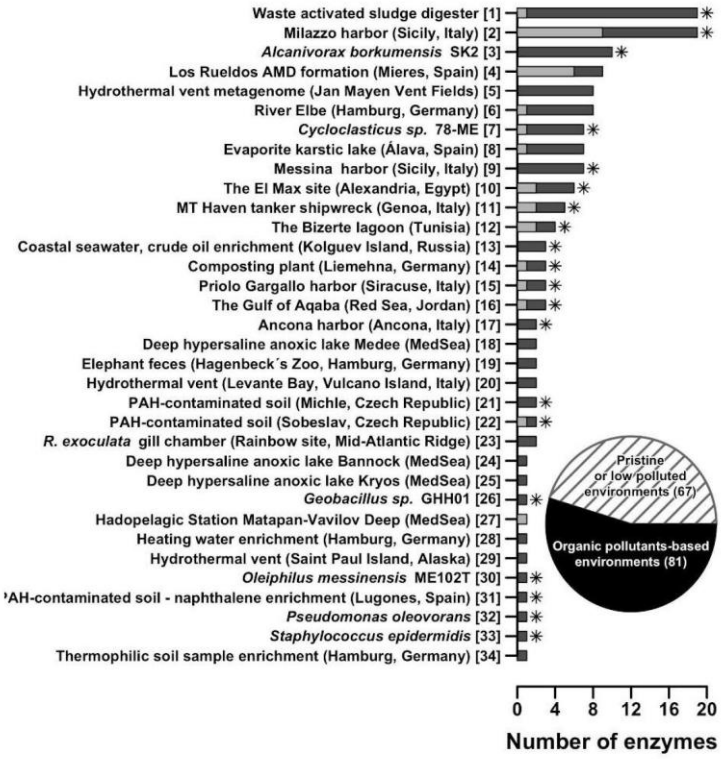


Figure S1. Bio-resources (habitats or microbes) of all 145 EHs in this study. This figure is generated from data in Supporting Information, Tables S1. The bars summarized the number of EHs characterized per each of the bio-resources screened. The number of highly promiscuous hits (i.e., using 30 or more esters) is shown by gray color in the bars. Two major bio-sources of enzymes were covered, namely organically polluted environments or bacteria inhabiting them, and pristine or low-pollution environments. The number of each of each of these bio-sources is summarized in the circular inset. The bio-sources defined as polluted are specifically marked with an asterisk in the bars. Each of the bio-resources is assigned an arbitrary number in square brackets, which is further used as ID in Fig. 2.

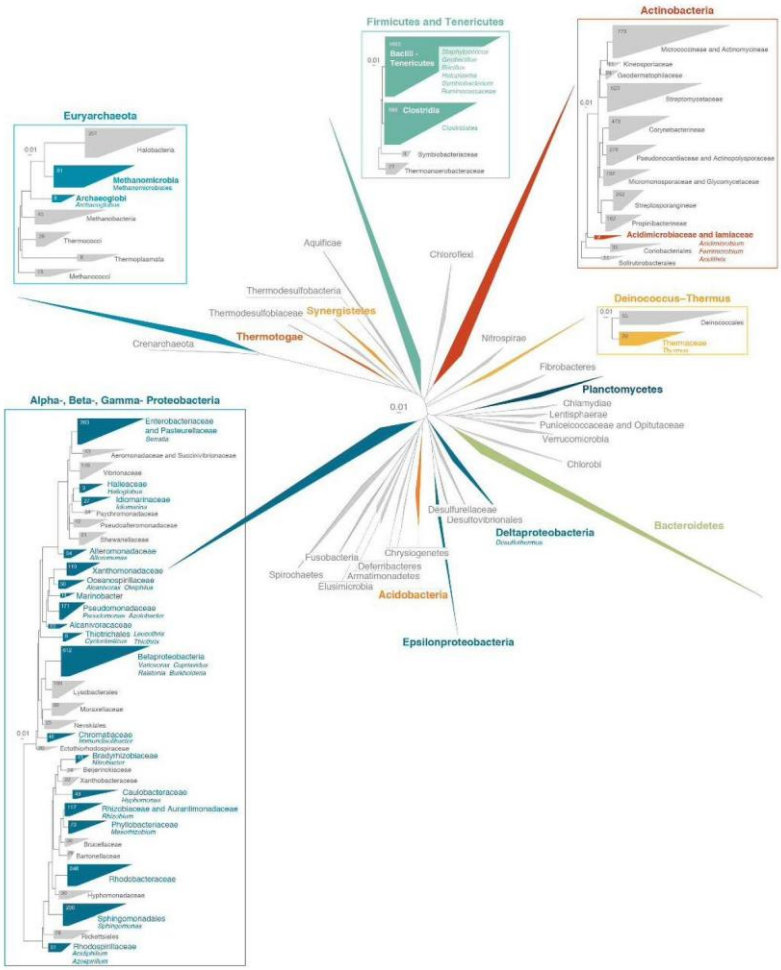


Figure S2. The associations of EHs to cultured and uncultured bacteria from multiple phylogenetic lineages. This figure is generated from data in Supporting Information, Tables S1. Maximum likelihood phylogenetic analysis of the 16S rRNA genes of all species present in the most recent

release of the Living Tree Project database. The groups in which enzymes have been characterized are presented with a color code. Microbial clades not covered in this study are colored in gray. The subtrees represent in detail the groups that contain EHs with broad substrate spectra (i.e., using 30 esters or more) described at the family and genus levels. The numbers in each collapsed clade from the subtrees designate the number of 16S rRNA gene sequences included in each group. Only major taxonomic groups with more than 5 type species present in the database were included in the phylogeny. The tree scales represent the nucleotide substitutions per site. Note that the phylogenetic binning of sequences encoding EHs (for details see Supporting Information, Table S1) was performed using a genome linguistics approach. Briefly, metagenomic fragments were searched for oligonucleotide compositional similarity (frequencies of tetranucleotides) against all sequenced bacterial chromosomes, plasmids and phages using the GOHTAM web tool.[1] For short DNA fragments, compositional analysis could not be performed, and a comprehensive analysis of the TBLASTX results was conducted. Both methods have been proven successful for suggesting the origin of metagenomics sequences.[1,2] However, we are aware that deep assignations (including at the species level) cannot be obtained for short DNA fragments and that this tool may not be appropriate for phylogenetic analysis of sequences with no homology in databases.

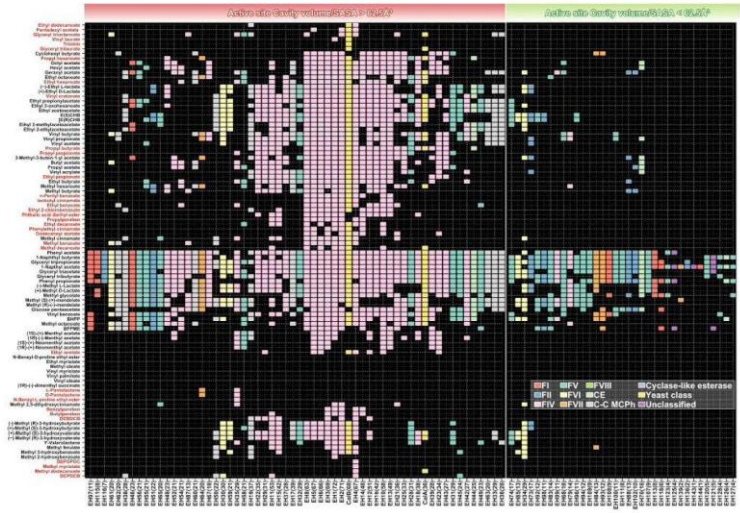


Figure S3. Active site effective volume and phylogeny may be also indicative of specific conversions. This figure is generated from data in Supporting Information, Tables S1 and S3. The figure is constructed as described in Fig. 2, although distance matrixes are not shown. Note that in Fig. 2 only EHs capable of converting 10 or more esters are given whatever their active site effective volume. In this case, the figure summarizes the substrate spectra of EHs from which active site effective volume could be unambiguously identified (96 in total) independently of the number of esters they can hydrolyze. The figure summarizes the substrate spectra of EHs with effective volumes ≥62.5 $Å^3$ (on the left side) or < 62.5 $Å^3$ (on the right side). The 34 esters found to be hydrolyzed under our assay conditions by EHs with effective volumes ≥62.5 $Å^3$ are indicated in the left side with a red color. Note that in all cases the threshold of 62.5 $Å^3$ was defined when the active site cavity volume was computed in $Å^3$ and SASA as a dimensionless ratio from 0 to 100 using the GetArea server (http://curie.utmb.edu/getarea.html).

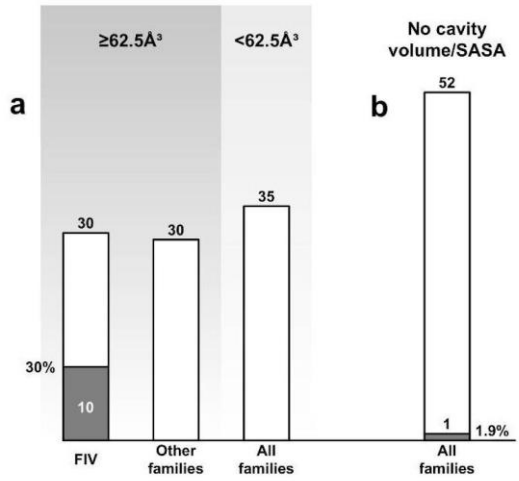


Figure S4. The active site effective volume and phylogeny are predictive markers of the capacity to convert benzyl-, butyl- and propyl-paraben esters. This figure is generated from data in Supporting Information, Tables S3 and S6. (a) Numbers (shown on the top of the bars) of EHs, assigned to FIV or other families, conforming to the ≥62.5 $Å^3$-threshold, or to all families and <62.5 $Å^3$-threshold. (b) Numbers of EHs (shown on the top of the bars) for which effective volume could not be unambiguously measured because sequence identities of less than 25% (to an existing crystal) or because not suitable alignments. In all cases, the number and percentage of EHs capable of converting paraben esters is shown by gray color in the bars. The threshold of 62.5 $Å^3$ was defined when the active site cavity volume was computed in $Å^3$ and SASA as a dimensionless ratio from 0 to 100 using the GetArea server (http://curie.utmb.edu/getarea.html).

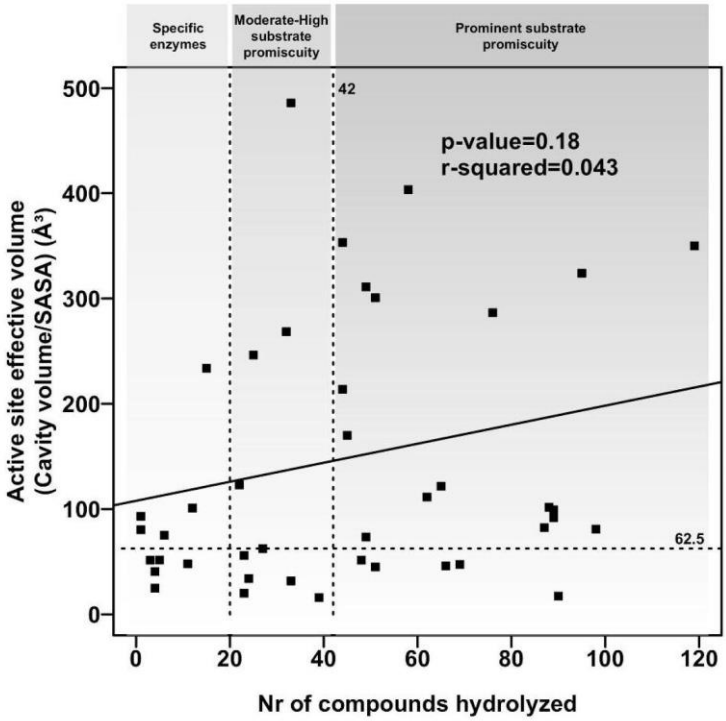


Figure S5. Relationships between the active site effective volume (in $Å^3$) and enzyme promiscuity (number of substrates hydrolyzed) of C2 members of HAD phosphatases. The number of substrates converted by each HAD phosphatase was obtained from Huang *et al.*[6] and is summarized in Supporting Information, Table S5. The panels contain information for HAD phosphatases for both crystal structures and homology models were available or could be unambiguously established (sequence identity ≥25%) and the catalytic residues identified. The threshold of substrates being converted, exemplifying the level of promiscuity as for EHs, is indicated. The active site effective volume (in $Å^3$) was calculated as described in Fig. 5.

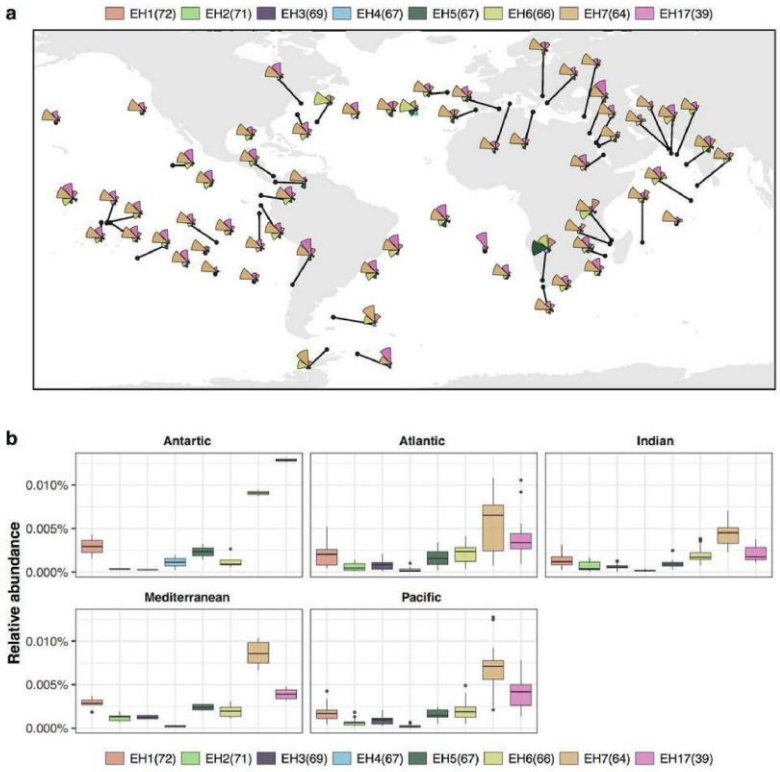


Figure S6. Exploring the sequence space for sequences that encode the most promiscuous EHs. The number in brackets close to the enzyme ID indicates the number of esters hydrolyzed by each enzyme, as described in the legend of Fig. 2. (a) Geographic distribution of the 8 selected promiscuous EHs in the 242 TARA Oceans samples. (b) Relative abundance of the 8 promiscuous EHs in five regions.

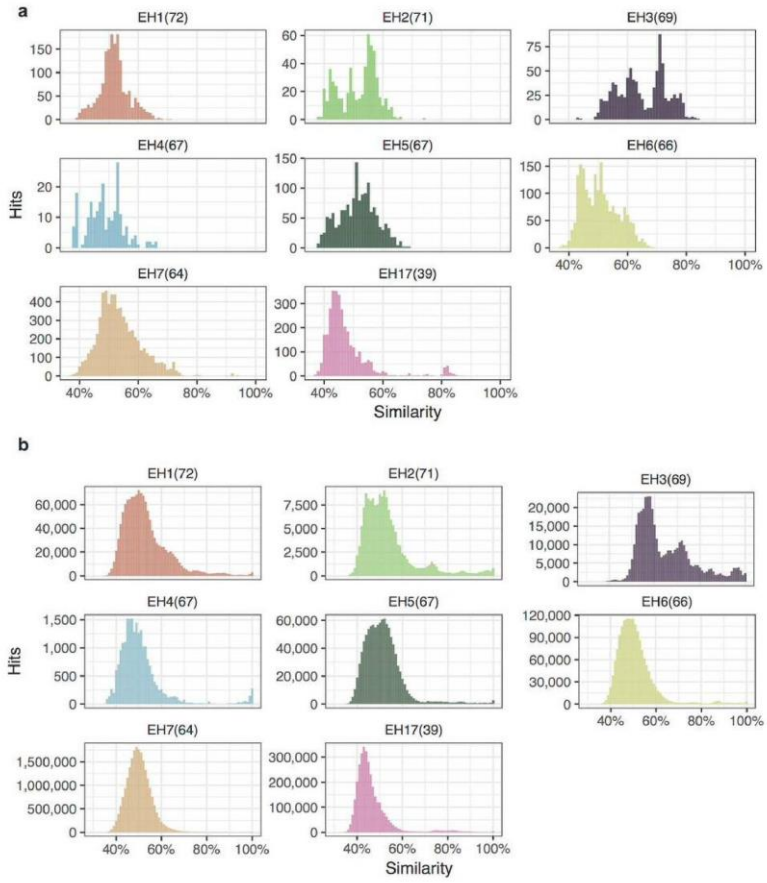


Figure S7. Distribution of the similarity values. The number in the bracket close to the enzyme ID indicates the number of esters hydrolyzed by each enzyme, as described in the legend of Fig. 2. (a) Blastp similarity distribution for every hit against each of the 8 selected promiscuous EHs. (b) All-vs-all blastp similarity distributions for all EH homologs found in the TARA Oceans samples.

Table S1. General information about all 145 EHs and the 2 commercial preparations investigated in this study. The database is an Excel table that provides the following information: ID code used in this study, accession number, PDB code if crystals were available, family to which each sequence was assigned, number of esters hydrolyzed (as found in Supporting Information, Table S3), active site cavity volume/SASA, average maximum specific activity (in U g^{-1} wet cells) (see Supporting Information, Table S3), sequence information, screening technique (sequence or naïve screen), enzyme source (with a full description and short description as in Figure 1), habitat type (as described in Figure 1 inset), cloning vector and expression host, expression conditions (i.e., antibiotic, inductor and temperature), a reference describing the identification, cloning and expression, the top hit in NCBI and the sequence identity, the theoretical molecular weight (ranging from 22 to 103 kDa) and the isoelectric point (ranging from 3.81 to 11.02). *Note 1*: To unambiguously assign sequences to families, a phylogenetic tree was generated using the FastTree v2.1.7 algorithm[3] implemented with the Shimodaira-Hasegawa test. Reference sequences that were unambiguously assigned to each of the esterase/lipase families[4,5] were used to help classify each of the 145 EHs. The final sequence alignments and the tree are available from the authors upon request. *Note 2*: The source organisms of selected polypeptides were identified by a search of oligonucleotide patterns against the GOHTAM database and TBLASTX to reveal compositional similarities between the gene sequences and/or DNA fragments containing the genes that encoded EHs and all analyzed sequenced bacterial chromosomes, plasmids and phages.[1,2] Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S2. Pairwise sequence similarities for all 147 EHs (including CalA/B) as calculated using Needleman-Wunsch alignments performed against all the other candidates ("all-vs.-all"). Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S3. The specific activity is given as U g^{-1} wet cells expressing esterases tested against a set of 96 structurally different esters. The table is an Excel table in which the following information is provided: name of ester, compound chemical class, log P value with standard deviation (calculated using the ACD/ChemSketch 2015.2.5 software), SMILES code, source or brand (with link to the source), molecular mass (g/mol), reaction conditions (including assay method, substrate and pH indicator concentration, buffer, shaking, temperature [30 °C] and pH [8.0], reaction time, additives and concentrations, cell amount), and average specific activity values for all 145 EH preparations in whole-cell assays given in U g^{-1} wet cell pellet. For the Cal A and Cal B preparations, the specific activity values are given in U g^{-1} total protein. The assays were performed as replicates, with the

average value given, and the standard deviation was less than 1% in all cases. Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S4. X-ray diffraction data collection and refinement statistics.

Protein	EH1	EH102
PDB code	5JD4	5DJ3
Data collection		
Space group	P 1	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	90.28, 90.10, 110.76	66.59, 129.45,
α, β, γ (°)	68.02, 79.60, 67.57	129.85
		90, 97.49, 90
Resolution, Å	25.00 – 2.05	29.73 – 2.30
R_{merge} [a]	0.072 (0.606)[b]	0.125 (0.543)
$I / \sigma(I)$	21.66 (2.52)	9.86 (2.50)
Completeness, %	94.9 (91.2)	96.9 (93.4)
Redundancy	3.0 (2.9)	3.6 (3.6)
Refinement		
Resolution, Å	24.95 – 2.05	29.73 – 2.30
No. of unique reflections:	16392, 1875	86908, 4375
working, test		
R-factor/free *R*-factor[c]	16.9/21.1 (27.9/30.0)	18.4/23.6
		(22.4/28.1)
No. of refined atoms,		
molecules	19 262, 8	13 721, 8
Protein	324	23
Solvent	2463	1572
Water		
B-factors		
Protein	37.1	37.1
Solvent	63.1	60.5

Water	48.8	47.0
r.m.s.d.		
Bond lengths, Å	0.003	0.004
Bond angles, °	0.730	0.684

[a] $R_{sym} = \Sigma_h\Sigma_i|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)$, in which $I_i(h)$ and $\langle I(h)\rangle$ are the ith and mean measurement of the intensity of reflection h.

[b] The values for the outer shells of the data are indicated in parentheses.

[c] $R = \Sigma|F_p^{obs} - F_p^{calc}|/\Sigma F_p^{obs}$, in which F_p^{obs} and F_p^{calc} are the observed and calculated structure factor amplitudes, respectively.

* = molecules in the active site cleft.

Table S5. General information about all HAD phosphatases analyzed in this study. The database is an Excel table that provides the following information: ID code, Uniprot ID and number of substrates being converted as reported,[6] and active site cavity volume/SASA. In case of protein for which a crystal structure was not available, the homology to the PDB template (in %) is given. Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S6. The specific activity is given as U g^{-1} pure protein tested against a set of 96 structurally different esters. The data for the Cal A and Cal B commercial preparations are also included. The table is an Excel table in which the following information is provided: name of ester, compound chemical class, log P value with standard deviation (calculated using the ACD/ChemSketch 2015.2.5 software), SMILES code, source or brand (with link to the source), molecular mass (g/mol), reaction conditions (including assay method, substrate and pH indicator concentration, buffer, shaking, temperature [30°C] and pH [8.0], reaction time, additives and concentrations, enzyme amount), and average specific activity values for all tested EHs in pure protein assays given in U g^{-1} protein. The assays were performed as replicates, with the average value given, and the standard deviation was less than 0.5% in all cases. It should be noted that all preparations were tested in pure form as His-tagged proteins, but for the sake of simplicity, the specific activities of only the most promiscuous EHs (i.e., those that reacted with 42 or more esters) are shown. All datasets are available upon request from the authors. Due to the extensive size, the table is submitted as a separate file in Excel format.

RESULTS

Source of EH. A search of oligonucleotide patterns against the GOHTAM database[1,2] and TBLASTX analysis indicated that all 145 EHs were distributed across the entire phylogenetic tree (i.e., in at least 10 phyla and 40 genera), although a bias (70%) toward proteobacterial EH was noted (Supporting Information, Table S1, Fig. S2). No clear affiliation other than *Bacteria* was observed for 9 EHs, and an additional set of 43 EHs could not be assigned at the genus level. Within the 31 highly promiscuous EHs (i.e., using 30 or more esters; excluding CalA/B), an unambiguous affiliation was found for 21 EHs that were associated with at least 2 phyla (the most abundant was Proteobacteria (94%)) and 15 genera. These genera included *Sphingomonas*, *Rhizobium*, *Pseudomonas*, *Alteromonas* and *Acidiphilium*, which have been well-explored with respect to their enzymatic contents, as well as genera that have been largely neglected, such as *Acidithrix*, *Acidimicrobium*/*Ferrimicrobium*, *Alcanivorax*, *Cycloclasticus*, *Immundisolibacter*, *Idiomarina*, *Hyphomonas,* and *Halioglobus* (Supporting Information, Fig. S2). Of the known 24 bacterial and 5 archaeal phyla with cultured representatives,[7] the present study covered EHs from 10 and 1 phyla, respectively (Supporting Information, Fig. S2); much diversity remains to be uncovered in the coming years. The reason that enzyme diversity is so biased is intriguing, even though a very broad diversity was sampled in this study. Note that because many microbial lineages others than those herein analyzed may contain promiscuous EHs a statistical significance of the enrichment of substrate-promiscuous EHs in the above bacterial species cannot be assessed, which is out of the scope of the present study. Whatever the case, we have provided first preliminary evidences that a number of underexplored microbial phylogenetic lineages contained EHs with prominent substrate range. Similarly, we noticed a high percentage of EHs with prominent substrate promiscuity in the chronically polluted seashore area of Milazzo harbor in Sicily (Italy) compared to other sites (Supporting Information, Fig. S1), but again the statistical significance cannot be obtained because the number of enzymes examined was over-represented in this site and because habitats others than those herein examined remains to be explored.

Substrate profiling: general considerations and dynamic range of the assay. Identifying EHs with a broad substrate profile remains a major bottleneck for biocatalysis and biotechnological processes in general.[5] Rapid screening methods will facilitate the identification of such enzymes. Medium-to-high throughput colorimetric protocols have been previously developed that enable EH hydrolytic activity testing with pure proteins.[1,8-11] However, protein purification on a large scale is a very time-consuming and expensive process.[12,13] Consequently, the methods in the present study were adapted to use whole cells, which allowed broad sampling and reduced the time, effort and

cost of identifying enzymes with ample or restricted substrate spectra. Prior to substrate profiling, all clones were tested for activity on agar plates using the three model esters 1-naphthyl acetate, glyceryl tri-acetate and tri-propionate, and their activity was confirmed (see Supporting Information, Methods).

The activity protocol established and used in this study included growing *E. coli* cultures expressing the enzyme of interest overnight in the presence of the appropriate antibiotics and an inducer, followed by activity analysis of the pellet fractions against 96 esters via a pH indicator assay in 384-well plates, at pH 8.0 and 30 °C. Acid is produced after ester bond cleavage by the action of the hydrolytic enzyme contained in the cell pellet,[8] which induces a color change of the pH indicator that can be measured spectrophotometrically at 550 nm. Following the recommendation of Janes *et al.*[8] the concentrations of the pH indicator (0.45 mM) and each of the esters (from 1.3 to 13.2 mM; see concentrations in Methods) were chosen to maximize the accuracy and sensitivity. The number of cells should be as high as possible to maximize sensitivity and to ensure accurate "activity data." In our assay, activity data refers to experimental "time course data" of substrate conversion using 0.4 mg of wet cells per ester, in which time course data might indicate specific activity (in U per g wet cells) from time-course conversion after a maximum of 24 h. This amount of cells (0.4 mg per ester) and time frame (up to 24 h) were found to ensure sensitivity and detection of all esters hydrolyzed by any given EH, as determined by studying the frequency of each ester considered as a hit using different numbers of cells and different time frames and comparing the data from whole-cell assays with pure proteins. Thus, the substrate utilization for all pure proteins supported the validity of the whole-cell screening assay in this study. Under our assay conditions, all esters hydrolyzed by whole cells (Supporting Information, Table S3) were also hydrolyzed when using pure proteins and *vice versa* (Supporting Information, Table S6), which demonstrated that substrate limitation problems using whole cells were not anticipated and that the rapid assay used in this study can be applied to screen substrate ambiguity for any type of cells containing EHs (i.e., fosmid clones, expression clones, microbial cells, etc.). However, we cannot rule out the possibility that unambiguous detection of the conversion of a substrate at a significantly low rate may not be possible using whole cells.

The rapid whole-cell assay used in this study to analyze the substrate profile may be of interest in future genomic and metagenomics screening programs to help identify and prioritize clones with a broad substrate range, whether they contain one or more sequences that encode EHs in a cloned DNA fragment. To screen for promiscuous enzymes, cells from pure cultures, enrichments or clones that express a particular DNA fragment can be screened with standard substrates. Those found to be active against a common standard ester (e.g., 1-napthyl-acetate or glyceryl tri-butyrate)

but do not show a restricted promiscuous profile in a subsequent substrate profiling test will not merit *a priori* sequencing and/or cloning efforts if the objective of the screening program is to find substrate-promiscuous enzymes. By contrast, clones active against a broad range of esters at a high rate merit further sequencing, cloning, expression and characterization. This will substantially reduce the reagent and labor cost while using modest resources. If the activity tests, which can be extended to a broader set of chemical blocks of industrial interest, are also directly performed under conditions resembling industrial requirements, for example in the presence of solvents or at high temperature, efforts can be made toward identifying new, versatile and robust biocatalysts fulfilling industry criteria. Indeed, preliminary tests indicated that the assay method used in this study can be performed in the presence of solvents such as methanol, ethanol, acetonitrile and DMSO at final concentrations of up to 30% (*v*/*v*).

Assays were performed at pH 8.0 and 30 °C in the absence of any chemical other than the ester and a small amount (4.5% v/v) of solvent (DMSO or acetonitrile) needed to dissolve the ester. Therefore, no surfactant was added. Although this may compromise the solubility of very hydrophobic molecules, the majority of large hydrophobic molecules, such as triolein, 2,4-dichlorophenyl 2,4-dichlorobenzoate, 2,4-dichlorobenzyl 2,4-dichlorobenzoate, and diethyl-2,6-dimethyl 4-phenyl-1,4-dihydro pyridine-3,5-dicarboxylate, were hydrolyzed by some EHs under our assay conditions, suggesting that our assay conditions did not introduce any bias in the detection of substrate-enzyme pairs.

Reactions were performed at pH 8.0 and 30 °C. Because we sampled a very broad diversity of habitats from moderately cold to thermophilic, we decided to use an assay temperature of 30 °C, the temperature at which we anticipated most EHs would show appreciable activity. A pH of 8.0 was selected because it was the pH required for the Phenol Red pH indicator used, although pH 7.0 with *p*-nitrophenol as the pH indicator may also be used.[8] We used pH 8.0 because most EHs have been reported to show neutral-to-slightly alkaline pH optima. We are aware that we also included samples from environments ranging from sea-water like habitats to acid mine drainage and that the enzymes from these environments may show different pH optima. Consequently, in many cases (not shown), pH and temperature values for optimal activity were compiled (from previous studies or after the evaluation of optimal parameters for the enzymes investigated herein for the first time) to ensure that the EHs were most active at 30 °C and pH 8.0. Having said that the assay herein used can be extended to pH 7.0 (using *p*-nitrophenol as pH indicator) and at any temperature, preferably below 70°C to avoid evaporation; a minimum concentration of 0.5 mM substrate is recommended to ensure detection of pH shift.[8] The concentration of buffer (5 mM EPPS) and pH indicator

(Phenol red 0.45 mM) were empirically determined as being optimal to ensure that changes in pH shift during reaction gives linear changes in absorbance.[8]

The predictive capacity of cavity volume/SASA is not influenced by the presence of flexible elements in the structure. The predictive capacity of the active site effective volume may also sensitively depend on the conformational state of flexible elements such as the lid.[5] Indeed, some EH proteins contain a lid domain that covers the active site, preventing substrate binding and requiring structural rearrangement to attain an open conformation. To evaluate whether the homology model-derived active site effective volume value of an EH differs if the template is in the open or closed state, we used the crystal structures of CalA and CalB in closed and open conformations (3GUU/2VEO and 4K6G/5A71). The variations were minimal. Since CalA and CalB do not have the classic lid, which consists of one or two helices that move considerably as rigid bodies, we extended the applicability of this measure to sequences of typical-lid fungal-like lipases such as lipases from *Candida rugosa* (open/close PDB: 1CRL/1THR) and *Thermomyces lanuginosa* (1DTE/1DT5). In all cases, the effective volume ($\geq$76.9 Å^3) was above the 62.5 Å^3-threshold, which is indicative of high substrate promiscuity, with minimal variations between the open and closed conformations. Therefore, the presence of lid domains does not have an observable effect on the substrate promiscuity level, which is mostly defined by the sequence-defined topology of the active site environment, herein exemplified by the active site effective volume.

An example of the potential of the effective volume measure for bioprospecting of EH with prominent substrate promiscuity. This study also produced a subset of EHs with large substrate ranges, competitive with the best industrial prototypes. Their sequences can be used as targets for bioprospecting similar sequences in large-scale datasets. As example, we used MMSEQS2[14] to screen the occurrence of promiscuous EHs against the predicted open reading frames from the TARA Oceans project assemblies,[15] which was used as a study case. All hits with an E-value threshold of 10^{-10} and sequence coverage $\geq$0.6 were selected (see Supporting Information, Methods). For simplicity, the search was restricted to sequences encoding EH1 to EH6 (assigned to FIV), EH7 (assigned to FVIII), and EH17 (assigned to MCP hydrolase family). All of these EHs were among the most substrate-promiscuous EHs and encompassed phylogenetically and environmentally diverse sequences (see Clusters 1 and 2 in Fig. 2). The data presented in Supporting Information, Fig. S6a reveal that not all sequences encoding promiscuous EHs were equally abundant. The homologues to FVIII serine beta-lactamase EH7 and MCP hydrolase EH17 were the most abundant in the majority of the sites, which is most noticeably when examining their accumulative abundances in all 5 oceanic regions (Supporting Information, Fig. S6b). At this stage we would like to notice that obtaining statistically significant differences to environmental

metadata and to global distributions, albeit of interest, is out the scope of the present study. Rather, we studied the degree of novelty of the EH homologs found in the TARA Oceans samples, by performing an all-vs-all blastp search.[16] Supporting Information, Fig. S7a shows the distribution of the similarity values of every hit against our reference sequences, and Supporting Information, Fig. S7b shows the values for all comparisons. The results revealed that the sequences of EH herein reported are distantly related to those in the TARA Oceans Project, which are also quite diverse. The active site effective volume calculation for all homologs (sequence identity >40%) in Tara Ocean samples indicated that all have volumes ranging from approx. 500 to 66.7 $Å^3$, which are above the 62.5 $Å^3$-throshold, and indicative of a broad substrate range, yet to be elucidated.

METHODS

Protein samples. Of the 145 EHs, the screening, cloning and expression of 69 were previously reported. The expression systems and purification conditions used in this study were as reported previously, in order to proceed with their characterization. These 69 EHs were identified by screening meta-genome clone libraries with the short esters α-naphthyl acetate and/or glyceryl tri-butyrate or, in one case, *p*-nitrophenyl-octanoate. Full details of these enzymes are extensively provided in Supporting Information, Table S1. The remaining 77 sequences encoding EHs are reported in this study for the first time and include 15 from Milazzo harbor (Sicily, Italy),[11,17] 9 from the Los Rueldos acid mine (Mieres, Spain),[18] 8 from a hydrothermal vent metagenome (Jan Mayen Vent Fields) (the meta-sequences are available from the National Center for Biotechnology Information (NCBI) database under the ID PRJNA296938, SAMN04111445), 8 from the River Elbe (Hamburg, Germany),[19] 6 from the El Max site (Alexandria, Egypt),[17] 4 from the Bizerte lagoon (Tunisia),[17] 3 from Messina harbor (Sicily, Italy),[17] 3 from the Gulf of Aqaba (Red Sea, Jordan),[17] 2 from Ancona harbor (Ancona, Italy),[20] 2 from elephant feces (Hagenbeck's Zoo, Hamburg, Germany),[21] 1 from Priolo Gargallo harbor (Syracuse, Italy),[17] 6 from *A. borkumensis* SK2,[22] 5 from *Cycloclasticus* sp. 78-ME,[23] 1 from *Oleiphilus messinensis* ME102T,[24] 1 from *Staphylococcus epidermidis*, 1 from *Pseudomonas oleovorans* DSM-1045 (genome not yet published), and 1 from *Geobacillus* sp. GHH01 (the genome sequence has been deposited in GenBank under accession no. CP004008).[25] In all cases, DNA extraction from the corresponding material,[11,17-25] preparation of pCCFOS1 libraries and naïve screening using α-naphthyl acetate and glyceryl tri-butyrate were performed as described elsewhere.[26] Positive clones containing presumptive EHs were selected, and their DNA inserts were sequenced using a MiSeq Sequencing System (Illumina, San Diego, USA) with a 2 × 150-bp sequencing kit. After sequencing, the reads

were quality-filtered and assembled to generate non-redundant meta-sequences, and genes were predicted and annotated as described previously.[27] The meta-sequences used for *in silico* screening are available from the NCBI database.[17,18,20] Protein-coding genes from metagenomes (sequence-based mining) and from the DNA inserts from positive clones (naïve screening) were screened (score > 45; e-value < $10e^{-3}$) using BLASTP and PSI-BLAST searches[28] for enzymes of interest against the ESTHER (*ESTerases and alpha/beta-Hydrolase Enzymes and Relatives*) and LED (*Lipase Engineering*) databases.[29,30]

As summarized in Supporting Information, Table S1, 128 of 145 genes encoding EHs were available or cloned in this study in common expression vectors (Ek/LIC 46, pET21a, pET22b, pVLT31, pCR-XL-TOPO, and p15Tv-L, among others) using a PCR-based approach and appropriate DNA samples as templates. The remaining set of 18 genes (accession codes KY483640-KY483649 and KY203030-KY203037) were synthesized by GenScript (Hong Kong) Limited in expression vectors. Of these 18 genes, the genes KY483640–KY483649 were synthesized in the pCDFDuet expression vector. The genes KY203030 to KY203037 were cloned using a recently developed vector suite that facilitates sub-cloning based on fragment exchange (FX) into multiple expression vectors.[31,32] Proteins were His-tagged at the C-terminus and purified as previously described (see Supporting Information, Table S1) with slight modifications. Briefly, selected *E. coli* clones that expressed each protein were grown at 37 °C on solid Luria Bertani (LB) agar medium supplemented with the appropriate antibiotics (Table S1), and one colony was picked and used to inoculate 10 mL of LB broth plus antibiotic in a 0.25-L flask. The cultures were then incubated at 37 °C and 200 rpm overnight. Afterward, 10 mL of this culture was used to inoculate 0.5 L of LB medium, which was then incubated to an OD_{600nm} of approximately 0.7 (ranging from 0.55 to 0.75) at 37 °C. Protein expression was induced by adding the appropriate inducer to a final concentration of approx. 1 mM (see Supporting Information, Table S1), followed by incubation for 16 h at 16°C. The cells were harvested by centrifugation at 5000 × *g* for 15 min to yield a pellet of 2-3 g/L pellet (wet weight). The wet cell pellet was frozen at -86 °C overnight, thawed and resuspended in 15 mL of 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0. Lysonase Bioprocessing Reagent (Novagen, Darmstadt, Germany) was added (4 µL/g wet cells) and incubated for 60 min on ice with rotating mixing. The cell suspension was sonicated for a total of 5 min and centrifuged at 15000 × *g* for 15 min at 4°C, and the supernatant was retained. The His-tagged proteins were purified at 4 °C after binding to a Ni-NTA His·Bind resin (Sigma Chemical Co. (St Louis, MO, USA)), followed by extensive dialysis of the protein solutions against 20 mM HEPES buffer (pH 7.0) by ultra-filtration through low-adsorption hydrophilic 10,000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) and storage at -86°C.

Purity was assessed as >98% using SDS-PAGE analysis[33] in a Mini PROTEAN electrophoresis system (Bio-Rad). The protein concentration was determined according to Bradford with bovine serum albumin as the standard.[34]

Quick tests for the production of active proteins and preparation of protein samples for hydrolysis activity assessment. Prior to substrate profiling, an activity test was performed to verify that each protein was active when expressed in *E. coli*. Protein activity was followed by monitoring the hydrolytic activity directly in the induced cells. Selected *E. coli* clones that expressed each protein were grown at 37°C on solid LB agar media supplemented with the appropriate antibiotics (Supporting Information, Table S1), and one colony was picked and used to inoculate 500 µL of LB broth plus antibiotic in a 2-mL Eppendorf tube. The culture was then incubated at 37 °C and 700 rpm in a Thermomixer (Eppendorf, Hamburg, Germany) for 7 h. To obtain uniform colonial growth, 5 µL of each culture was spotted on the surface of an LB agar plate (90 cm ø) supplemented with the appropriate antibiotic and expression inducer (Supporting Information, Table S1). The plates were incubated overnight at 37°C, and the agar surface was then covered with three different esters, 1-naphthyl acetate, glyceryl tri-acetate and tri-propionate, which are common EH substrates. Because EHs are commonly active at neutral or slightly alkaline pH, the activity tests were performed at pH 7.0. Briefly, the agar surface was covered with a layer of 20 mL of 1-naphthyl acetate/Fast Blue RR salt in HEPES buffer, pH 7.0, containing 0.4% (w/v) agar.[26] The hydrolysis of 1-naphthyl acetate was monitored by following the formation of an intense brown precipitate because of the release of naphthol, which was further oxidized. In parallel, a second plate was covered with a layer of 20 mL of 5 mM *N*-(2-hydroxyethyl)piperazine-*N*'-(3-propanesulfonic acid (EPPS) buffer supplemented with 0.45 mM Phenol Red as a pH indicator, 1 mL of a glyceryl tri-acetate and tri-propionate stock solution (200 mg/mL in acetonitrile each), and 0.4% (w/v) agar. If EH activity occurred, a yellow halo was evident around the colony due to acid formation. Clones containing all 145 EHs investigated in this study tested positive in in at least one of the two screening methods, indicating the production of soluble active proteins, as further confirmed by SDS-PAGE analysis in a Mini PROTEAN electrophoresis system (Bio-Rad).[33]

Selected *E. coli* clones that expressed active and soluble proteins were grown at 37 °C on solid LB agar supplemented with the appropriate antibiotics (Supporting Information, Table S1), and one colony was picked and used to inoculate 500 µL of LB broth plus antibiotic in a 2-mL Eppendorf tube. The culture was then incubated at 37 °C and 700 rpm in a Thermomixer (Eppendorf, Hamburg, Germany) for 7 h, after which 300 µL of each culture was used to seed LB agar petri dishes (90 cm ø) supplemented with the appropriate antibiotic and expression inducer (i.e., isopropyl β-D-1-thiogalactopyranoside or arabinose; see Table S1). The 300-µL culture was spread

on the plate to obtain uniform growth that covered the entire surface of the plate. The plates were incubated overnight at 37 °C. After incubation, 5 mL of 40 mM HEPES buffer, pH 7.0, was added to the surface of each plate. The bacterial cells were detached using sterile disposable Drigalsky spatulas, and the cell suspensions were transferred to a 5-mL Eppendorf tube and pelleted by centrifugation at 8000 rpm for 10 min at 4 °C. The supernatant was discarded, and the pellets were washed and centrifuged twice in the same buffer. The washed, wet pellets were weighed. For the activity tests, 100 mg of wet intact cells was re-suspended in 0.5 mL of 5 mM EPPS buffer, pH 8.0. This suspension was mixed by vortexing for 1 min and 2 μL of these suspensions were used immediately for the activity test as described in Methods. The activity was further confirmed using purified His-tagged proteins. Prior to use, stock solutions of 1-5 mg mL^{-1} protein in 5 mM EPPS buffer, pH 8.0, were prepared after extensive dialysis against this buffer by ultra-filtration through low-adsorption hydrophilic 10,000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) and 2 μL of protein solutions (from 1.0-17 μg of total protein depending on the specific activity of the particular EH) were immediately used for the activity tests.

Structural determinations. The proteins EH1 and EH102 were expressed and purified according to previously described procedures.[10] The His_6 tags were removed by TEV protease cleavage. The proteins were crystallized using the sitting-drop method in Intelliplate 96-well plates and a Mosquito liquid-handling robot (TTP LabTech), which mixed 0.5 μL of 20-25 mg mL^{-1} protein with 0.5 μL of the following reservoir solutions: EH1 - 0.1 M Tris pH 8.5, 0.2 M ammonium sulfate, 25% (w/v) PEG3350, EH102 - 0.1 M Tris pH 8.5, 0.2 M ammonium sulfate, 25% (w/v) PEG3350, 1/70 units of chymotrypsin. The crystal was cryoprotected with the reservoir solution supplemented with Paratone-N oil prior to flash freezing in an Oxford Cryosystems Cryostream. For the EH1 crystal, diffraction data were collected at 100 K and the Cu K° emission wavelength using a Rigaku HF-0007 rotating anode with a Rigaku R-AXIS IV++ detector. For the EH102 crystal, diffraction data were collected at 100 K and the Se absorption edge wavelength at the Structural Biology Center, Advanced Photon Source, beamline 19-ID using an ADSC Quantum 315R CCD detector. All diffraction data were reduced with HKL3000.[35] The structures were determined by molecular replacement using the structures from the Est2 Protein Databank (PDB) codes 1EVQ and 3RJT for EH1 and EH102, respectively, and the software Phenix.phaser.[36] Refinement was completed with Phenix.refine.[37] All B-factors were refined as isotropic with TLS parameterization. All geometries were verified using Phenix and the wwPDB server.

Protein Energy Landscape Exploration (PELE) sampling. We used Protein Energy Landscape Exploration (PELE) software to sample the binding modes of glyceryl tri-acetate with EH1 and EH102.[38,39] PELE is a Monte Carlo algorithm composed of a sequence of perturbation,

relaxation, and Metropolis acceptance tests. In the first step, the ligand is subjected to random rotations and translations, while the protein is perturbed based on the anisotropic network model (ANM).[40] The maximum allowed translation for ligand perturbation was 1.5 Å, and the maximum rotation was 20°. During the protein perturbation, all atoms were displaced by a maximum of 0.5 Å by moving the α-carbons following a random linear combination of the 6 lowest eigenvectors obtained in the ANM model. The relaxation step included the repositioning of all amino acid side chains within 6 Å of the ligand and the 5 side chains with the highest energy increase along the previous ANM step. The relaxation stage ended with a truncated Newton minimization using the OPLS all-atom force field and an implicit surface-generalized Born continuum solvent. The new proposed minima were then accepted or rejected based on a Metropolis test. The substrate binding plots contained all accepted conformations for three 12-h simulations using 200 processors.

Homology modeling. Homology models were developed using Prime software from Schrödinger. Prime uses BLAST (with BLOSUM62 matrix) for homology search and alignment and refines the results using the Pfam database and pairwise alignment with ClustalW.

Cavity Volume and Solvent Accessible Surface Area (SASA) calculation. The relative Solvent Accessible Surface Area (SASA) for a residue was obtained using the GetArea web server.[41] This service allows a user to submit a PDB file and retrieve the relative SASA or the solvation energy in a variety of formats. This server has several options and allows the user to compute individual residue exposure under the "Select desired level of output" option. Thus, the "exposure" of the active site can be computed using the catalytic amino acids, e.g. the conserved catalytic Ser/Asp/His triad in esterases and the two conserved aspartic catalytic residues in HAD phosphatases. Cavity volumes were computed with Fpocket,[42] a very fast open-source protein pocket (cavity) detection algorithm based on Voronoi tessellation. Fpocket includes two other programs (dpocket and tpocket) that allow the extraction of pocket descriptors and the testing of owned scoring functions, respectively.

SUPPORTING INFORMATION REFERENCES

(1) Ménigaud, S., Mallet, L., Picord, G., Churlaud, C., Borrel, A., Deschavanne, P. (2012) GOHTAM: a website for 'Genomic Origin of Horizontal Transfers, Alignment and Metagenomics'. *Bioinformatics 28*, 1270-1271.

(2) Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Rico Eguizabal, E., Guerrero, Mdel C., Yakunin, A. F., and Ferrer, M. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol.* 79, 3553-3562.

(3) Price, M.N., Dehal, P.S., Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5, e9490.

(4) Arpigny, J. L., and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J. 343*, 177-183.

(5) Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., and Golyshin, P. N. (2015) Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransform. 33*, 235-249.

(6) Huang, H., Pandya, C., Liu, C., Al-Obaidi, N.F., Wang, M., Zheng, L., Toews Keating, S., Aono, M., Love, J. D., Evans, B., Seidel, R. D., Hillerich, B. S., Garforth, S. J., Almo, S. C., Mariano, P. S., Dunaway-Mariano, D., Allen, K. N., and Farelli, J. D. (2015) Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. USA 112,* E1974-1983.

(7) Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol. 12*, 635-645.

(8) Janes, L. E., Löwendahl, C., and Kazlauskas, R. J. (1998) Rapid quantitative screening of hydrolases using pH indicators. Finding enantioselective hydrolases. *Chem. Eur. J. 4*, 2317-2324.

(9) Alcaide, M., Tornés, J., Stogios, P.J., Xu, X., Gertler, C., Di Leo, R., Bargiela, R., Lafraya, A., Guazzaroni, M. E., López-Cortés, N., Chernikova, T. N., Golyshina, O. V., Nechitaylo, T.Y ., Plumeier, I., Pieper, D. H., Yakimov, M.M., Savchenko, A., Golyshin, P. N., and Ferrer, M. (2013) Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochem. J. 454*,157-166.

(10) Alcaide, M., Stogios, P. J., Lafraya, Á., Tchigvintsev, A., Flick, R., Bargiela, R., Chernikova, T. N., Reva, O. N., Hai, T., Leggewie, C.C., Katzke, N., La Cono, V., Matesanz, R., Jebbar, M., Jaeger, K. E., Yakimov, M. M., Yakunin, A. F., Golyshin, P. N., Golyshina, O. V., Savchenko, A., Ferrer, M. (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol. 17*, 332-345.

(11) Popovic, A., Hai, T., Tchigvintsev, A., Hajighasemi, M., Nocek, B., Khusnutdinova, A. N., Brown, G., Glinos, J., Flick, R., Skarina, T., Chernikova, T. N., Yim, V., Brüls, T., Paslier, D.L., Yakimov, M. M., Joachimiak, A., Ferrer, M., Golyshina, O. V., Savchenko, A., Golyshin, P. N., and Yakunin, A. F. (2017) Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. *Sci. Rep. 7*, 44103.

(12) Barak, Y., Nov, Y., Ackerley, D. F., and Matin, A. (2008) Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J. 2*,171-179.

(13) Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol. 9*, 22-34

(14) Steinegger, M., Soeding, J. (2016) Sensitive protein sequence searching for the analysis of massive

data sets. *bioRxiv* 079681.

(15) Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B. T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S.; Tara Oceans coordinators, Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M. B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., and Bork P. (2015) Structure and function of the global ocean microbiome. *Science 348*, 1261359.

(16) Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. L (2009) BLAST+: architecture and applications. *BMC Bioinformatics 10*, 421.

(17) Bargiela, R., Mapelli, F., Rojo, D., Chouaia, B., Tornés, J., Borin, S., Richter, M., Del Pozo, M. V., Cappello, S., Gertler, C., Genovese, M., Denaro, R., Martínez-Martínez, M., Fodelianakis, S., Amer, R. A., Bigazzi, D., Han, X., Chen, J., Chernikova, T. N., Golyshina, O. V., Mahjoubi, M., Jaouanil, A., Benzha, F., Magagnini, M., Hussein, E., Al-Horani, F., Cherif, A., Blaghen, M., Abdel-Fattah, Y. R., Kalogerakis, N., Barbas, C., Malkawi, H. I., Golyshin, P. N., Yakimov, M. M., Daffonchio, D., and Ferrer, M. (2015) Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. *Sci Rep 5*, 11651.

(18) Méndez-García, C., Mesa, V., Sprenger, R. R., Richter, M., Diez, M. S., Solano, J., Bargiela, R., Golyshina, O. V., Manteca, Á., Ramos, J. L., Gallego, J. R., Llorente, I., Martins dos Santos, V. A., Jensen, O. N., Peláez, A. I., Sánchez, J., and Ferrer, M. (2014) Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. *ISME J. 8*, 1259-1274.

(19) Rabausch, U., Juergensen, J., Ilmberger, N., Böhnke, S., Fischer, S., Schubach, B., Schulte, M., Streit, W. R. (2013) Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl. Environ. Microbiol. 79*, 4551-4563.

(20) Bargiela, R., Gertler, C., Magagnini, M., Mapelli, F., Chen, J., Daffonchio, D., Golyshin, P. N., and Ferrer, M. (2015) Degradation network reconstruction in uric acid and ammonium amendments in oil-degrading marine microcosms guided by metagenomic data. *Front. Microbiol. 6*, 1270.

(21) Güllert, S., Fischer, M. A., Turaev, D., Noebauer, B., Ilmberger, N., Wemheuer, B., Alawi, M., Rattei, T., Daniel, R., Schmitz, R. A., Grundhoff, A., and Streit, W. R. (2016) Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnol. Biofuels 9*, 121.

(22) Schneiker, S., Martins dos Santos, V. A., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., Chernikova, T. N., Denaro, R., Ferrer, M., Gertler, C., Goesmann, A., Golyshina, O. V., Kaminski, F., Khachane, A. N., Lang, S., Linke, B., McHardy, A. C., Meyer, F., Nechitaylo, T., Pühler, A., Regenhardt, D., Rupp, O., Sabirova, J. S., Selbitschka, W., Yakimov, M. M., Timmis, K. N., Vorhölter, F. J., Weidner, S.,

Kaiser, O., and Golyshin, P. N. (2006) Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nat. Biotechnol. 24*, 997-1004.

(23) Messina, E., Denaro, R., Crisafi, F., Smedile, F., Cappello, S., and Genovese, M. (2016) Genome sequence of obligate marine polycyclic aromatic hydrocarbons-degrading bacterium *Cycloclasticus* sp. 78-ME, isolated from petroleum deposits of the sunken tanker Amoco Milford Haven, Mediterranean Sea. *Mar. Genomics* 25, 11-13.

(24) Golyshin, P. N., Chernikova, T. N., Abraham, W. R., Lünsdorf, H., Timmis, K. N., Yakimov, M. M. (2002) Oleiphilaceae fam. nov., to include *Oleiphilus messinensis* gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. *Int. J. Syst. Evol. Microbiol. 52*, 901-911.

(25) Wiegand, S., Rabausch, U., Chow, J., Daniel, R., Streit, W. R., Liesegang, H. (2013) Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. *Genome Announc.* 1, e0009213.

(26) Reyes-Duarte, D., Ferrer, M., and García-Arellano, H. (2012) Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. *Methods Mol Biol 861*, 101-113.

(27) Placido, A., Hai, T., Ferrer, M., Chernikova, T. N., Distaso, M., Armstrong, D., Yakunin, A. F., Toshchakov, S. V., Yakimov, M. M., Kublanov, I. V., Golyshina, O. V., Pesole, G., Ceci, L. R., Golyshin, P. N. (2015) Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay, Vulcano Island (Aeolian archipelago) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Appl. Microbiol. Biotechnol. 99*, 10031-10046.

(28) Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389-3402.

(29) Fischer, M., and Pleiss, J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res. 31*, 319-321.

(30) Barth, S., Fischer, M., Schmid, R. D., and Pleiss J. (2004) The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. *Bioinformatics 20*, 2845-2847.

(31) Geertsma, E. R., and Dutzler, R. (2011) A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry 50*, 3272-3278.

(32) Bjerga, G. E. K., Arsın, H., Larsen, Ø., Puntervoll, P., Kleivdal, H. T. (2016) A rapid solubility-optimized screening procedure for recombinant subtilisins in *E. coli*. *J. Biotechnol.* 20, 38-46.

(33) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 680-685.

(34) Bradford, M. M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248-254.

(35) Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr. 62*, 859-866.

(36) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D: Biol. Crystallogr. 66*, 213-221.

(37) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D: Biol. Crystallogr. 60*, 2126-2132.

(38) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B 105*, 6474-6487.

(39) Borrelli, K. W., Vitalis, A., Alcantara, R., and Guallar, V. (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Chem. Theory Comput. 1*, 1304-1311.

(40) Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Hall, M. D., Zhang, J., and Friesner, R. A. (2013) Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum. Chem. 113*, 2110-2142.

(41) Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem. 19*, 319.

(42) Guilloux, V. L., Schmidtke, P., and Tuffery, P. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics 10*, 168.

3.2. Computer-aided laccase engineering: toward biological oxidation of arylamines

Gerard Santiago†, Felipe de Salas‡, M. Fátima Lucas†§, Emanuele Monza†, Sandra Acebes†, Ángel T. Martinez‡, Susana Camarero*‡, and Víctor Guallar*†‖

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

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

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

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

ACS Catalysis

Research Article

pubs.acs.org/acscatalysis

Computer-Aided Laccase Engineering: Toward Biological Oxidation of Arylamines

Gerard Santiago,[†,⊥] Felipe de Salas,[‡,⊥] M. Fátima Lucas,[†,§,⊥] Emanuele Monza,[†] Sandra Acebes,[†] Ángel T. Martinez,[‡] Susana Camarero,[*,‡] and Víctor Guallar[*,†,∥]

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

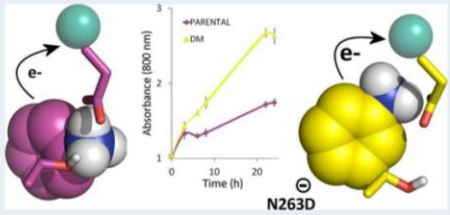
[‡]Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

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

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

Supporting Information

ABSTRACT: Oxidation of arylamines, such as aniline, is of high industrial interest, and laccases have been proposed as biocatalysts to replace harsh chemical oxidants. However, the reaction is hampered by the redox potential of the substrate at acid pH, and enzyme engineering is required to improve the oxidation. In this work, instead of trying to improve the redox potential of the enzyme, we aim toward the (transient) substrate's potential and propose this as a more reliable strategy. We have successfully combined a computational approach with experimental validation to rationally design an improved biocatalyst. The in silico protocol combines classical and quantum mechanics to deliver atomic and electronic level detail on the two main processes involved: substrate binding and electron transfer. After mutant expression and comparison to the parent type, kinetic results show that the protocol accurately predicts aniline's improved oxidation (2-fold k_{cat} increase) in the engineered variant for biocatalyzed polyaniline production.



KEYWORDS: *laccase, aniline, PELE, QM/MM, protein engineering, PANI, arylamines*

INTRODUCTION

There is a rising interest in replacing harsh and environmentally unfriendly industrial chemical reactions by clean enzyme-based processes.[1] Although the practical use of proteins goes back to ancient times, the large-scale usage of well-characterized enzymes in the textile, detergent, and starch industries is the result, in most cases, of considerable advances in biotechnology.[2] Protein engineering based on rational design, directed evolution, and/or computer simulations has shown considerable success, but altering proteins remains an arduous process.[3]

Laccases (EC 1.10.3.2) are oxidases that contain four copper ions: a trinuclear (T2/T3) cluster buried in the protein matrix, where molecular oxygen is reduced to water, and a T1 site, close to the surface of the protein, where substrates are oxidized. The promiscuity of these proteins toward a large number of organic compounds, the use of oxygen as a final electron acceptor (with no need for expensive cofactors), and production of water as the sole byproduct converts them into ideal biocatalysts for diverse technological purposes.[4] The relatively low redox potential (E° = 0.4−0.8 V), in comparison to those of other oxidoreductases such as peroxidases, limits their application,[5] but the use of mediators has been found to expand their action toward more difficult substrates.[6] Trying to improve laccase's redox potential is a current strategy for increasing the oxidation capability of these proteins,[7] but the observation that K_M and k_{cat} values of different laccases are ligand dependent evidence that oxidation is not reliant merely on the redox potential difference.[7a,c]

To advance in this matter, we have used a recently developed computational protocol[8] that combines the protein energy landscape exploration (PELE)[9] software with quantum mechanics/molecular mechanics (QM/MM)[10] techniques. PELE is an all-atom Monte Carlo based algorithm routinely used in our laboratories to map long-time-scale dynamic processes such as global protein−substrate exploration or locally induced fit events.[11] In redox processes, for example, we use it to identify important parameters such as donor−acceptor distance (DAD)[12] and the solvent-accessible surface area (SASA) of the substrate.[12a,13] Changes in these two parameters are expected to correlate with key variables in electron transfer theory: (i) a shorter DAD points to a larger electronic coupling; (ii) a smaller SASA suggests smaller solvent reorganization energy.[12] Then, using QM/MM we calculate spin densities on

Received: May 24, 2016
Revised: July 1, 2016
Published: July 8, 2016

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

selected structures, obtaining an estimation of the electron transfer driving force.[10a,14] In our previous study[8] we found that 2,6-dimethoxyphenol's enhanced oxidation by an evolved laccase was the result of the substrate's rearrangement in the active site, with no important change in the redox potential of the T1 copper. In this study we challenge this computational protocol by rationally engineering a laccase for aniline (ANL) oxidation under preferred conditions for industrial applications.

Aniline and its derivatives have been traditionally used in the dyestuff industry, as precursors of aniline dyes and azo dyes for textiles, hair, leather, or printing,[15] and also as components for engineering polymers, composites, and rubbers. In particular, the synthesis of conductive polyaniline (PANI) has been widely investigated during the last two decades due to its wide range of applications (sensor devices, rechargeable batteries, etc.[16]). Conducting PANI macromolecules have regular head-to-tail linked monomers (over 95%), but highly conductive polymer is only produced under strongly acidic conditions (pH <2.5) using oxidants such as ammonium peroxydisulfate.[17] To replace these processes by enzymatic bioconversion is thus a main goal;[18] however, aniline (pK_a = 4.6) is mostly protonated at low pH. Since the anilinium cation ($E°$ = 1.05 V) is much less oxidable than nonprotonated aniline (ANL) ($E°$ = 0.63 V), it becomes clear why laccases normally cannot oxidize this compound or reactions occur too slowly.[19] The high potential barrier for oxidizing aniline in acid medium[20] demands the use of high-redox-potential laccases ($E°$ ≈ +0.8 V) such as the one used in this study, whereas other fungal laccases with lower redox potential are unable to catalyze the reaction.

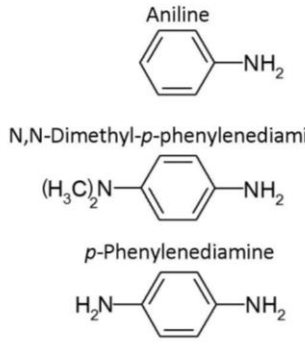
The strategy, presented in this work, for improving the activity of a high-redox-potential recombinant laccase previously obtained by directed evolution,[21] aims at precision enzyme design instead of screening a large number of candidates.[22] By means of our computational protocol, two-point mutations were predicted and experimentally validated in one single protein variant. The double mutant (DM) was produced in *S. cerevisiae* and compared with the parental laccase. Kinetic analysis showed a 2-fold k_{cat} increase for ANL oxidation. Furthermore, oxidation of *p*-phenylenediamine (PPD), 2,2′-azinobis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), and *N,N*-dimethyl-*p*-phenylenediamine (DMPD) (see Scheme 1 for chemical structures) was also improved 1.4-, 2.0-, and 1.6-fold, respectively. DMPD, in particular, is a precursor of the high-value reagent methylene blue, with applications in a wide range of fields (including as a potential antimalarial agent),[23] which nowadays is produced by oxidation of DMPD with ferric chloride (a highly toxic compound) and hydrogen sulfide dissolved in hydrochloric acid.

Overall, we demonstrate the laccase's enhancement toward the aforementioned noteworthy technological applications, reinforcing the potential for synergetic experimental and computational protocols in predictive studies.[24]

RESULTS AND DISCUSSION

The work is presented in three sections. We begin by confirming the importance of the ligand binding event[8,25] in ANL oxidation by a laccase developed in a previous directed evolution experiment.[21] After obtaining these molecular details, we rationally designed the protein binding site for improved ANL oxidation. Second, the precision designed variant was experimentally obtained and compared against the parent type for oxidation of ANL, DMPD, and PPD, thermostability, and pH range. Finally, computational cross-validation for the observed improvement in DMPD oxidation was performed.

Scheme 1. Chemical Structures of All Studied Arylamines: Aniline (ANL), *N,N*-Dimethyl-*p*-phenylenediamine (DMPD), and *p*-Phenylenediamine (PPD)



In Silico Study and Rational Design for ANL Oxidation. The computational protocol employed, described in Methodology, begins by determining the binding event. For this, PELE, a Monte Carlo based molecular simulation software, was employed. The program performs random substrate and protein perturbations aiming at reproducing fully flexible protein−substrate binding, including induced-fit effects. Here 120 independent 48 h trajectories were produced to investigate the interactions between the laccase variants and each of the studied substrates. By analyzing the laccase−substrate interaction energy profiles produced by PELE (Figure 1), we were capable of identifying the lowest energy minima and thus the

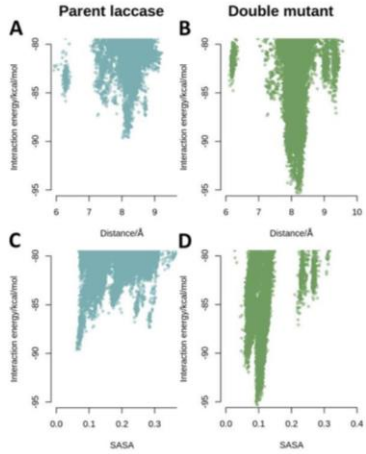


Figure 1. Protein−substrate interaction energies vs distance between the center of mass of ANL and the copper T1 atom (top) and vs SASA (bottom) for parent laccase (A, C) and double mutant (B, D).

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

main binding modes at the T1 copper site. Once these were identified, QM/MM calculations were performed to estimate the fraction of the spin density localized on the substrate. QM/MM calculations split the system into two regions: quantum and classical. In the first, which includes the T1 copper with its first coordination sphere, electronic-based methods were used to study the electron transfer. The second region, the classical one, includes the rest of the system, which provides the correct geometric and electrostatic environment to the quantum region. For these calculations, 50 complexes were randomly selected using a distance and interaction energy threshold. Selection was restrained to the substrate positions within 10 Å of the T1 copper atom site and 5 kcal/mol range from the lowest interaction energy. Interaction energy plots for ANL diffusion in the parent type laccase are shown in Figure 1A.

Observation of the lowest interacting energy complexes in the parent type protein (Figure 1A) shows ANL placed in the T1 cavity close to N207, N263, and S264 and H-bonded to D205 (Figure 2A). QM/MM calculations performed on randomly selected structures (see Methodology) show that these exhibit 12% spin density (Table 1; average over 50 snapshots).

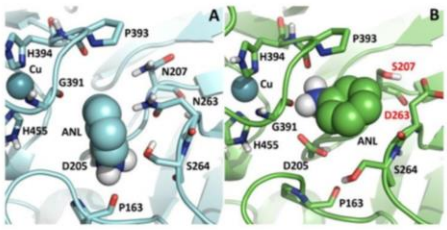


Figure 2. Representative minima for ANL interaction with (A) parent laccase and (B) double mutant.

Table 1. Average Spin Densities (SP) and Experimental Kinetic Data for Parent Laccase and Double Mutant (DM) Oxidizing ANL, DMPD, PPD, and ABTS

system	SP (%)	k_{cat} (s^{-1})	K_M (mM)
Lac + ANL	12 ± 1	10.1 ± 1.1	28 ± 7.2
DM + ANL	26 ± 5	22.6 ± 3.4	59.3 ± 15.5
Lac + DMPD	39 ± 7	459 ± 18	1.7 ± 0.2
DM + DMPD	68 ± 14	741 ± 48	1.2 ± 0.2
Lac + PPD	32 ± 5	14.7 ± 0.6	3.7 ± 0.3
DM + PPD	46 ± 14	20.8 ± 2.1	3.5 ± 0.8
Lac + ABTS	23 ± 14	291 ± 18	0.0042 ± 0.001
DM + ABTS	40 ± 7	570 ± 26	0.01 ± 0.001

The next step was to redesign the binding site to improve ANL's oxidation rate (at pH 3) by this laccase. Since changing the redox potential of the copper site is a complex task, in particular in a high-redox-potential laccase,[26] and conscious of the effect of the binding event on the overall oxidation rate,[7c,8,25b,27] we have instead targeted to locally modify the ANL (anilinium cation) redox potential. We wish to stabilize the oxidized form of ANL, boosting electron transfer to the Cu T1 site which, in theory, can be accomplished by introducing a second negatively charged residue (in addition to D205) next to the ANL binding site. To test this hypothesis, we probed alternative single-point mutations. Initially, we tested mutations in position 205, since the lowest energy structures showed a direct interaction between ANL and that residue in the parent type laccase. Furthermore, the literature[28] pointed to a potential benefit from D205E mutation but computational QM/MM scoring indicated poorer activity. Other mutations, D205Y and D205A, led to similar results (all computed spin densities are available in the Supporting Information). In addition to D205, two other residues are seen in direct contact with the substrate: residues 207 and 263, both asparagines. Since our criteria for selecting the mutations was to improve the electrostatic environment around the substrate, we started by testing single-point mutations to aspartic acid (side-chain length identical with that of asparagine) on each residue (N207D/N263D). This double mutant showed considerable spin density improvement (32%), but two aspartic acids so close each to other could be troublesome for protein stability. For this reason, we tested which of these positions would have the most beneficial effect on ANL oxidation (Table S1 in the Supporting Information) and found N263D to be the best candidate. Finally, positon 207 was mutated to serine because this mutation was found to be a good compromise among ligand binding, activity, and protein stability. Ligand binding on the double mutant N207S/N263D (DM) displays an interaction energy profile slightly improved from the parent protein (Figure 1B), with a shorter DAD; in particular the amine group is on average closer to the Cu atom (Figure 2B) and has a similar SASA (Figure 1C,D). Activity was scored with QM/MM calculations, which showed an enhanced 26% spin density, in comparison with 12% for the parent type laccase reaction (Table 1). Finally, the likelihood of disrupting protein stability was assessed through sequence analysis (see below).

Introduction of N207S/N263D modifies the substrate−protein interactions as seen in Figures 1B and 2B. The new minima show that the substrate is now located with a different orientation and the amine group closer to the copper atom thanks to the smaller side chain of S207 (Figure 2B), enclosed between the two negatively charged D205 and D263 residues. The presence of these two residues so close to each other could also alter their protonation states (especially at pH 3); thus, to verify this possibility we have used an H++ server to assess the pK_a of these amino acids in the mutant protein. Predictions indicate a clear deprotonated state for both residues. Furthermore, DM interacts with ANL through the backbone carbonyl atoms of residues G391 and P393, the latter being connected to the copper-coordinated H394.

We have also performed a sequence space search to check if other proteins carry the proposed mutations. We found, using HotSpot Wizard (loschmidt.chemi.muni.cz/hotspotwizard), that position N207 contains 34 out of 50 times an asparagine while 5 other sequences have serines in that position (default settings used). In the case of N263 also 34 out of 50 sequences have an asparagine but only two have a glutamic and one has an aspartic acid. We have further verified these results by performing a BLAST search using the Prime module of Maestro software. After multiple sequence alignment with ClustalW (in the Supporting Information) we found that out of 28 sequences (with available crystal structures) 1 contains the substitutions proposed in the DM (Figure 3). This multicopper oxidase (MCO), sharing 29% sequence identity with our parent type, is not a laccase but the membrane Fet3 protein from *S. cerevisiae*, which catalyzes the oxidation of Fe(II) to Fe(III) (PDB entry 1ZPU).[29]

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

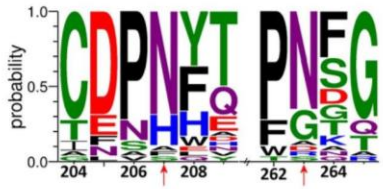


Figure 3. Sequence logo including the residues mutated in this work. The positions 207 and 263 are identified by a red arrow The sequences belong to multicopper oxidases with at least 30% sequence identity to the parent type used in this work.

The fact that we find the N263S mutation in 5 other sequences and the 2 N207D/N263S mutations in another MCO helps determine the probability of obtaining a stable variant.

Experimental Validation. In order to verify the effect of the mutations predicted in silico, the N207S/N263D mutations were introduced in the parent laccase. The resulting DM was produced in *S. cerevisiae*, purified to 100% homogeneity, and its activity with ANL at pH 3 compared with that of the pure parent type (Table 1). The enzyme's turnover for aniline increased 2-fold with respect to the parental laccase. In addition, DM showed 1.6- and 1.4-fold improvement of k_{cat} with the aniline analogues DMPD and PPD and no increment in K_M, resulting in an enhancement of the catalytic efficiency with regard to the parent type. The notably higher turnover rates of both enzymes on DMPD (with respect to ANL) can be explained by the fact that, assuming that for each DMPD molecule only one nitrogen is positively charged, the reciprocal 1−4-position of the amine groups stabilizes the radical more than in ANL (since an electron-donating group is in *para* position with respect to the double positive charge).

It has been shown that a conserved acid residue (D206 in *Trametes versicolor* laccase; Glu235 in *Myceliophtora albomyces* laccase) has a key role in the oxidation of phenols by fungal laccases. This residue assists the transfer of the proton while the electron is transferred through one His coordinating T1 copper (H458 in *T. versicolor* laccase), giving rise to the phenoxy radical. The concerted electron−proton transfer first described by Galli and co-workers[30] was confirmed in a recent work.[27c] Nevertheless, improved oxidation of ABTS (Table 1), which does not require a proton transfer, in the mutated variants favors the idea that changes observed in k_{cat} are most likely due to an improved local oxidation of the substrate (thanks to an additional negatively charged residue) and not in a facilitated proton transfer. To validate this hypothesis, we have further performed quantum calculations on a model system; the energy difference between the oxidized and nonoxidized forms of ANL in the presence and absence of a negatively charged carboxylic acid (in the same position of D263) was calculated (Table S2 in the Supporting Information).[31] Results indicate that the energy difference, and therefore oxidation, is more favorable when the negative charge is added, corroborating our hypothesis (for further details please see the Supporting Information).

The correlation between k_{cat} and spin density (percent) values for both laccases (and identical results seen in other studies[8,27c]) confirmed the usefulness of simulation studies for the rational design of these enzymes.

To better assess the possible benefit of this kinetic improvement for laccase application, we compared the activity and stability of the DM with those of the parent laccase under the working conditions for the enzymatic synthesis of polyaniline, which is pH 3 in the presence of anionic surfactants. DM presented a pH activity profile similar to that of the parental laccase (Figure 4A). As for the stability at acid

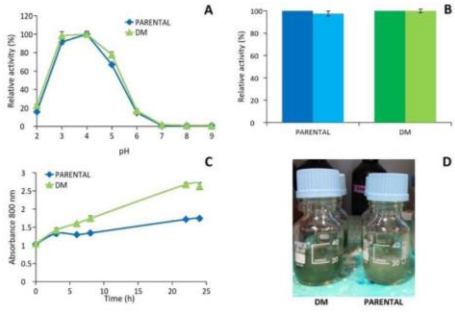


Figure 4. Comparison of double mutant (DM) and parental laccase for (A) optimum pH for aniline oxidation (300 mM), (B) laccase stability at pH 3 (initial activity, dark color, and residual activity after 5 h of incubation, light color, are shown), (C) enzymatic polymerization of 15 mM aniline along time (with 5 mM SDBS as template) as shown by the increase of absorbance at 800 nm, and (D) green PANI produced in (C) by parental and DM laccases.

pH, DM maintained the high stability of the parent laccase at pH 3, retaining 100% initial activity after 5 h of incubation (Figure 4B). We also checked the laccase's thermostability since, very frequently, mutations with beneficial effects on enzyme activity might be detrimental for its stability. In this case, T_{50} (the temperature at which the activity is 50% of the initial activity) was lowered marginally by 1.5 °C upon mutation of laccase (T_{50} = 63.5 °C for DM versus T_{50} = 65 °C for the parental laccase). Finally, we assayed both laccases as biocatalysts for the synthesis of water-soluble polyaniline (PANI). The enzymatic polymerization of aniline was carried out in citrate−phosphate buffer pH 3, using dodecyl benzenesulfonate (SDBS) as a doping template to provide the para coupling of the protonated ANL monomers. The synthesis of PANI was followed by monitoring the increment of absorbance at 800 nm. This absorbance band is the distinctive signal for emeraldine salt that is the green, protonated, and conductive form of PANI.[32] When DM catalyzed the reaction, the absorbance at 800 nm was significantly raised. Up to 36% higher absorbance was obtained after 22 h of reaction with DM with respect to the value obtained with the parental laccase (Figure 4C). This increment correlated with superior production of green color during the polymerization of ANL catalyzed by DM (Figure 4D).

DMPD in Silico Cross-Validation. The aniline derivative, DMPD, was also chosen to compare the activity of DM and parent laccase. Since kinetic analysis showed a significant k_{cat} and K_M improvement and a literature survey reveals this compound to be a precursor for the production of methylene blue, a high-value reagent, we have cross-validated the computational method with DMPD. The same protocol described for ANL was used for DMPD (keeping in mind

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

that at pH 3 two possible protonation states can coexist). PELE simulations for the parent type laccase and DM evidence the positive effect of both single-point mutations in DAD (as conformations closer to the copper atom are possible; Figures S1 and S2 in the Supporting Information). N263D mutation allows DMPD to adopt a new position more favorable for the oxidation (Figure 5; results for the protonated tertiary amine

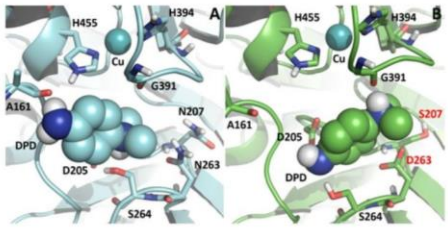


Figure 5. Representative minima for DMPD (protonated tertiary amine) interaction with (A) parent laccase and (B) DM.

where changes in DM are more pronounced). DMPD is initially interacting with H455 and A161 (Figure 5A), while in DM it is more buried. DMPD in DM is interacting with the two acid side chains (D205 and D263) which generate a more favorable electrostatic environment for electron abstraction (Figure 5B).

The presence of N207S/N263D mutations in the engineered laccase also displays an increased spin density of 68% (as opposed to 39% for the parent laccase; Table 1), confirming that the computational method is consistently capable of reproducing the observed increase in k_{cat} for both laccase substrates. It should be noted that spin densities must be taken as relative quantities: that is to say, a reference must always be available, preferentially the parent protein. Depending on the nature of the substrate and the local environment, spin densities vary greatly but in relative terms are accurate predictors for improved k_{cat}.

■ CONCLUSIONS

On the basis of a computational strategy optimizing DAD, SASA, and protein ligand recognition, we rationally introduced two active site point mutations in a high-redox-potential laccase, aiming at a more favorable oxidation of ANL. Following this (classical force field) conformational space search, improved oxidation was confirmed with mixed QM/MM techniques. This combined effort provides a reliable in silico screening, reducing experimental validation to precisely designed mutants. In particular, the introduction of a negatively charged residue (N263D) has increased ANL's local oxidation (stabilizing the anilinium cation), thus favoring electron transfer. Experimental results confirm that the recalcitrant oxidation of ANL shows a change in k_{cat} from 10 to 23 s^{-1}. The increment of K_M is not troublesome in terms of direct industrial application, since high substrate concentrations are employed. Screening for improved oxidation of DMPD evidenced the enhancement of the catalytic efficiency, thanks to a simultaneous improvement of k_{cat} and similar K_M. For these particular cases, the computational method appears to provide not only a qualitative prediction to k_{cat} enhancement but also a notable quantitative agreement with the computationally estimated spin density (2.2- and 1.6-fold for ANL and DMPD, respectively). Moreover, in line with earlier findings[8,25a,33] on the importance of the binding event, our results confirm that targeting the substrate's local (transient) oxidation is a reliable strategy for laccase activity improvement. These results are encouraging for future synergic computational and experimental studies aiming at bio-based oxidation of small arylamines with vast applications.

■ METHODOLOGY

Systems Setup. The high redox potential laccase from basidiomycete PM1 (PDB entry 5ANH, provided by Dr. Javier Medrano) was used as a template (96% sequence identity) for the evolved laccase.[21] For purposes of clarity, the protein redox potentials referenced along the text, unless otherwise mentioned, refer to the copper T1 site. The protein structure was prepared with assistance of Protein Preparation Wizard,[34] PROPKA,[35] and the H++ web server[36] to determine the protonation state of all ionizable amino acids at pH 3.

Four substrates have been employed in the computational study: ANL and DMPD, PPD, and ABTS. Substrates were fully optimized with the density functional M06[37] with the 6-31G** basis set in an implicit solvent, and the electrostatic potential charges computed at the same level of theory were used in the following force field based PELE simulations. ANL was modeled as an anilinium cation, since it is the dominating form at low pH needed for PANI production, while DMPD (also studied at pH 3) was considered in two possible protonation states: charged on the primary or the tertiary amine.

Protein−Substrate Binding Modes. To determine the different binding modes of substrates in the T1 copper site, PELE[9a,38] simulations were performed following the protocol described in detail by Monza et al.[8] Since the ligand's binding site in laccases is well-known, each substrate was manually placed close to the entrance of the copper site and was then free to move in a region 20 Å within the T1 copper atom. Local conformational exploration was accomplished by perturbing the ligand and the protein in a sequence of steps followed by side-chain readjustment (using rotamer libraries) and all-atom minimization. It should be noted that this procedure is unlike most docking approaches, since not only the dynamics of the ligand are taken into consideration but also the protein's side chains and backbone. The new protein−substrate conformations are then accepted or rejected through a Metropolis test based on energy (E) differences computed using an OPLS force field[39] and a surface-generalized Born (SGB) implicit continuum solvent.[40] Interaction energies have been computed as $IE = E_{PS}^{SGB} - (E_{P}^{SGB} + E_{S}^{SGB})$, where PS refers to the protein (P) substrate (S) complex. From the local laccase−substrate sampling several important parameters (that may affect electron transfer) can be extracted. In the first place, by analyzing the overall interaction energy profile, we identify the best binding poses. The average DAD is also computed since changes in electronic coupling can be qualitatively estimated through this quantity.[12] Then, the SASA gives indications of changes in the electronic reorganization energy (its solvent component), as a more exposed ligand will have a higher energetic penalty.[12a,13,41]

QM/MM Electron Transfer. After selection of the lowest interaction energy complexes obtained in the PELE simulations these were used to estimate the amount of spin density transferred from the substrate to the copper T1 site. For this,

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

we have used QM/MM methods, which employ both quantum mechanics and molecular mechanics.[10,14] QSite[42] was used, including in the quantum region the entire copper site (with equatorial and axial ligands) as well as the substrate. The M06-L[43] density functional with lacvp* basis set was used (LANL2DZ effective core basis set for the copper atom and 6-31G* for the rest of the atoms). The remaining part of the protein was modeled in the molecular mechanics section through an all-atom OPLS force field. Mulliken[44] populations were computed to characterize the spin density transferred from the substrate to the copper site. Previous studies have shown that these are insensitive to the functional employed and the initial guess used for the wave function (specifying localized spin on either the donor/acceptor or neither at the beginning of the calculation does not change the results).[8] In addition, previous experimental and theoretical studies have shown that the spin density correlates well with oxidation potentials of donors in ET proteins, establishing if an unpaired electron is energetically more stable on the donor or acceptor molecular orbitals.[8,45]

The rationale behind the choice of the point mutations presented in this work consists of increasing spin density along with decreasing SASA and DAD, which is expected to lead to an improvement in k_{cat}.[8] The aim of simultaneously reducing the solvent reorganization energy and improving the driving force is justified by the high reorganization energy of laccases (above 30 kcal/mol[7c]), which safely places our designs far from the Marcus inverted region.[13] Otherwise, more attention should be paid to the driving force−reorganization energy balance.

Reagents and Enzymes. DMPD, ANL, PPD, and the *S. cerevisiae* transformation kit were purchased from Sigma-Aldrich. ABTS and the high pure plasmid isolation kit was purchased from Roche. Zymoprep yeast plasmid miniprep II kit was purchased from Zymo Research, and the plasmid midi kit and QIAquick gel extraction kit were obtained from QIAGEN. PFU Ultra High-Fidelity DNA polymerase was purchased from Agilent Technologies.

Strains and Culture Media. The uracil-independent and ampicillin-resistant shuttle pJRoC30 vector carrying the parent laccase construct under the control of the GAL1 promoter was generated as previously described.[46] The protease-deficient *S. cerevisiae* BJ5465 strain was grown in YPD medium. Minimal expression medium contained 100 mL of 67 g/L sterile yeast nitrogen base, 100 mL of 19.2 g/L sterile yeast synthetic dropout medium supplement without uracil, 100 mL of sterile galactose 20%, 67 mL of KH_2PO_4 1 M pH 6.0 buffer, 1 mL of sterile $CuSO_4$ 1 M, 1 mL of sterile chloramphenicol 25 mg/mL in ethanol, and 631 mL of ddH_2O.

Site-Directed Mutagenesis/IVOE. Mutations N207S and N263D were introduced in the laccase sequence by in vivo overlap extension (IVOE).[47] Four mutagenic primers were designed to obtain overlapping ends: N207SFW (5′-TGCGACCCGTCTACACG-3′), N207SRV (5′-CGTGTAAGACGGGTCGCA-3′), N263DFW (5′-CCTTCCCGACTCCGGGACCA-3′), and N263DRV (5′-TGGTCCCGGAGTCGGGAAGG-3′). In addition, we used primer RMLN sense (5′-CCTCTATACTTTAACGTCAAGG-3′), which binds to nucleotides 160−180 of the plasmid, and primer RMLC antisense (5′-GGGAGGGCGTGAATGTAAGC-3′), which binds to nucleotides 2031−2050 of the plasmid. The 5′-end gene fragment was amplified with RMLN and reverse mutagenic N207SRV primer, the center fragment with N207SFW and N263DRV primers, and the 3′-end fragment with RMLC and forward mutagenic N263DFW primer.

All polymerase chain reactions (PCR) mixtures were prepared in a final volume of 50 μL containing 0.25 μm of each primer, 1 μm of dNTPs, and 2.5 U of *Pfu* ultra DNA polymerase, and 100 ng of pJRαlac[21] was used as a template. Reactions were performed as follows: 95 °C for 2 min (one cycle), 94 °C for 0.5 min, 55 °C for 0.5 min, and 74 °C for 2 min (28 cycles) and 72 °C for 10 min (one cycle).

Purified PCR products were cotransformed in yeast together with the linearized expression vector. Transformed cells were plated in SC dropout plates[46a] and incubated for 2 days at 28 °C.

Laccase Production and Purification. Laccase was produced in 1 L flask cultures, in minimal expression medium, at 30 °C and 200 rpm. Crude extracts were separated by centrifugation at 12000 rpm at 4 °C and then filtered using a 0.45 μm pore size membrane and concentrated by ultrafiltration through 10000 MWCO membranes and dialyzed against 20 mM Tris-HCl buffer, pH 7. Laccases were purified by HPLC (AKTA purifier, GE Healthcare) in two anion-exchange steps and one molecular exclusion step: first, using a HiPrep Q FF 16/10 column and a 100 mL gradient of 0−40% elution buffer (20 mM Tris-HCl + 1 M NaCl, pH 7); second, using a Mono Q HR 5/5 column and a 30 mL gradient of 0−25% elution buffer; finally, using a HiLoad 16/600 Superdex 75 pg column and 20 mM Tris-HCl + 150 mM NaCl, pH 7 (all columns from GE Healthcare). Fractions containing laccase activity were pooled, dialyzed, and concentrated between each chromatographic step.

Laccase Characterization. *Kinetic Constants.* Substrate oxidation was measured by the increment of absorbance at 550 nm for DMPD (ε_{550} = 4134 M^{-1} cm^{-1}) and 410 nm for aniline (ε_{410} = 1167 M^{-1} cm^{-1}) using the plate reader Spectramax Plus (Molecular Devices). Reactions were carried out in triplicate, in 50 mM citrate−phosphate buffer pH 3.0, in 250 μL final volume. Initial rates were represented versus substrate concentration and fitted to a single rectangular hyperbola function in SigmaPlot (version 10.0) software, where the parameter a was equal to k_{cat} and the parameter b was equal to K_M.

Optimum pH. Plates containing 10 μL of laccase samples with 0.1 U/mL activity (measured with 3 mM ABTS) (ε_{418} = 36000 M^{-1} cm^{-1}) and 180 μL of 100 mM Britton and Robinson buffer were prepared at pH values of 2, 3, 4, 5, 6, 7, and 8. The assay commenced when 10 μL of 60 mM ABTS was added to each well to give a final substrate concentration of 3 mM. Activities were measured in triplicate in kinetic mode, and the relative activity was calculated as a percentage of the maximum activity of each variant in the assay.

Stability at pH 3. A 40 μL portion of laccase with 0.1 U/mL activity (measured as mentioned above) was added to 150 μL of citrate−phosphate buffer 50 mM, pH 3. Plates were incubated for 5 h at room temperature. The residual activity was measured in triplicate in kinetic mode in the plate reader, and the relative activity was calculated as a percentage of the initial activity. One activity unit was defined as the amount of enzyme needed to transform 1 μmol of ABTS/min.

Thermostability. T_{50} was evaluated as previously described.[21]

Enzymatic Polymerization of Aniline. Oxidation of 15 mM aniline was carried out in 25 mL of 50 mM citrate−phosphate buffer pH 3, using 5 mM sodium dodecylbenzene-

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

sulfonate (SDBS) as a template. The reaction was performed with 0.1 U/mL laccase activity (measured with 3 mM DMPD pH 3) of crude enzyme, in Pyrex bottles maintaining a 1/1 medium/air ratio (v/v), at room temperature, for 24 h with constant stirring. Increase in 800 nm absorbance was followed in triplicate.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscatal.6b01460.

Average spin densities for unsuccessful mutations, PELE interaction energy profiles for DMPD, and protein sequence alignment for multicopper proteins with available crystal structures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail for S.C.: susanacam@cib.csic.es.

*E-mail for V.G.: victor.guallar@bsc.es.

Author Contributions

⊥These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by the INDOX (KBBE-2013-7-613549) EU-project and the NOESIS (BI0201456388-R) and OxiDesign (CTQ2013−48287-R) Spanish project. G.S. acknowledges an FPI grant from the Spanish Ministry of Competitiveness.

■ ABBREVIATIONS

ABTS, 2,2′-azinobis(3-ethylbenzothiazoline-6-sulfonic acid); ANL, aniline; DAD, donor−acceptor distance; DM, double mutant; DMPD, *N*,*N*-dimethyl-*p*-phenylenediamine; PANI, polyaniline; PELE, protein energy landscape exploration; PPD, *p*-phenylenediamine; QM/MM, quantum mechanics/molecular mechanics; SASA, solvent accessible surface area; SDBS, sodium dodecylbenzenesulfonate

■ REFERENCES

(1) Waites, M. J.; Morgan, N. L.; Rockey, J. S.; Higton, G. *Industrial Microbiology: An Introduction*; Wiley: Hoboken, NJ, 2001.

(2) Kirk, O.; Borchert, T. V.; Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **2002**, *13*, 345−351.

(3) (a) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700−5725. (b) Kries, H.; Blomberg, R.; Hilvert, D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **2013**, *17*, 221−228. (c) Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **2013**, *82*, 447−470. (d) Ye, L. J.; Toh, H. H.; Yang, Y.; Adams, J. P.; Snajdrova, R.; Li, Z. Engineering of Amine Dehydrogenase for Asymmetric Reductive Amination of Ketone by Evolving Rhodococcus Phenylalanine Dehydrogenase. *ACS Catal.* **2015**, *5*, 1119−1122. (e) Verges, A.; Cambon, E.; Barbe, S.; Salamone, S.; Le Guen, Y.; Moulis, C.; Mulard, L. A.; Remaud-Siméon, M.; André, I. Computer-Aided Engineering of a Transglycosylase for the Glucosylation of an Unnatural Disaccharide of Relevance for Bacterial Antigen Synthesis. *ACS Catal.* **2015**, *5*, 1186−1198. (f) Mellot-Draznieks, C.; Valayannopoulos, V.; Chrétien, D.; Munnich, A.; de Lonlay, P.; Toulhoat, H. Relative Enzymatic Activity Levels from In Silico Mutagenesis. *ACS Catal.* **2012**, *2*, 2673−2686. (g) Bloom, J. D.; Arnold, F. H. In the light of directed evolution: Pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9995−10000.

(4) (a) Mayer, A. M.; Staples, R. C. Laccase: new functions for an old enzyme. *Phytochemistry* **2002**, *60*, 551−565. (b) Minussi, R. C.; Pastore, G. M.; Durán, N. Potential applications of laccase in the food industry. *Trends Food Sci. Technol.* **2002**, *13*, 205−216. (c) Osma, J. F.; Toca-Herrera, J. L.; Rodríguez-Couto, S. Uses of Laccases in the Food Industry. *Enzyme Res.* **2010**, *2010*, 1−8. (d) Riva, S. Laccases: blue enzymes for Green Chem. *Trends Biotechnol.* **2006**, *24*, 219−226. (e) Cañas, A. I.; Camarero, S. Laccases and their natural mediators: Biotechnological tools for sustainable eco-friendly processes. *Biotechnol. Adv.* **2010**, *28*, 694−705. (f) Witayakran, S.; Ragauskas, A. J. Synthetic Applications of Laccase in Green Chem. *Adv. Synth. Catal.* **2009**, *351*, 1187−1209.

(5) (a) Martínez, Á., High Redox Potential Peroxidases. In *Industrial Enzymes*; Polaina, J., MacCabe, A., Eds.; Springer: Amsterdam, 2007; pp 477−488. (b) Ayala, M., Redox Potential of Peroxidases. In *Biocatalysis Based on Heme Peroxidases*; Torres, E., Ayala, M., Eds.; Springer: Berlin, Heidelberg, 2010; pp 61−77.

(6) (a) Bourbonnais, R.; Paice, M. G. Oxidation of non-phenolic substrates. *FEBS Lett.* **1990**, *267*, 99−102. (b) Camarero, S.; Ibarra, D.; Martínez, Á. T.; Romero, J.; Gutiérrez, A.; del Río, J. C. Paper pulp delignification using laccase and natural mediators. *Enzyme Microb. Technol.* **2007**, *40*, 1264−1271.

(7) (a) Li, K.; Xu, F.; Eriksson, K.-E. L. Comparison of Fungal Laccases and Redox Mediators in Oxidation of a Nonphenolic Lignin Model Compound. *Appl. Environ. Microbiol.* **1999**, *65*, 2654−2660. (b) Morozova, O. V.; Shumakovich, G. P.; Shleev, S. V.; Yaropolov, Y. I. Laccase-mediator systems and their applications: A review. *Appl. Biochem. Microbiol.* **2007**, *43*, 523−535. (c) Tadesse, M. A.; D'Annibale, A.; Galli, C.; Gentili, P.; Sergi, F. An assessment of the relative contributions of redox and steric issues to laccase specificity towards putative substrates. *Org. Biomol. Chem.* **2008**, *6*, 868−878. (d) Xu, F. Oxidation of Phenols, Anilines, and Benzenethiols by Fungal Laccases: Correlation between Activity and Redox Potentials as Well as Halide Inhibition†. *Biochemistry* **1996**, *35*, 7608−7614. (e) Xu, F.; Kulys, J. J.; Duke, K.; Li, K.; Krikstopaitis, K.; Deussen, H.-J. W.; Abbate, E.; Galinyte, V.; Schneider, P. *Appl. Environ. Microbiol.* **2000**, *66*, 2052−2056.

(8) Monza, E.; Lucas, M. F.; Camarero, S.; Alejaldre, L. C.; Martínez, A. T.; Guallar, V. Insights into Laccase Engineering from Molecular Simulations: Toward a Binding-Focused Strategy. *J. Phys. Chem. Lett.* **2015**, *6*, 1447−1453.

(9) (a) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **2005**, *1*, 1304. (b) Cossins, B. P.; Hosseini, A.; Guallar, V. Exploration of Protein Conformational Change with PELE and Meta-Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 959−965.

(10) (a) Friesner, R. A.; Guallar, V. Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427. (b) Gao, J.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

(11) (a) Babot, E. D.; del Río, J. C.; Cañellas, M.; Sancho, F.; Lucas, F.; Guallar, V.; Kalum, L.; Lund, H.; Gröbe, G.; Scheibner, K.; Ullrich, R.; Hofrichter, M.; Martínez, A. T.; Gutiérrez, A. Steroid hydroxylation by basidiomycete peroxygenases: A combined experimental and computational study. *Appl. Environ. Microbiol.* **2015**, *81*, 4130−4142. (b) Lucas, M. F.; Guallar, V. An Atomistic View on Human Hemoglobin Carbon Monoxide Migration Processes. *Biophys. J.* **2012**, *102*, 887−896. (c) Linde, D.; Pogni, R.; Cañellas, M.; Lucas, F.; Guallar, V.; Baratto; Maria, C.; Sinicropi, A.; Sáez-Jiménez, V.; Coscolín, C.; Romero, A.; Medrano, F. J.; Ruiz-Dueñas, F. J.; Martínez, A. T. Catalytic surface radical in dye-decolorizing peroxidase: a computational, spectroscopic and site-directed mutagenesis study. *Biochem. J.* **2015**, *466*, 253−262. (d) Hosseini, A.; Espona-Fiedler, M.;

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

Soto-Cerrato, V.; Quesada, R.; Pérez-Tomás, R.; Guallar, V. Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members. *PLoS One* **2013**, *8*, e57562.
(12) (a) Winkler, J. R.; Gray, H. B. Electron Flow through Metalloproteins. *Chem. Rev.* **2014**, *114*, 3369−3380. (b) Christensen, N. J.; Kepp, K. P. Setting the stage for electron transfer: Molecular basis of ABTS-binding to four laccases from Trametes versicolor at variable pH and protein oxidation state. *J. Mol. Catal. B: Enzym.* **2014**, *100*, 68−77.
(13) Marcus, R. A. Electron transfer reactions in chemistry. Theory and experiment. *Rev. Mod. Phys.* **1993**, *65*, 599−610.
(14) Senn, H.; Thiel, W. QM/MM Methods for Biological Systems. In *Atomistic Approaches in Modern Biology*; Reiher, M., Ed.; Springer: Berlin, Heidelberg, 2007; Vol. *268*, pp 173−290.
(15) Morris, P.; Travis, A. History of the International Dyestuff Industry. *Am. Dyest. Rep.* **1992**, *81*, 59−59.
(16) (a) Bhadra, S.; Khastgir, D.; Singha, N. K.; Lee, J. H. Progress in preparation, processing and applications of polyaniline. *Prog. Polym. Sci.* **2009**, *34*, 783−810. (b) Ćirić-Marjanović, G. Recent advances in polyaniline research: Polymerization mechanisms, structural aspects, properties and applications. *Synth. Met.* **2013**, *177*, 1−47.
(17) Sapurina, I. Y.; Stejskal, J. Oxidation of aniline with strong and weak oxidants. *Russ. J. Gen. Chem.* **2012**, *82*, 256−275.
(18) (a) Kobayashi, S.; Makino, A. Enzymatic Polymer Synthesis: An Opportunity for Green Polymer Chemistry. *Chem. Rev.* **2009**, *109*, 5288−5353. (b) Hollmann, F.; Arends, I. W. C. E.; Buehler, K.; Schallmey, A.; Buhler, B. Enzyme-mediated oxidations for the chemist. *Green Chem.* **2011**, *13*, 226−265. (c) Gross, R. A.; Kumar, A.; Kalra, B. Polymer Synthesis by In Vitro Enzyme Catalysis. *Chem. Rev.* **2001**, *101*, 2097−2124. (d) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185−194.
(19) Zhang, J.; Zou, F.; Yu, X.; Huang, X.; Qu, Y. Ionic liquid improves the laccase-catalyzed synthesis of water-soluble conducting polyaniline. *Colloid Polym. Sci.* **2014**, *292*, 2549−2554.
(20) Yang, L. Y. O.; Chang, C.; Liu, S.; Wu, C.; Yau, S. L. Direct Visualization of an Aniline Admolecule and Its Electropolymerization on Au(111) with in Situ Scanning Tunneling Microscope. *J. Am. Chem. Soc.* **2007**, *129*, 8076−8077.
(21) Pardo, I.; Vicente, A. I.; Mate, D. M.; Alcalde, M.; Camarero, S. Development of chimeric laccases by directed evolution. *Biotechnol. Bioeng.* **2012**, *109*, 2978−2986.
(22) (a) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3790−3795. (b) Frushicheva, M. P.; Cao, J.; Chu, Z. T.; Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16869−16874. (c) Chen, C.-Y.; Georgiev, I.; Anderson, A. C.; Donald, B. R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3764−3769.
(23) (a) Nematollahi, D.; Maleki, A. Electrochemical oxidaton of N,N-dialkyl-p-phenylenediamines in the presence of arylsulfinic acids. An efficient method for the synthesis of new sulfonamide derivatives. *Electrochem. Commun.* **2009**, *11*, 488−491. (b) Meissner, P.; Mandi, G.; Coulibaly, B.; Witte, S.; Tapsoba, T.; Mansmann, U.; Rengelshausen, J.; Schiek, W.; Jahn, A.; Walter-Sack, I.; Mikus, G.; Burhenne, J.; Riedel, K.-D.; Schirmer, R. H.; Kouyate, B.; Muller, O. Methylene blue for malaria in Africa: results from a dose-finding study in combination with chloroquine. *Malar. J.* **2006**, *5*, 84. (c) Schirmer, R. H.; Coulibaly, B.; Stich, A.; Scheiwein, M.; Merkle, H.; Eubel, J.; Becker, K.; Becher, H.; Müller, O.; Zich, T.; Schiek, W.; Kouyaté, B. Methylene blue as an antimalarial agent. *Redox Rep.* **2003**, *8*, 272−275. (d) Coulibaly, B.; Zoungrana, A.; Mockenhaupt, F. P.; Schirmer, R. H.; Klose, C.; Mansmann, U.; Meissner, P. E.; Müller, O. Strong Gametocytocidal Effect of Methylene Blue-Based Combination Therapy against Falciparum Malaria: A Randomised Controlled Trial. *PLoS One* **2009**, *4*, e5318.
(24) (a) Karplus, M.; Lavery, R. Significance of Molecular Dynamics Simulations for Life Sciences. *Isr. J. Chem.* **2014**, *54*, 1042. (b) Lucas, M. F.; Cabeza de Vaca, I.; Takahashi, R.; Rubio-Martínez, J.; Guallar, V. Atomic Level Rendering of DNA-Drug Encounter. *Biophys. J.* **2014**, *106*, 421−429.
(25) (a) Pardo, I.; Santiago, G.; Gentili, P.; Lucas, F.; Monza, E.; Medrano, F. J.; Galli, C.; Martinez Ferrer, A. T.; Guallar, V.; Camarero, S. RE-DESIGNING THE SUBSTRATE BINDING POCKET OF LACCASE FOR ENHANCED OXIDATION OF SINAPIC ACID. *Catal. Sci. Technol.* **2016**, *6*, 3900. (b) Toscano, M. D.; De Maria, L.; Lobedanz, S.; Østergaard, L. H. Optimization of a Small Laccase by Active-Site Redesign. *ChemBioChem* **2013**, *14*, 1209−1211.
(26) (a) Xu, F.; Berka, R. M.; Wahleithner, J. A.; Nelson, B. A.; Shuster, J. R.; Brown, S. H.; Palmer, A. E.; Solomon, E. I. Site-directed mutations in fungal laccase: effect on redox potential, activity and pH profile. *Biochem. J.* **1998**, *334*, 63−70. (b) Xu, F.; Palmer, A. E.; Yaver, D. S.; Berka, R. M.; Gambetta, G. A.; Brown, S. H.; Solomon, E. I. Targeted Mutations in a Trametes villosa Laccase: axial perturbations of the T1 copper. *J. Biol. Chem.* **1999**, *274*, 12372−12375. (c) Cambria, M. T.; Gullotto, D.; Garavaglia, S.; Cambria, A. In silico study of structural determinants modulating the redox potential of Rigidoporus lignosus and other fungal laccases. *J. Biomol. Struct. Dyn.* **2012**, *30*, 89−101.
(27) (a) Koschorreck, K.; Richter, S. M.; Swierczek, A.; Beifuss, U.; Schmid, R. D.; Urlacher, V. B. Comparative characterization of four laccases from Trametes versicolor concerning phenolic C−C coupling and oxidation of PAHs. *Arch. Biochem. Biophys.* **2008**, *474*, 213−219. (b) Galli, C.; Gentili, P.; Jolivalt, C.; Madzak, C.; Vadalà, R. How is the reactivity of laccase affected by single-point mutations? Engineering laccase for improved activity towards sterically demanding substrates. *Appl. Microbiol. Biotechnol.* **2011**, *91*, 123−131. (c) Pardo, I.; Santiago, G.; Gentili, P.; Lucas, F.; Monza, E.; Medrano, F. J.; Galli, C.; Martinez, A. T.; Guallar, V.; Camarero, S. Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. *Catal. Sci. Technol.* **2016**, *6*, 3900.
(28) Madzak, C.; Mimmi, M. C.; Caminade, E.; Brault, A.; Baumberger, S.; Briozzo, P.; Mougin, C.; Jolivalt, C. Shifting the optimal pH of activity for a laccase from the fungus Trametes versicolor by structure-based mutagenesis. *Protein Eng., Des. Sel.* **2005**, *19*, 77−84.
(29) Taylor, A. B.; Stoj, C. S.; Ziegler, L.; Kosman, D.; Hart, P. J. The copper-iron connection in biology: structure of the metallo-oxidase Fet3p. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15459−15464.
(30) Galli, C.; Madzak, C.; Vadalà, R.; Jolivalt, C.; Gentili, P. Concerted Electron/Proton Transfer Mechanism in the Oxidation of Phenols by Laccase. *ChemBioChem* **2013**, *14*, 2500−2505.
(31) Méndez-Hernández, D. D.; Tarakeshwar, P.; Gust, D.; Moore, T. A.; Moore, A. L.; Mujica, V. Simple and accurate correlation of experimental redox potentials and DFT-calculated HOMO/LUMO energies of polycyclic aromatic hydrocarbons. *J. Mol. Model.* **2013**, *19*, 2845−2848.
(32) Junker, K.; Gitsov, I.; Quade, N.; Walde, P. Preparation of aqueous polyaniline-vesicle suspensions with class III peroxidases. Comparison between horseradish peroxidase isoenzyme C and soybean peroxidase. *Chem. Pap.* **2013**, *67*, 1028−1047.
(33) Toscano, M. D.; De Maria, L.; Lobedanz, S.; Ostergaard, L. H. Optimization of a Small Laccase by Active-Site Redesign. *ChemBioChem* **2013**, *14*, 1209−1211.
(34) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.
(35) Olsson, M.; Sondergaard, C.; Rostkowski, M.; Jensen, J. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(36) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537−W541.

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

(37) Zhao, Y.; Truhlar, D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(38) Madadkar-Sobhani, A.; Guallar, V. PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic Acids Res.* **2013**, *41*, W322−W328.
(39) 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* **2001**, *105*, 6474−6487.
(40) Bashford, D.; Case, D. A. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(41) Bortolotti, C. A.; Siwko, M. E.; Castellini, E.; Ranieri, A.; Sola, M.; Corni, S. The Reorganization Energy in Cytochrome c is Controlled by the Accessibility of the Heme to the Solvent. *J. Phys. Chem. Lett.* **2011**, *2*, 1761−1765.
(42) Murphy, R.; Philipp, D.; Friesner, R. A mixed quanum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* **2000**, *21*, 1442−1457.
(43) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157−167.
(44) Mulliken, R. S. Electronic Population Analysis on LCAO−MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(45) (a) Artz, K.; Williams, J. C.; Allen, J. P.; Lendzian, F.; Rautter, J.; Lubitz, W. Relationship between the oxidation potential and electron spin density of the primary electron donor in reaction centers from Rhodobacter sphaeroides. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 13582−13587. (b) Meisel, D.; Neta, P. One-electron redox potentials of nitro compounds and radiosensitizers. Correlation with spin densities of their radical anions. *J. Am. Chem. Soc.* **1975**, *97*, 5198−5203.
(46) (a) Camarero, S.; Pardo, I.; Cañas, A. I.; Molina, P.; Record, E.; Martínez, A. T.; Martínez, M. J.; Alcalde, M. Engineering Platforms for Directed Evolution of Laccase from Pycnoporus cinnabarinus. *Appl. Environ. Microbiol.* **2012**, *78*, 1370−1384. (b) Maté, D.; García-Burgos, C.; García-Ruiz, E.; Ballesteros, A. O.; Camarero, S. Alcalde, M., Laboratory Evolution of High-Redox Potential Laccases. *Chem. Biol.* **2010**, *17*, 1030−1041.
(47) Alcalde, M., Mutagenesis Protocols in Saccharomyces cerevisiae by In Vivo Overlap Extension. In *In Vitro Mutagenesis Protocols*; Braman, J., Ed.; Humana Press: New York, 2010; Vol. *634*, pp 3−14.

DOI: 10.1021/acscatal.6b01460
ACS Catal. 2016, 6, 5415−5423

Supporting information.

Computer-aided laccase engineering: toward biological oxidation of arylamines

Gerard Santiago[†‡], Felipe de Salas[§‡], M. Fátima Lucas[†‖‡], Emanuele Monza[†], Sandra Acebes[†], Ángel T. Martinez[§], Susana Camarero[§*], Victor Guallar[†¥*]

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

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

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

[¥] ICREA, Passeig Lluis Companys 23, E-08010 Barcelona, Spain.

Corresponding Author

* victor.guallar@bsc.es
* susanacam@cib.csic.es

This supporting document contains the following information:

1. Average spin densities for unsuccessful mutations
2. PELE interaction energy profiles for DMPD
3. Estimation of the effect of a negative charge in aniline's oxidation
4. Protein sequence alignment for multi-copper proteins with available crystal structures

1. Average spin densities for unsuccessful mutations

Table S1. Extra single point mutations tested in this work.

Single-point mutations			
Position	Original residue	Mutation	Spin density average (%)
205	D	E	≈ 0
		A	≈ 0
		Y	≈ 0
207	N	D	8
		S	3
263	N	D	10
		S	1

2. PELE interaction energy profiles for DMPD

Since DMPD can co-exist in two protonation states at pH 3: charged on the primary or the tertiary amine, PELE simulations were performed with both molecules.

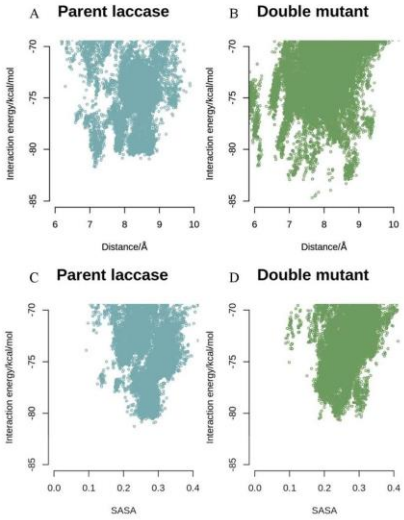


Figure S1. Interaction energies vs. distance between the center of mass of DMPD (protonated on the **primary amine**) and the copper T1 atom for: A) parent laccase and B) DM. Interaction energy vs. SASA for: C) parent laccase and D) DM.

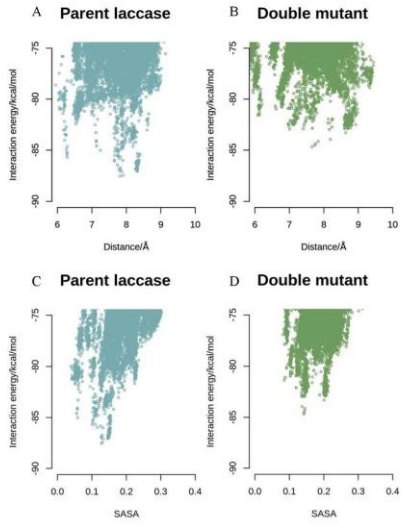


Figure S2. Interaction energies vs. distance between the center of mass of DMPD (protonated on the **tertiary amine**) and the copper T1 atom for: A) parent laccase and B) DM. Interaction energy vs. SASA for: C) parent laccase and D) DM.

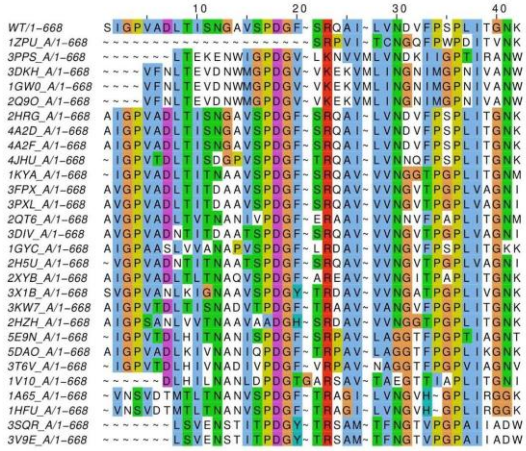
3. Estimation of the effect of a negative charge in aniline's oxidation

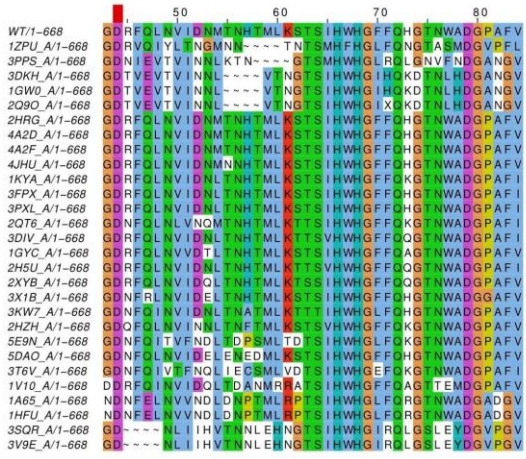
Single point energy calculations were performed using Jaguar software, with HF/6-311G-3DF-3PD++ for all systems. The model systems include ANL and ANL together with the side chain of D263. Energies were computed for the oxidized and non-oxidized states for both systems and energy differences indicates that it is easier to abstract an electron from aniline (oxidize) in the presence of a nearby negative charge. All energies are in hartrees.

Table S2. Negative charge addition effect on highest occupied molecular orbitals.

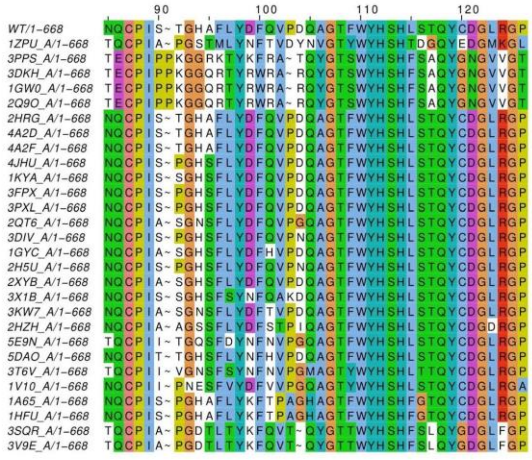
	Oxidized	Non-oxidized	DE
ANL	-285.7117	-286.1773	0.47
ANL + acetic acid	-552.4824	-552.7302	0.25

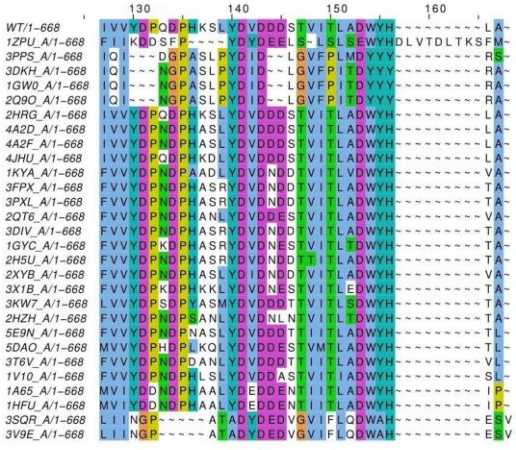
4. Protein sequence alignment



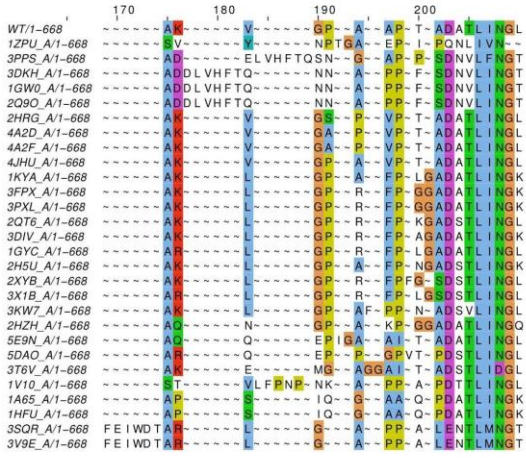


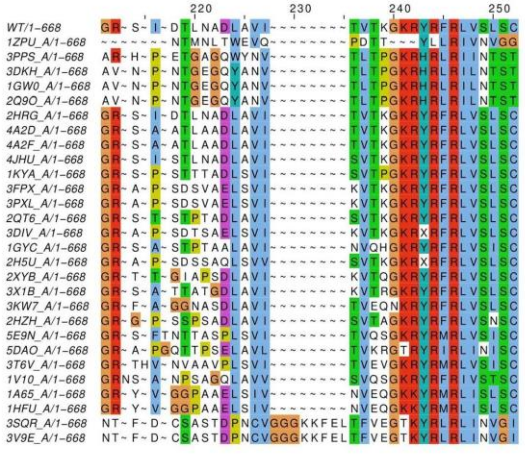
5


90
100
110
120
WT/1–668 NQCPIS~TGHAFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
1ZPU_A/1–668 TQCPIA~PGSTMLYNFTVDYNVGTYWYHSHTDGQYEDGMKGL
3PPS_A/1–668 TECPIPPKGGRKTYKFRA~TQYGTSWYHSHFSAQYGNGVVGT
3DKH_A/1–668 TECPIPPKGGQRTYRWRA~RQYGTSWYHSHFSAQYGNGVVGT
1GW0_A/1–668 TECPIPPKGGQRTYRWRA~RQYGTSWYHSHFSAQYGNGVVGT
2Q9O_A/1–668 TECPIPPKGGQRTYRWRA~RQYGTSWYHSHFSAQYGNGVVGT
2HRG_A/1–668 NQCPIS~TGHAFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
4A2D_A/1–668 NQCPIS~TGHAFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
4A2F_A/1–668 NQCPIS~TGHAFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
4JHU_A/1–668 NQCPIS~PGHSFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
1KYA_A/1–668 NQCPIS~SGHSFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
3FPX_A/1–668 NQCPIS~PGHSFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
3PXL_A/1–668 NQCPIS~PGHSFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
2QT6_A/1–668 NQCPIA~SGNSFLYDFQVPGQAGTFWYHSHLSTQYCDGLRGP
3DIV_A/1–668 NQCPIS~PGHSFLYDFQVPNQAGTFWYHSHLSTQYCDGLRGP
1GYC_A/1–668 NQCPIA~SGHSFLYDFHVPDQAGTFWYHSHLSTQYCDGLRGP
2H5U_A/1–668 NQCPIS~PGHSFLYDFQVPNQAGTFWYHSHLSTQYCDGLRGP
2XYB_A/1–668 NQCPIA~SGHSFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGP
3X1B_A/1–668 NQCPIS~SGHSFSYNFQAKDQAGTFWYHSHLSTQYCDGLRGP
3KW7_A/1–668 NQCPIA~SGNSFLYDFTVPDQAGTFWYHSHLSTQYCDGLRGP
2HZH_A/1–668 NQCPIA~AGSSFLYDFSTPIQAGTFWYHSHLSTQYCDGDRGP
5E9N_A/1–668 TQCPII~TGQSFDYNFNVPGQAGTFWYHSHLSTQYCDGLRGP
5DAO_A/1–668 NQCPIT~TGHSFLYNFHVPDQAGTFWYHSHLSTQYCDGLRGP
3T6V_A/1–668 TQCPII~VGNSFSYNFNVPGMAGTYWYHSHLTTQYCDGLRGP
1V10_A/1–668 NQCPII~PNESFVYDFVVPGQAGTYWYHSHLSTQYCDGLRGA
1A65_A/1–668 NQCPIS~PGHAFLYKFTPAGHAGTFWYHSHFGTQYCDGLRGP
1HFU_A/1–668 NQCPIS~PGHAFLYKFTPAGHAGTFWYHSHFGTQYCDGLRGP
3SQR_A/1–668 TQCPIA~PGDTLTYKFQVT~QYGTTWYHSHFSLQYGDGLFGP
3V9E_A/1–668 TQCPIA~PGDTLTYKFQVT~QYGTTWYHSHFSLQYGDGLFGP

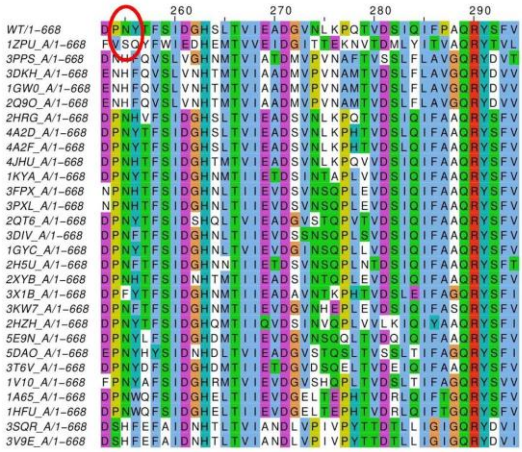

130
140
150
160
WT/1–668 IVVYDPQDPHKSLYDVDDDSTVITLADWYH~~~~~~~~LA~
1ZPU_A/1–668 FIIKDDSFP~~~~YDYDEELS~LSLSEWYHDLVTDLTKSFM~
3PPS_A/1–668 IQI~~~DGPASLPYDID~~LGVFPLMDYYY~~~~~~~~RS~
3DKH_A/1–668 IQI~~~NGPASLPYDID~~LGVFPITDYYY~~~~~~~~RA~
1GW0_A/1–668 IQI~~~NGPASLPYDID~~LGVFPITDYYY~~~~~~~~RA~
2Q9O_A/1–668 IQI~~~NGPASLPYDID~~LGVFPITDYYY~~~~~~~~RA~
2HRG_A/1–668 IVVYDPQDPHKSLYDVDDDSTVITLADWYH~~~~~~~~LA~
4A2D_A/1–668 IVVYDPNDPHASLYDVDDDSTVITLADWYH~~~~~~~~LA~
4A2F_A/1–668 IVVYDPNDPHASLYDVDDDSTVITLADWYH~~~~~~~~LA~
4JHU_A/1–668 IVVYDPQDPHKDLYDVDDDSTVITLADWYH~~~~~~~~LA~
1KYA_A/1–668 FVVYDPNDPAADLYDVDNDDTVITLVDWYH~~~~~~~~VA~
3FPX_A/1–668 FVVYDPNDPHASRYDVDNDDTVITLADWYH~~~~~~~~TA~
3PXL_A/1–668 FVVYDPNDPHASRYDVDNDDTVITLADWYH~~~~~~~~TA~
2QT6_A/1–668 FVVYDPNDPHANLYDVDDESTVITLADWYH~~~~~~~~VA~
3DIV_A/1–668 FVVYDPNDPHASRYDVDNDDTVITLADWYH~~~~~~~~TA~
1GYC_A/1–668 FVVYDPKDPHASRYDVDNESTVITLTDWYH~~~~~~~~TA~
2H5U_A/1–668 FVVYDPNDPHASRYDVDNDDTTITLADWYH~~~~~~~~TA~
2XYB_A/1–668 FVVYDPNDPHASLYDIDNDDTVITLADWYH~~~~~~~~VA~
3X1B_A/1–668 FVVYDPKDPHKKLYDVDNESTVITLEDWYH~~~~~~~~TA~
3KW7_A/1–668 LVVYDPSDPYASMYDVDDDTTVITLSDWYH~~~~~~~~TA~
2HZH_A/1–668 FVVYDPNDPSANLYDVDNLNTVITLTDWYH~~~~~~~~TA~
5E9N_A/1–668 FVVYDPNDPNASLYDVDDDTTIITLADWYH~~~~~~~~TL~
5DAO_A/1–668 MVVYDPHDPLKQLYDVDDESTVMTLADWYH~~~~~~~~TL~
3T6V_A/1–668 FVVYDPNDPDANLYDVDDDTTIITLADWYH~~~~~~~~VL~
1V10_A/1–668 FVVYDPNDPHLSLYDVDDASTVITIADWYH~~~~~~~~SL~
1A65_A/1–668 MVIYDDNDPHAALYDEDDENTIITLADWYH~~~~~~~~IP~
1HFU_A/1–668 MVIYDDNDPHAALYDEDDENTIITLADWYH~~~~~~~~IP~
3SQR_A/1–668 LIINGP~~~~~ATADYDEDVGVIFLQDWAH~~~~~~~~ESV
3V9E_A/1–668 LIINGP~~~~~ATADYDEDVGVIFLQDWAH~~~~~~~~ESV

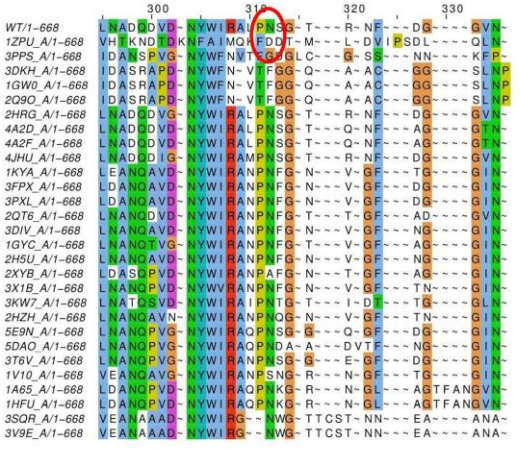
6


170
180
190
200
WT/1–668
1ZPU_A/1–668
3PPS_A/1–668
3DKH_A/1–668
1GW0_A/1–668
2Q9O_A/1–668
2HRG_A/1–668
4A2D_A/1–668
4A2F_A/1–668
4JHU_A/1–668
1KYA_A/1–668
3FPX_A/1–668
3PXL_A/1–668
2QT6_A/1–668
3DIV_A/1–668
1GYC_A/1–668
2H5U_A/1–668
2XYB_A/1–668
3X1B_A/1–668
3KW7_A/1–668
2HZH_A/1–668
5E9N_A/1–668
5DAO_A/1–668
3T6V_A/1–668
1V10_A/1–668
1A65_A/1–668
1HFU_A/1–668
3SQR_A/1–668
3V9E_A/1–668

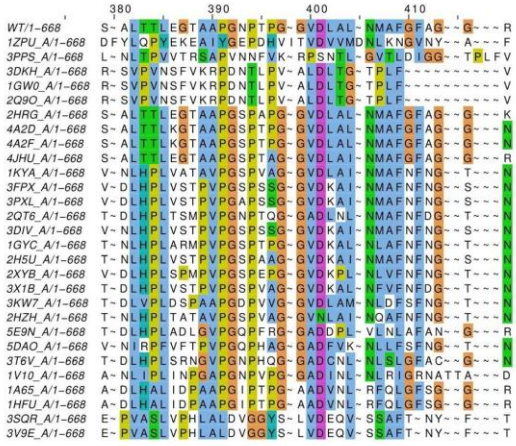

220
230
240
250
WT/1–668
1ZPU_A/1–668
3PPS_A/1–668
3DKH_A/1–668
1GW0_A/1–668
2Q9O_A/1–668
2HRG_A/1–668
4A2D_A/1–668
4A2F_A/1–668
4JHU_A/1–668
1KYA_A/1–668
3FPX_A/1–668
3PXL_A/1–668
2QT6_A/1–668
3DIV_A/1–668
1GYC_A/1–668
2H5U_A/1–668
2XYB_A/1–668
3X1B_A/1–668
3KW7_A/1–668
2HZH_A/1–668
5E9N_A/1–668
5DAO_A/1–668
3T6V_A/1–668
1V10_A/1–668
1A65_A/1–668
1HFU_A/1–668
3SQR_A/1–668
3V9E_A/1–668

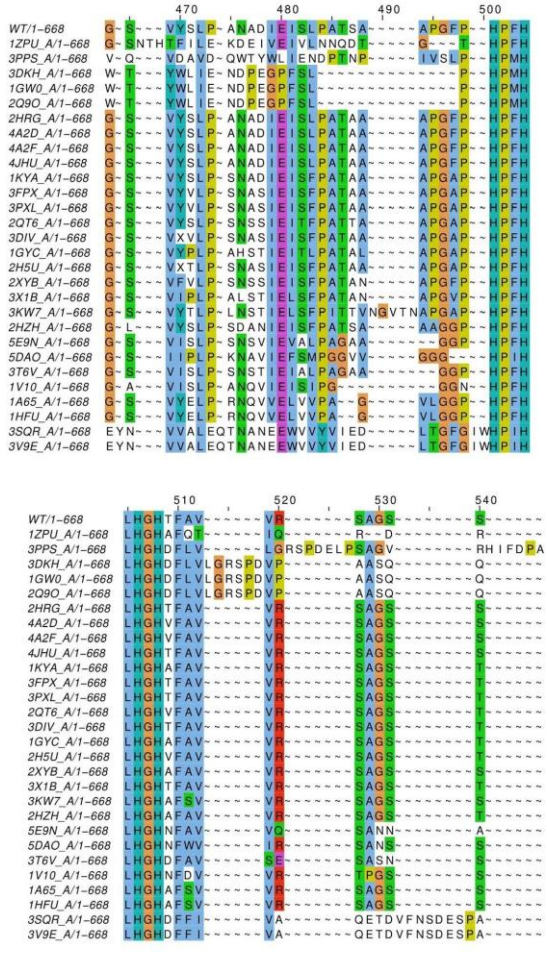
7


260
270
280
290
WT/1-668
1ZPU_A/1-668
3PPS_A/1-668
3DKH_A/1-668
1GW0_A/1-668
2Q9O_A/1-668
2HRG_A/1-668
4A2D_A/1-668
4A2F_A/1-668
4JHU_A/1-668
1KYA_A/1-668
3FPX_A/1-668
3PXL_A/1-668
2QT6_A/1-668
3DIV_A/1-668
1GYC_A/1-668
2H5U_A/1-668
2XYB_A/1-668
3X1B_A/1-668
3KW7_A/1-668
2HZH_A/1-668
5E9N_A/1-668
5DAO_A/1-668
3T6V_A/1-668
1V10_A/1-668
1A65_A/1-668
1HFU_A/1-668
3SQR_A/1-668
3V9E_A/1-668

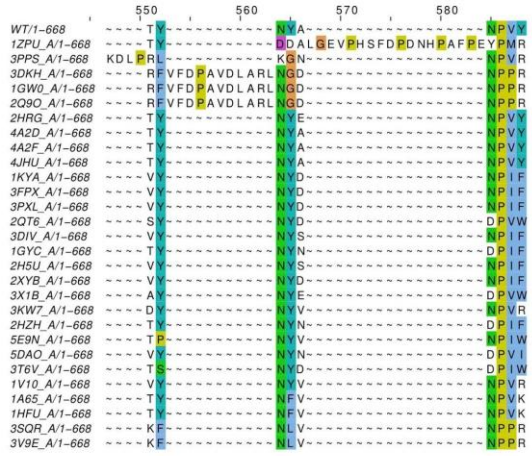

300
310
320
330
WT/1-668
1ZPU_A/1-668
3PPS_A/1-668
3DKH_A/1-668
1GW0_A/1-668
2Q9O_A/1-668
2HRG_A/1-668
4A2D_A/1-668
4A2F_A/1-668
4JHU_A/1-668
1KYA_A/1-668
3FPX_A/1-668
3PXL_A/1-668
2QT6_A/1-668
3DIV_A/1-668
1GYC_A/1-668
2H5U_A/1-668
2XYB_A/1-668
3X1B_A/1-668
3KW7_A/1-668
2HZH_A/1-668
5E9N_A/1-668
5DAO_A/1-668
3T6V_A/1-668
1V10_A/1-668
1A65_A/1-668
1HFU_A/1-668
3SQR_A/1-668
3V9E_A/1-668

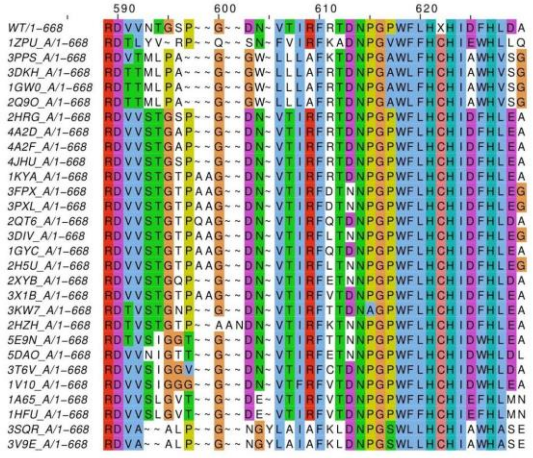
8

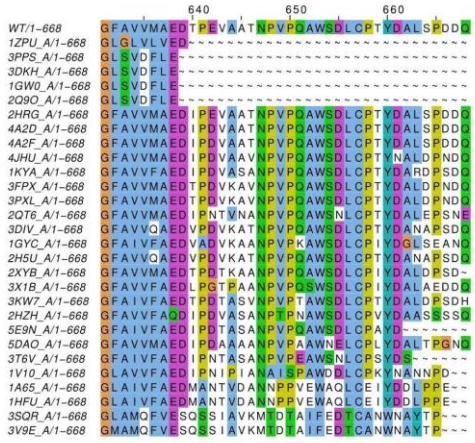

380
390
400
410
WT/1-668 S~ALTTLEGTAAPGNPTPG~GVDLAL~NMAFGFAG~~G~~~R
1ZPU_A/1-668 DFYLQPYEKEAIYGEPDHVITVDVVMDNLKNGVNY~~A~~~F
3PPS_A/1-668 L~NLTPVVTRSAPVNNFVK~RPSNTL~GVTLDIGG~~TPLFV
3DKH_A/1-668 R~SVPVNSFVKRPDNTLPV~ALDLTG~TPLF~~~~~~~~~V
1GW0_A/1-668 R~SVPVNSFVKRPDNTLPV~ALDLTG~TPLF~~~~~~~~~V
2Q9O_A/1-668 R~SVPVNSFVKRPDNTLPV~ALDLTG~TPLF~~~~~~~~~V
2HRG_A/1-668 S~ALTTLEGTAAPGSPAPG~GVDLAL~NMAFGFAG~~G~~~K
4A2D_A/1-668 S~ALTTLKGTAAPGSPTPG~GVDLAL~NMAFGFAG~~G~~~N
4A2F_A/1-668 S~ALTTLKGTAAPGSPTPG~GVDLAL~NMAFGFAG~~G~~~N
4JHU_A/1-668 S~ALTTLEGTAAPGSPTAG~GVDLAI~NMAFGFAG~~G~~~R
1KYA_A/1-668 V~NLHPLVATAVPGSPVAG~GVDLAI~NMAFNFNG~~T~~~N
3FPX_A/1-668 V~DLHPLVSTPVPGSPSSG~GVDKAI~NMAFNFNG~~S~~~N
3PXL_A/1-668 V~DLHPLVSTPVPGAPSSG~GVDKAI~NMAFNFNG~~S~~~N
2QT6_A/1-668 T~DLHPLTSMPVPGNPTQG~GADLNL~NMAFNFDG~~T~~~N
3DIV_A/1-668 V~NLHPLVSTPVPGSPSSG~GVDKAI~NMAFNFNG~~S~~~N
1GYC_A/1-668 T~NLHPLARMPVPGSPTPG~GVDKAL~NLAFNFNG~~T~~~N
2H5U_A/1-668 T~NLHPLVSTPVPGSPAAG~GVDKAI~NMAFNFNG~~S~~~N
2XYB_A/1-668 V~DLHPLSPMPVPGSPEPG~GVDKPL~NLVFNFNG~~T~~~N
3X1B_A/1-668 T~DLHPLVSTPVPGSPVAG~GVDKAL~NFVFNFDG~~T~~~N
3KW7_A/1-668 T~DLVPLDSPAAPGDPVVG~GVDLAM~NLDFSFNG~~T~~~N
2HZH_A/1-668 T~NLHPLTATAVPGSPVAG~GVNLAI~NQAFNFNG~~T~~~N
5E9N_A/1-668 T~DLHPLADLGVPGQPFRG~GADDPL~VLNLAFAN~~G~~~R
5DAO_A/1-668 V~NIRPFVFTPVPGQPHAG~GADFVK~NLLFSFNG~~T~~~N
3T6V_A/1-668 T~DLHPLSRNGVPGNPHQG~GADCNL~NLSLGFAC~~G~~~N
1V10_A/1-668 A~NLIPLINPGAPGNPVPG~GADINL~NLRIGRNATTA~~~D
1A65_A/1-668 A~DLHALIDPAAPGIPTPG~AADVNL~RFQLGFSG~~G~~~R
1HFU_A/1-668 A~DLHALIDPAAPGIPTPG~AADVNL~RFQLGFSG~~G~~~R
3SQR_A/1-668 E~PVASLVPHLALDVGGYS~LVDEQV~SSAFT~NY~~F~~~T
3V9E_A/1-668 E~PVASLVPHLALDVGGYS~LVDEQV~SSAFT~NY~~F~~~T
430
440
450
460
WT/1-668 FTIN~GASFTPP~TVPVLLQILSGAQS~~~~~~~~AQDLLPS
1ZPU_A/1-668 F~~N~NITYTAP~KVPTLMTVLSSGDQ~~~~~~~~ANNSEIY
3PPS_A/1-668 WKVN~GSAINVDWGKPILDYVMSGNTS~~~~~~~~VPVSDNI
3DKH_A/1-668 WKVN~GSDINVDWGKPIIDYILTGNTSYPVSDNIVQVDAVDQ
1GW0_A/1-668 WKVN~GSDINVDWGKPIIDYILTGNTSYPVSDNIVQVDAVDQ
2Q9O_A/1-668 WKVN~GSDINVDWGKPIIDYILTGNTSYPVSDNIVQVDAVDQ
2HRG_A/1-668 FTIN~GASFTPP~TVPVLLQILSGAQS~~~~~~~~AQDLLPS
4A2D_A/1-668 FTIN~GASFTPP~TVPVLLQILSGAQS~~~~~~~~AADLLPA
4A2F_A/1-668 FTIN~GASFTPP~TVPVLLQILSGAQS~~~~~~~~AADLLPA
4JHU_A/1-668 FTIN~GASFTPP~TVPVLLQILSGAQN~~~~~~~~AQDLLPT
1KYA_A/1-668 FFIN~GASFTPP~TVPVLLQIISGAQN~~~~~~~~AQDLLPS
3FPX_A/1-668 FFIN~GASFVPP~TVPVLLQILSGAQT~~~~~~~~AQDLLPS
3PXL_A/1-668 FFIN~GASFVPP~TVPVLLQILSGAQT~~~~~~~~AQDLLPS
2QT6_A/1-668 FFIN~GESFTPP~TVPVLLQIISGANT~~~~~~~~AQDLLPS
3DIV_A/1-668 FFIN~GASFVPP~SVPVLLQILSGAQT~~~~~~~~AQDLLPS
1GYC_A/1-668 FFIN~NASFTPP~TVPVLLQILSGAQT~~~~~~~~AQDLLPA
2H5U_A/1-668 FFIN~GASFTPP~SVPVLLQILSGAQT~~~~~~~~AQDLLPS
2XYB_A/1-668 FFIN~DHTFVPP~SVPVLLQILSGAQA~~~~~~~~AQDLVPE
3X1B_A/1-668 FFIN~DATFTPP~SVPVLLQILSGAQA~~~~~~~~AQDLLPS
3KW7_A/1-668 FFIN~NETLIPP~TVPVLLQILSGAQS~~~~~~~~ASDLLPT
2HZH_A/1-668 HFVD~GASFVPP~TVPVLSQIVSGAQS~~~~~~~~AADLLAS
5E9N_A/1-668 FSID~GVSFVPP~TVPVLLQILSGAQN~~~~~~~~AQDLLPA
5DAO_A/1-668 FQVD~NVSFVPP~TVPILLQILSGAHT~~~~~~~~AQDLMPA
3T6V_A/1-668 FVIN~GVSFTPP~TVPVLLQICSGANI~~~~~~~~AADLLPS
1V10_A/1-668 FTIN~GAPFIPP~TVPVLLQILSGVTN~~~~~~~~PNDLLPG
1A65_A/1-668 FTIN~GTAMESP~SVPTLLQIMSGAQS~~~~~~~~ANDLLPA
1HFU_A/1-668 FTIN~GTAMESP~SVPTLLQIMSGAQS~~~~~~~~ANDLLPA
3SQR_A/1-668 WTINSSSLLLDW~SSPTTLKIFNN~~~~~~~~~~~ETIFPT
3V9E_A/1-668 WTINSSSLLLDW~SSPTTLKIFNN~~~~~~~~~~~ETIFPT


470
480
490
500
WT/1-668
1ZPU_A/1-668
3PPS_A/1-668
3DKH_A/1-668
1GW0_A/1-668
2Q9O_A/1-668
2HRG_A/1-668
4A2D_A/1-668
4A2F_A/1-668
4JHU_A/1-668
1KYA_A/1-668
3FPX_A/1-668
3PXL_A/1-668
2QT6_A/1-668
3DIV_A/1-668
1GYC_A/1-668
2H5U_A/1-668
2XYB_A/1-668
3X1B_A/1-668
3KW7_A/1-668
2HZH_A/1-668
5E9N_A/1-668
5DAO_A/1-668
3T6V_A/1-668
1V10_A/1-668
1A65_A/1-668
1HFU_A/1-668
3SQR_A/1-668
3V9E_A/1-668
510
520
530
540
WT/1-668
1ZPU_A/1-668
3PPS_A/1-668
3DKH_A/1-668
1GW0_A/1-668
2Q9O_A/1-668
2HRG_A/1-668
4A2D_A/1-668
4A2F_A/1-668
4JHU_A/1-668
1KYA_A/1-668
3FPX_A/1-668
3PXL_A/1-668
2QT6_A/1-668
3DIV_A/1-668
1GYC_A/1-668
2H5U_A/1-668
2XYB_A/1-668
3X1B_A/1-668
3KW7_A/1-668
2HZH_A/1-668
5E9N_A/1-668
5DAO_A/1-668
3T6V_A/1-668
1V10_A/1-668
1A65_A/1-668
1HFU_A/1-668
3SQR_A/1-668
3V9E_A/1-668

10


550
560
570
580
WT/1-668
1ZPU_A/1-668
3PPS_A/1-668
3DKH_A/1-668
1GW0_A/1-668
2Q9O_A/1-668
2HRG_A/1-668
4A2D_A/1-668
4A2F_A/1-668
4JHU_A/1-668
1KYA_A/1-668
3FPX_A/1-668
3PXL_A/1-668
2QT6_A/1-668
3DIV_A/1-668
1GYC_A/1-668
2H5U_A/1-668
2XYB_A/1-668
3X1B_A/1-668
3KW7_A/1-668
2HZH_A/1-668
5E9N_A/1-668
5DAO_A/1-668
3T6V_A/1-668
1V10_A/1-668
1A65_A/1-668
1HFU_A/1-668
3SQR_A/1-668
3V9E_A/1-668


590
600
610
620
WT/1-668
1ZPU_A/1-668
3PPS_A/1-668
3DKH_A/1-668
1GW0_A/1-668
2Q9O_A/1-668
2HRG_A/1-668
4A2D_A/1-668
4A2F_A/1-668
4JHU_A/1-668
1KYA_A/1-668
3FPX_A/1-668
3PXL_A/1-668
2QT6_A/1-668
3DIV_A/1-668
1GYC_A/1-668
2H5U_A/1-668
2XYB_A/1-668
3X1B_A/1-668
3KW7_A/1-668
2HZH_A/1-668
5E9N_A/1-668
5DAO_A/1-668
3T6V_A/1-668
1V10_A/1-668
1A65_A/1-668
1HFU_A/1-668
3SQR_A/1-668
3V9E_A/1-668


640
650
660
WT/1-668 GFAVVMAEDTPEVAATNPVPQAWSDLCPTYDALSPDDQ
1ZPU_A/1-668 GLGLVLVED~~~~~~~~~~~~~~~~~~~~~~~~~~~~
3PPS_A/1-668 GLSVDFLE~~~~~~~~~~~~~~~~~~~~~~~~~~~~
3DKH_A/1-668 GLSVDFLE~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1GW0_A/1-668 GLSVDFLE~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2Q9O_A/1-668 GLSVDFLE~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2HRG_A/1-668 GFAVVMAEDIPEVAATNPVPQAWSDLCPTYDALSPDDQ
4A2D_A/1-668 GFAVVMAEDIPDVAATNPVPQAWSDLCPTYDALSPDDQ
4A2F_A/1-668 GFAVVMAEDIPDVAATNPVPQAWSDLCPTYDALSPDDQ
4JHU_A/1-668 GFAVVMAEDIPDVAAVNPVPQAWSDLCPTYNALDPNDQ
1KYA_A/1-668 GFAVVFAEDIPDVASANPVPQAWSDLCPTYDARDPSDQ
3FPX_A/1-668 GFAVVMAEDTPDVKAVNPVPQAWSDLCPTYDALDPNDQ
3PXL_A/1-668 GFAVVMAEDTPDVKAVNPVPQAWSDLCPTYDALDPNDQ
2QT6_A/1-668 GFAVVMAEDIPNTVNANPVPQAWSNLCPTYDALEPSNE
3DIV_A/1-668 GFAVVQAEDVPDVKATNPVPQAWSDLCPTYDANAPSDQ
1GYC_A/1-668 GFAIVFAEDVADVKAANPVPKAWSDLCPIYDGLSEANQ
2H5U_A/1-668 GFAVVQAEDVPDVKATNPVPQAWSDLCPTYDANAPSDQ
2XYB_A/1-668 GFAVVMAEDTPDTKAANPVPQAWSDLCPIYDALDPSD~
3X1B_A/1-668 GFAVVFAEDLPGTPAANPVPQSWSDLCPIYDALAEDDQ
3KW7_A/1-668 GFAIVFAEDTPDTASVNPVPTAWSDLCPTYDALDPSDH
2HZH_A/1-668 GFAVVFAQDIPDVASANPTPNAWSDLCPVYDAASSSSQ
5E9N_A/1-668 GFAIVFAEDIPDTASANPVPQAWSDLCPAYD~~~~~~~
5DAO_A/1-668 GFAVVMAEDIPDAAAANPVPAAWNELCPLYDALTPGNQ
3T6V_A/1-668 GFAIVFAEDIPNTASANPVPEAWSNLCPSYDS~~~~~~
1V10_A/1-668 GLAVVFAEDIPNIPIANAISPAWDDLCPKYNANNPD~~
1A65_A/1-668 GLAIVFAEDMANTVDANNPPVEWAQLCEIYDDLPPE~~
1HFU_A/1-668 GLAIVFAEDMANTVDANNPPVEWAQLCEIYDDLPPE~~
3SQR_A/1-668 GLAMQFVESQSSIAVKMTDTAIFEDTCANWNAYTP~~~
3V9E_A/1-668 GMAMQFVESQSSIAVKMTDTAIFEDTCANWNAYTP~~~

3.3. Rational Engineering of Multiple Active Sites in an Ester Hydrolase

Gerard Santiago†○ , Mónica Martínez-Martínez‡○, Sandra Alonso‡, Rafael Bargiela‡⊥, Cristina Coscolín‡, Peter N. Golyshin§‖ , Víctor Guallar*†# , and Manuel Ferrer*‡

† Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain

‡ Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain

⊥ School of Chemistry, §School of Biological Sciences, and ‖ Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, United Kingdom

Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Biochemistry

Article

Cite This: Biochemistry XXXX, XXX, XXX−XXX

pubs.acs.org/biochemistry

Rational Engineering of Multiple Active Sites in an Ester Hydrolase

Gerard Santiago,[†,○] Mónica Martínez-Martínez,[‡,○] Sandra Alonso,[‡] Rafael Bargiela,[‡,⊥] Cristina Coscolín,[‡] Peter N. Golyshin,[§,∥] Víctor Guallar,[*,†,#] and Manuel Ferrer[*,‡]

[†]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain

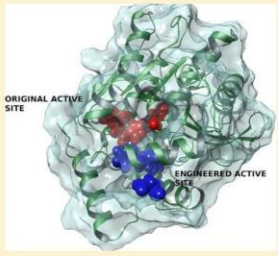
[‡]Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain

[⊥]School of Chemistry, [§]School of Biological Sciences, and [∥]Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, United Kingdom

[#]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Effects of altering the properties of an active site in an enzymatic homogeneous catalyst have been extensively reported. However, the possibility of increasing the number of such sites, as commonly done in heterogeneous catalytic materials, remains unexplored, particularly because those have to accommodate appropriate residues in specific configurations. This possibility was investigated by using a serine ester hydrolase as the target enzyme. By using the Protein Energy Landscape Exploration software, which maps ligand diffusion and binding, we found a potential binding pocket capable of holding an extra catalytic triad and oxyanion hole contacts. By introducing two mutations, this binding pocket became a catalytic site. Its substrate specificity, substrate preference, and catalytic activity were different from those of the native site of the wild type ester hydrolase and other hydrolases, due to the differences in the active site architecture. Converting the binding pocket into an extra catalytic active site was proven to be a successful approach to create a serine ester hydrolase with two functional reactive groups. Our results illustrate the accuracy and predictive nature of modern modeling techniques, opening novel catalytic opportunities coming from the presence of different catalytic environments in single enzymes.



A number of studies have shown the utility of placing multiple functional groups randomly or uniformly on the surface of solid heterogeneous catalysts to enhance activity and selectivity.[1] Functional groups are positioned in those catalysts at varying distances and spatial arrangements. However, these advances have not been as well developed for enzymatic catalysts. Thus, with few exceptions,[2−4] all enzymes contain one active site. Although introducing single active sites into noncatalytic proteins has been successfully achieved,[5−16] no evidence has been gathered that demonstrates that introducing a second or more active sites into an enzyme is feasible. If one could use molecular modeling to discover extra binding pockets and convert them into catalytic sites, then one would facilitate designing enzymes with multiple reactive groups. This would promote the competitiveness and catalytic opportunities of enzymes coming from different catalytic environments and be a considerable step forward in the field of *de novo* enzyme design.

We tested this hypothesis by using a serine ester hydrolase from the structural superfamily of α/β-hydrolases as a model enzyme. The activity of serine ester hydrolases relies mainly on a catalytic triad usually formed by Ser, Asp/Glu, and His.[17] The initial attack of an ester is mediated by the Ser residue that acts as nucleophile through which a covalent intermediate is formed in collaboration with two other residues (Asp and His) that help activate the nucleophile by forming a charge-relay network. At basic pH the His residues are δ-protonated, and the Asp residues are deprotonated, resulting in the formation of a His-Ser and His-Asp hydrogen-bonding network. The stabilization of the intermediate is increased by so-called oxyanion hole contacts with nitrogen atoms of adjacent Gly residues.[17]

This enzyme class was selected for a number of reasons. First, it is widely distributed in the environment, it has important physiological functions, it includes hydrolases which are among the most important industrial biocatalysts, and extensive biochemical knowledge has been accumulated.[18] Second, a number of studies have disclosed the utility of computational tools to introduce active sites supporting ester hydrolysis in a number of noncatalytic proteins. A first attempt used RosettaMatch to search for catalytically inert theozyme scaffold proteins that could accommodate a model catalytic triad (Cys-His-Glu/Asp) and oxyanion holes reassembling cysteine-like geometries.[11] Authors identified four protein scaffolds with minimalistic catalytic schema consisting of a Cys nucleophile, a nearby His and a backbone NH group, which were capable of hydrolyzing the model esters coumarin-(2-phenyl)-propanoate and *p*-nitrophenyl-(2-phenyl)-propa-

Received: March 6, 2018

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

noate (pN2PP) (k_2/K_S up to 309 M^{-1} s^{-1}). More recently, Moroz and colleagues took advantage of existing enzymatic folds to explore whether calmodulin (CaM) could assume esterase activity against pN2PP.[16] Docking was used to determine whether *p*N2PP accommodates into a hydrophobic cavity of CaM. Residues facing the substrate were then identified and further mutated to His, a residue that, per se, has been shown to catalyze hydrolysis of *p*N2PP, albeit with a modest activity. A mutant with a single Met144His substitution was capable of hydrolyzing *p*N2PP (k_{cat}/K_M up to 4800 M^{-1} min^{-1}). In another approach, the hydrolysis of *p*-nitrophenyl acetate (*p*NPA) was achieved by incorporating mutations favoring metal binding, with the metal being the catalytic component. By using this approach, Zn^{2+} and/or Hg^{2+} ions were introduced into the Rab-4-binding domain,[10] the TRI family of peptides,[12] and short amyloid-forming peptides.[14] The resulting metallo-proteins were capable of *p*NPA hydrolysis (k_{cat}/K_M up to 630, 23.3, and 18 M^{-1} s^{-1}). The question of where to put the functional group has been found to be critical for catalytic performance. For example, it has been demonstrated that the His orientation modulates local metal orientation, which in turn has functional consequences.[19,20]

It remains to be established whether the computational approaches described above can be implemented to introduce multiple active sites supporting ester hydrolysis, not only in noncatalytic proteins but also in ester hydrolases already containing a native functional group. The novelty of this study relies on using the Protein Energy Landscape Exploration (PELE) software. PELE offers one of the best modeling alternatives to map protein−ligand dynamics and induced fit.[21−25] It allows for a complete protein surface exploration, locating binding site pockets in only a few hours of a moderate computing cluster (~32 computing cores).[26,27] Thus, these technological developments are ideally suited to locate, without previous knowledge, potential binding sites that could be converted into active sites. Finding these pockets typically requires significant enzyme reorganization, both at the level of conformational sampling and induced fit.[27] Thus, similar analysis cannot be performed with simple docking techniques. Effective alternative techniques are mostly limited to molecular dynamic simulations,[28] although at a significantly higher computational cost.

The present study adds important insights and empirical and computational data proving, for first time to the best of our knowledge, that introducing extra catalytic reactive groups into a serine ester hydrolase is plausible. We would like to highlight that our approach is based on locating, by PELE software, extra binding pockets and converting them into catalytic active sites. We were interested not only in proving that introducing extra functional groups into a catalytic serine ester hydrolase is plausible, but also in examining the catalytic potential of the resulting variant compared to the wild type enzyme and other native ester hydrolases. In this line, we have unambiguously confirmed that the extra catalytic site is not only competitive with those of other native ester hydrolases, but also could confer catalytic changes when introduced in the wild type ester hydrolase already containing a native active site.

EXPERIMENTAL AND COMPUTATIONAL METHODS

Chemicals and Oligonucleotides. All chemicals used for enzymatic tests were of the purest grade available and were purchased from Sigma Chemical Co. (St Louis, MO, USA), Alfa Aesar (Karlsruhe, Germany) or Santa Cruz Biotechnology (Heidelberg, Germany). The oligonucleotides used for DNA amplification were synthesized by Sigma Genosys Ltd. (Pampisford, Cambs, UK).

$EH1_A$ Protein Source and Crystal Structure. The isolation of the enzyme $EH1_A$ used in the present study was reported previously.[29] The enzyme is available in the expression vector pET46 Ek/LIC plasmid and *Escherichia coli* BL21 as a host,[29,30] which was the source of the enzyme for the present study. The crystal structure of $EH1_A$ protein was recently solved (PDB code: 5JD4), and X-ray diffraction data collection and refinement statistics are available.[30]

Site Directed Mutagenesis. Mutagenic PCR was developed using the Quick Change Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies, Cheadle, UK), following manufacturer instructions. Briefly, 50 ng of pET46 Ek/LIC plasmid containing wild type $EH1_A$ DNA insert was mixed with a master mix containing 2.5 μL of 10× Multi reaction buffer, 0.5 μL of Quick solution, 1 μL of dNTP mix, 1 μL of Multi enzyme blend, and 100 ng of each primer. Distilled water was added to a final volume of 25 μL. PCR conditions were as follows: 2 min at 95 °C, 30 cycles of 20 s at 95 °C, 30 s at 55 °C, 195 s at 65 °C, and one cycle of 5 min at 65 °C. The resulting variant plasmids were transferred into *E. coli* BL21 and selected on the Luria−Bertani (LB) agar supplemented with 50 μg mL^{-1} ampicillin. To obtain the variant $EH1_B$, the following primers were used: Ser161Ala Fwd (GTG GGC GGC GAT GCG GCG GGC GGC G), Glu25Asp Fwd (CGG CCC CGG CTG GAT ACC CTG CCG CAT GC), and Leu214His Fwd (TTC CTC AGC AAG GCG CAC ATG GAC TGG TTC TGG G). To obtain the variant $EH1_{AB}$, only the primers Glu25Asp Fwd and Ile214His Fwd were used.

To obtain variants containing Ser211Ala, Ser161Ala, Asp25Gln, Asp256Gln, His214Phe, and His286Phe mutations, individually or in combination, the pET46 Ek/LIC plasmid containing $EH1_A$, $EH1_B$, or $EH1_{AB}$ DNA inserts and the primers Ser211Ala Fwd (GCC GAA GGC TAC TTC CTC GCC AAG GCG CAC ATG GAC TGG), Ser161Ala Fwd (GTG GGC GGC GAT GCG GCG GGC GGC G), Asp25Gln Fwd (CGG CCC CGG CTG CAG ACC CTG CCG CAT GC), Asp256Gln Fwd (ACC GCC GGC TAC CAA CCG CTG CGC GAC G), His214Phe Fwd (TTC CTC AGC AAG GCG TTC ATG GAC TGG TTC TGG G), and His286Phe Fwd (T CCC GGC ACC ATC TTC GGC TTC TTC TCG) were used. Mutagenic PCR conditions were as above.

Gene Expression and Protein Purification. Protein expression and purification of wild type and mutants were performed as previously described with slight modifications.[29,30] Briefly, selected *E. coli* clones that expressed each protein, His-tagged at the N-terminus, were grown at 37 °C on solid LB agar medium supplemented with 50 μg mL^{-1} ampicillin, and one colony was picked and used to inoculate 10 mL of LB broth plus antibiotic in a 0.25-l flask. The cultures were then incubated at 37 °C and 200 rpm overnight. Afterward, 10 mL of this culture was used to inoculate 0.5 L of LB medium, which was then incubated to an $OD_{600\ nm}$ to approximately 0.7 (ranging from 0.55 to 0.75) at 37 °C. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of approximately 1 mM, followed by incubation for 16 h at 16 °C. The cells were harvested by centrifugation at 5000*g* for 15 min to yield a pellet of 2−3 g L^{-1} pellet (wet weight). The wet cell pellet was frozen at −86 °C overnight, thawed, and resuspended in 15 mL of 40

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0. Lysonase Bioprocessing reagent (Novagen, Darmstadt, Germany) was added (4 μL g^{-1} wet cells) and incubated for 60 min on ice with rotating mixing. The cell suspension was sonicated using a pin Sonicator 3000 (Misonix, New Highway Farmingdale, NY, USA) for a total time of 5 min (10 W) on ice and centrifuged at 15000*g* for 15 min at 4 °C, and the supernatant was retained.

The His-tagged proteins, native and engineered variants, were purified at 4 °C after binding to a Ni-NTA His-Bind resin (Sigma Chemical Co. (St. Louis, MO, USA)), followed by ultrafiltration through low-adsorption hydrophilic 10000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) to concentrate the protein solution. An extensive dialysis of protein solutions against 40 mM HEPES buffer (pH 7.0) was then performed using Pur-A-LyzerTM Maxi 1200 dialysis kit ((Sigma Chemical Co. (St. Louis, MO, USA)), as follows. Five milliliters concentrated protein solution was dialyzed against 2 L buffer during 1 h at room temperature, after which the buffer was changed by other 2 L buffer and maintained 1 h more. Then, the buffer was changed, and the dialysis was kept overnight at 4 °C. The dialyzed protein solution was recovered and concentrated as before. The concentrated protein solution (10 mg mL^{-1}) was then subjected to size-exclusion chromatography by a fast protein liquid chromatography (FPLC) equipment (LCC-500CI, Amersham Bioscience, Barcelona, Spain). The protein sample was loaded onto the FPLC coupled with a Superdex 75 size exclusion column pre-equilibrated with buffer HEPES buffer (pH 7.0). The proteins were eluted with the same buffer at a flow rate of 1 mL min^{-1}. Fractions with hydrolytic activity were pooled, concentrated, and dialyzed against HEPES buffer (pH 7.0), as before. Throughout the purification protocol, the fractions were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on 12% gels, in a Mini PROTEAN electrophoresis system (Bio-Rad),[31] in which the proteins were stained with Coomassie brilliant blue (Protoblue Safe, National Diagnostics, GA, USA), and for hydrolytic activity using *p*-nitrophenyl propionate (*p*NPP) as described below. The protein concentration was determined according to Bradford with bovine serum albumin as the standard.[32]

Purity was assessed as >99% by SDS-PAGE, matrix-assisted laser desorption/ionization-time-of-flight/time-of-flight (MALDI-TOF/TOF) and other complementary techniques (see Supporting Results).

Ester Bond Hydrolysis Activity Assessment: Substrate Profiling Tests with 96 Esters. Hydrolytic activity was assayed at 550 nm using structurally diverse esters in 384-well plates as previously described.[30,33] All chemicals used were of the purest grade available. Briefly, before an assay, a concentrated ester stock solution was prepared in a 96-well plate by dissolving each of the 96 esters at a concentration of 25 mg mL^{-1} in acetonitrile or dimethyl sulfoxide (DMSO), depending on its solubility. Stock solutions were prepared immediately prior to use and maintained in a 96-deep-well plate at 4 °C.

The assays were conducted according to the following steps. First, a 384-well plate (Molecular Devices, LLC, CA, USA) was filled with 20 μL of 5 mM *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid (EPPS) buffer, pH 8.0, using a QFill3 microplate filler (Molecular Devices, LLC, CA, USA). Second, 2 μL of each ester stock solution was added to each well using a PRIMADIAG liquid-handling robot (EYOWN TECHNOLOGIES SL, Madrid, Spain). Each of the 96 esters was dispensed in four replicates in each 384-plate. After the esters were added, the 384-well plate was filled with 20 μL of 5 mM EPPS buffer, pH 8.0, containing 0.912 mM Phenol Red (used as a pH indicator) using a QFill3 microplate filler. The final ester concentration in each well was 1.14 mg mL^{-1}, and the final concentration of Phenol Red was 0.45 mM. A total of 2 μL of protein solution (from a 1 (for $EH1_A$/$EH1_{AB}$) or 8 (for $EH1_B$) mg mL^{-1} stock solution in 40 mM HEPES buffer pH 7.0) was immediately added to each well using an Eppendorf Repeater M4 pipet (Eppendorf, Hamburg, Germany) or a PRIMADIAG liquid-handling robot. Accordingly, the total reaction volume was 44 μL, with 4.5% (v/v) acetonitrile or DMSO in the reaction mixture. After incubation at 30 °C and 150 rpm in a Synergy HT Multi-Mode Microplate reader, ester hydrolysis was measured spectrophotometrically in continuous mode at 550 nm for a total time of 24 h. One unit (U) of enzyme activity was defined as the number of enzyme required to transform 1 μmol of substrate in 1 min under the assay conditions using the reported extinction coefficient ($\varepsilon_{\text{Phenol red}}$ at 550 nm = 8450 M^{-1} cm^{-1}).[29] All values were corrected for nonenzymatic transformation.

Kinetic Measurements. For determination of kinetic parameters, these were calculated by simple Michaelis−Menten kinetics.

Kinetics experiments for *p*NPP were initiated by the addition of a stock solution (100 mM in acetonitrile) of *p*NPP to 195 μL of HEPES buffer (pH 7.0) containing $EH1_A$, $EH1_B$, or $EH1_{AB}$ in 96-well microtiter plates. Kinetics experiments for phenyl propionate were initiated by the addition of a stock solution (100 mM in acetonitrile) of phenyl propionate to 40 μL of 5 mM EPPS buffer pH 8.0 with 0.45 mM Phenol Red (used as a pH indicator) containing $EH1_A$, $EH1_B$, or $EH1_{AB}$ in 96-well microtiter plates. For K_m determinations, the amount of protein was 0.1, 16, and 1.4 μg for $EH1_A$, $EH1_B$, and $EH1_{AB}$, respectively, or 100 μg for other mutants. For k_{cat} determinations, the amount of *p*NPP and phenyl propionate was 1 and 6 mM, respectively.

In all cases, reactions were followed at 410 nm (for *p*NPP hydrolysis) or 550 nm (for phenyl propionate hydrolysis) by UV−vis spectrophotometry in a Synergy HT Multi-Mode Microplate Reader. Initial rates were determined from linear fits of the absorbance versus time (<10% conversion) corrected for the rate of uncatalyzed hydrolysis. In all cases, one unit (U) of enzyme activity was defined as the number of enzyme required to transform 1 μmol substrate in 1 min under the assay conditions using the reported extinction coefficient ($\varepsilon_{\text{pNPP}}$ at 410 nm = 15200 M^{-1} cm^{-1}; $\varepsilon_{\text{Phenol red}}$ at 550 nm = 8450 M^{-1} cm^{-1}). In all cases, reactions (performed in triplicate) were maintained at 30 °C.

pH Optima Determination. Britton and Robinson buffer (50 mM; pH 4.0−9.5) and *p*NPP (1 mM) were used for determining optimal pH of the enzyme variants. A total of 2 μL of a stock solution (100 mM in acetonitrile) of *p*NPP was added to 195 μL of HEPES buffer (pH 7.0) containing $EH1_A$ (0.1 μg), $EH1_B$ (16 μg), or $EH1_{AB}$ (1.4 μg) in 96-well microtiter plates. The *p*NPP hydrolysis was monitored as above but at 346 nm ($\varepsilon_{\text{pNPP}}$ at 346 nm = 4800 M^{-1} cm^{-1} regardless of solution pH), in triplicates.

Protein Energy Landscape Exploration (PELE) Simulations. The initial structure was taken from the coordinates of the $EH1_A$ crystal structure (PDB code: 5JD4).[30] The protonation state of titratable residues was estimated with the

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

Protein Preparation Wizard (PROPKA)[34] and the H++ server (http://biophysics.cs.vt.edu/H++) followed by visible inspection. At pH 8 (the pH at which the activity assays were performed), the catalytic triad histidine residues were δ-protonated, and the catalytic triad aspartic acid residues were deprotonated, resulting in the formation of a histidine-serine and histidine-aspartic hydrogen-bonding network. The glyceryl tripropionate structure was fully optimized with Jaguar[27] in an implicit solvent, and the electrostatic potential charges were computed with the density functional M06 at the 6-31G* level of theory; ligand parameters were extracted from these for the classic simulations.

We used PELE software to sample the binding mode of glyceryl tripropionate with $EH1_A$.[21] PELE is a Monte Carlo algorithm composed of a sequence of perturbation, relaxation, and Metropolis acceptance tests. In the first step, the ligand is subjected to random rotations and translations, while the protein is perturbed based on the anisotropic network model (ANM).[27] The maximum allowed translation for the ligand perturbation was 1.5 Å, and the maximum rotation was 20°. During the protein perturbation, all atoms were displaced by a maximum of 0.5 Å by moving the α-carbons following a random linear combination of the six lowest eigenvectors obtained in the ANM model. The relaxation step included the repositioning of all amino acid side chains within 6 Å of the ligand and the five side chains with the highest energy increase along the previous ANM step. The relaxation stage ended with a truncated Newton minimization using the OPLS all-atom force field and an implicit surface-generalized Born continuum solvent.[26] The new proposed minima were then accepted or rejected based on a Metropolis test. The substrate binding plots contained all accepted conformations for three 12-h simulations using 200 processors.

Molecular Dynamics. 250 ns of molecular dynamics (MD) simulation with DESMOND[35] were performed to ensure the enzymatic stability. After appropriate preparation of the system, as explained before, an orthorhombic water box with a minimum distance of 10 Å was introduced. The systems were then neutralized and 150 mM NaCl added. Equilibration using the default protocol was performed followed by 20 ns NPT simulation at 300 K and 1 atm with the OPLS-2005 force field. The temperature was regulated with the Nosé−Hoover chain thermostat, while the pressure was controlled by the Martyna−artynae was cbarostat with isotropic coupling and a relaxation time of 2.0 ps.

Peptide Mass Fingerprinting by Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight/Time-Of-Flight (MALDI-TOF/TOF). Before MALDI-TOF/TOF analysis in-solution, protein digestion was performed.[36] Briefly, 20 μg of protein samples (5 mg mL^{-1} in 40 mM HEPES buffer pH 7.0) were diluted and denatured in 20 μL of 7 M urea/2 M thiourea/100 mM triethylammonium bicarbonate (TEAB), pH 7.5, reduced with 2 μL of 50 mM tris(2-carboxyethyl) phosphine (TCEP, AB SCIEX), pH 8.0, at 37 °C for 60 min, followed by addition of 2 μL of 200 mM cysteine-blocking reagent (methylmethanethiosulfonate (MMTS); Pierce, Appleton, WI, USA) for 10 min at room temperature. Samples were diluted up to 120 μL to reduce guanidine concentration with 50 mM TEAB. Digestions were performed using sequence grade-modified trypsin (Promega, Alcobendas, Spain) to each sample in a ratio 1/20 (w/w), which were then incubated at 37 °C overnight on a shaker. Sample digestions were evaporated to dryness and were cleaned-up/desalted using Stage-Tips with Empore 3 M C18 disks (Sigma Chemical Co.; St. Louis, MO, USA). The tryptic eluted peptides were dried by speed-vacuum centrifugation and resuspended in 4 μL of MALDI solution (30% acetonitrile/15% isopropanol/0.5% trifluoroacetic acid). A 0.8 μL aliquot of each peptide mixture was deposited onto a 384-well OptiTOF Plate (SCIEX, Foster City, CA) and allowed to dry at room temperature. A 0.8 μL aliquot of matrix solution (3 mg mL^{-1} α-cyano-4-hydroxycinnamic acid in MALDI solution) was then deposited onto dried digest and allowed to dry at room temperature.

For MALDI-TOF/TOF analysis, samples were automatically acquired in an ABi 4800 MALDI TOF/TOF mass spectrometer (SCIEX, Foster City, CA) in positive ion reflector mode (the ion acceleration voltage was 25 kV to MS acquisition and 2 kV to MSMS), and the obtained spectra were stored into the ABi 4000 Series Explorer Spot Set Manager. PMF and MSMS fragment ion spectra were smoothed and corrected to zero baseline using routines embedded in ABi 4000 Series Explorer Software v3.6. Each PMF spectrum was internally calibrated with the mass signals of trypsin autolysis ions to reach a typical mass measurement accuracy of <25 ppm. Known trypsin and keratin mass signals, as well as potential sodium and potassium adducts (+21 Da and +39 Da) were removed from the peak list. To submit the combined PMF and MS/MS data to MASCOT software v.2.6.0 (Matrix Science, London, UK), GPS Explorer v4.9 was used, searching in a custom protein database with the sequences encoding $EH1_A$, $EH1_B$ and $EH1_{AB}$. The following search parameters were used: enzyme, trypsin; allowed missed cleavages, 1; methylthiolation cysteine as fixed modification by the treatment with MMTS; variable modifications, oxidation of methionine; mass tolerance for precursors was set to ±50 ppm and for MS/MS fragment ions to ±0.3 Da. The confidence interval for protein identification was set to ≥95% ($p < 0.05$) and only peptides with an individual ion score above the identity threshold were considered correctly identified.

Estimation of Molecular Mass by MALDI-TOF/TOF. Protein samples were diluted at 1:1 ratio (v/v) with matrix solution (50% saturated sinapinic acid in 70% aqueous acetonitrile and 0.1% trifluoroacetic acid). A 1.0 μL aliquot of this mixture was manually deposited onto a 386-well OptiTOF Plate (ABSciex, Framingham, MA, USA) and allowed to dry at room temperature. For MALDI-TOF/TOF analysis, samples were automatically acquired in an ABi 4800 MALDI TOF/TOF mass spectrometer (SCIEX, Foster City, CA) in positive ion linear mode (the ion acceleration voltage was 25 kV for MS acquisition). The detection mass range was set between 1500 and 80000 m/z.

RESULTS AND DISCUSSION

Model Ester Hydrolase for This Study. We chose a serine ester hydrolase, herein referred to as $EH1_A$, with a typical α/β hydrolase fold as a model. In a recent study, it was identified as the most promiscuous ester hydrolase among a total of 147 when tested with a set of 96 chemically and structurally different esters (Supporting Table S1).[30] $EH1_A$ was isolated from the metagenomic DNA of microbial communities inhabiting a karstic lake,[29] and its structure was recently solved (PDB code 5JD4).[30] The active site is formed by a catalytic triad formed by Ser161, Asp256, and His286 (Figure 1) and an oxyanion hole formed by Gly88, Gly89, and Gly90.[30] This active site supports the hydrolysis of a broad range of 72 esters

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

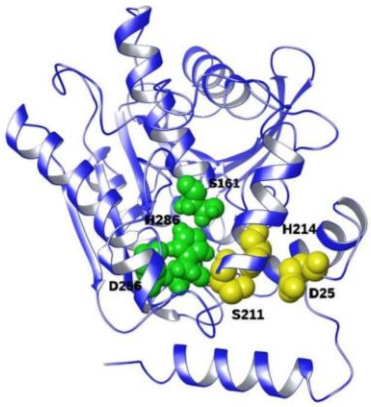


Figure 1. Relative position of the two active sites, with their catalytic triads. Original $EH1_A$ catalytic triad and oxyanion hole are shown in green and the designed $EH1_B$ catalytic triad and oxyanion hole in yellow.

(Figure 2, inset), with vinyl butyrate and phenyl propionate serving as best substrates (Figure 2).

PELE Simulations for Locating an Extra Binding Pocket. Using the PELE software, which allows mapping ligand diffusion and binding,[21−25] we performed an exploration of $EH1_A$ with glyceryl tripropionate, a ligand with high activity in multiple ester hydrolases, including $EH1_A$ (~62000 ± 7400 U g^{-1}; Figure 2), and appropriately sized for binding into defined cavities. The rationale was to identify (alternative) potential binding sites to accommodate a new active site (Figure 3). PELE simulations revealed a potential second binding site located ~13 Å from the native catalytic position at Ser161 (Figure 4a). This second site (Figure 4b) already contains a serine residue (Ser211). Thus, we computationally designed additional mutations, adding Asp and His residues to build a proper catalytic triad, taking special care of distances between the residues and substrate accommodation. PELE results for the Glu25Asp and Leu214His double mutant revealed good enzyme−substrate interaction energies (Supporting Figure S1) and a suitable catalytic position for the glyceryl tripropionate substrate. Figure 4 summarizes the catalytic triad environment of the newly introduced active site compared to the wild type. In addition, our results show that residues Gly207, Tyr208 and Phe209 act as a potential oxyanion hole, a key element in ester hydrolase catalysis (Figure 4b). Moreover, extensive MD simulations indicate proper stabilization of the double mutant (Supporting Figure S2).

Converting the Binding Pocket into a Functional Catalytic Site by Site-Directed Mutagenesis. By using site-directed mutagenesis and with the above considerations, an enzyme variant with the new presumptive active site and an inactive original site was designed to determine whether the extra active site was, per se, functional. We introduced Glu25Asp, Leu214His, and Ser161Ala substitutions, so that this variant, named $EH1_B$, would presumably employ a new catalytic triad (Ser211, Asp25, and His214) with Ser211 as the nucleophile and a new oxyanion hole (Gly207, Tyr208, and Phe209). The corresponding gene was cloned, and the His_6-tagged protein was expressed and purified by Ni-NTA affinity and size-exclusion chromatography (Supporting Figure S3 and Table S2). A number of control experiments were performed to ensure the purity and identity of the $EH1_B$ protein (Supporting Figures S3−S8). These analyses were also performed with purified $EH1_A$ and other variants (see below). With the sensitivity allowing MALDI-TOF/TOF and other complementary techniques (see Supporting Results), we can conclude that there is no contamination and that the purity of the $EH1_B$ (as well as $EH1_A$ and further variant $EH1_{AB}$) was higher than 99%. The introduced mutations were confirmed in all cases by sequencing of the corresponding gene and by peptide mass fingerprinting by MALDI-TOF/TOF of purified proteins (for details see Supporting Results, and Figures S5 and S6).

Once the purity and nature of $EH1_B$ was confirmed, its hydrolytic activity was evaluated against the set of 96 esters as for the wild-type ester hydrolase (see Experimental Section). We found that $EH1_B$ with the new active site showed not only measurable activity but also catalyzed reactions with 24 different substrates (Figure 2), including cyclohexyl butyrate and the large aromatic ester benzoic acid, 4-formyl-, phenylmethyl ester, which are rarely hydrolyzed by many native ester hydrolases (Figure 5; esters ID nos. 59 and 67).[30] Note that in $EH1_B$ the original active site (Ser161) was silenced by mutating it by Ala, the mutation per se found to completely abolish the activity of the original active site (Supporting Figure S9D). Thus, the capacity of $EH1_B$ to convert these 24 esters could be unambiguously assigned to the new active site.

To evaluate the catalytic potential of the newly introduced active site, we compared the number of esters being hydrolyzed with that of the wild-type protein ($EH1_A$), 144 naturally occurring ester hydrolases and two commercial preparations tested against the same set of 96 esters.[30]

As shown in Figure 6, we observed that $EH1_B$ had a restricted substrate spectrum (24 esters) compared to $EH1_A$ (72 esters), the most promiscuous ester hydrolase among those examined, but it was broader than that of the other hundred native ester hydrolases. The ester hydrolase with the new active site designed herein would fall into the category of ester hydrolases with moderate substrate promiscuity, thus suggesting the hydrolytic potential of the newly introduced active site in terms of the substrate spectrum.

We further evaluated whether $EH1_B$ showed different substrate preferences compared to $EH1_A$. We considered only those 24 substrates converted by both variants (Figure 2) and calculated the relative activity for each. As shown in Figure 2, distinct profiles were observed that highlight the difference in binding capacity for each of the catalytic environments. As an example, the new active site ($EH1_B$) showed a higher relative preference for the hydrolysis of methyl-, propyl-, and ethyl-hexanoate compared to the native site ($EH1_A$). The difference in substrate binding and preference between the new ($EH1_B$) and wild type ($EH1_A$) active sites was also noticeably when comparing their capacity to hydrolyze phenyl propionate and *p*NPP (Supporting Figure S9 and Table S3). Thus, in terms of k_{cat} $EH1_B$ preferred phenyl propionate over *p*NPP (5.7-fold), whereas $EH1_A$ slightly preferred *p*NPP (1.4-fold). A local refinement simulation with PELE revealed that the changes in preference for ethyl-hexanoate correlate with substrate placement. In particular, the difference in number of catalytic events in both sites, which define the poses where the substrate is readily placed for catalysis, is significantly reduced for ethyl-hexanoate, when compared to one of the substrates with

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

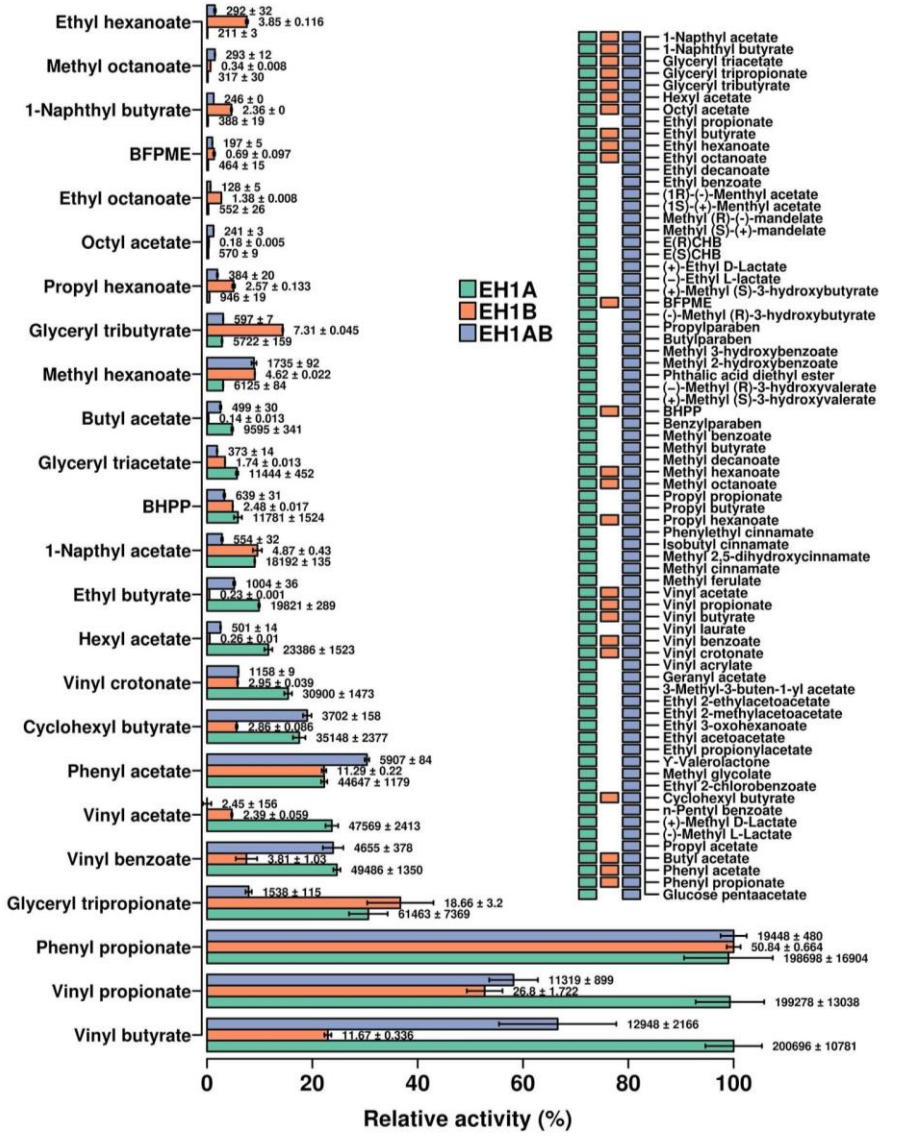


Figure 2. Substrate spectra of EH1 variants. The 72 out 96 esters for which activity was detected are listed on the right side panel, with indication of substrates being converted by each of the three variants. The ID code representing each EH variant is color coded. On the left main panel the relative activity (%) for each ester referred to the best substrate is indicated, with specific activity (U g^{-1}) and standard deviation (triplicates) indicated in the bars. Abbreviations as follows: E(R/S)CHB: ethyl (*R*/*S*)-(+)-4-chloro-3-hydroxybutyrate; BFPME: benzoic acid, 4-formyl-, phenylmethyl ester; BHPP: benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate. The activity protocol established and used to identify the esters hydrolyzed by each variant was based on a 550 nm follow-up pH indicator assay at pH 8.0 and 30 °C (see Experimental Section).

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

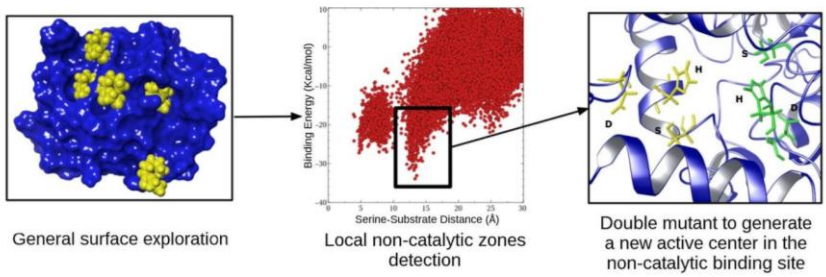


Figure 3. Molecular modeling protocol used for adding a second catalytic active site: (i) Complete surface exploration with the probe ligand (glyceryl tripropionate); (ii) alternative binding site location; (iii) substrate placement simulations (local refinement on the new site) with mutations.

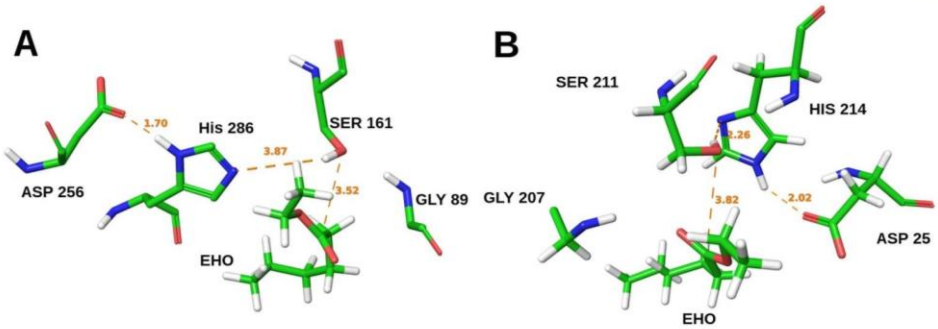


Figure 4. Catalytic position adopted by ethyl hexanoate (EHO) in the wild type active site (A) and in the Glu25Asp and Leu214His double mutant (B). Along with the substrate, the catalytic triad and the oxyanion hole residue are shown; main catalytic distances are underlined in angstroms.

maximum activity, phenyl propionate (Figure 7). Thus, differences in substrate profile may be due to distinct active site architectures (Figures 1 and 4). Additionally, there are different active site effective volumes (the active site cavity volume corrected by the relative solvent accessible surface area (SASA) of the catalytic triad) for the $EH1_B$ active site (82.2 $Å^3$) and the $EH1_A$ active site (167.7 $Å^3$). This volume has been recently correlated with changes in substrate promiscuity in ester hydrolases.[30]

Finally, we compared the hydrolytic rate of $EH1_B$ with that of $EH1_A$ and 225 previously reported ester hydrolases. We chose glyceryl tributyrate as a representative ester substrate because its specific activity values are commonly reported. As shown in Figure 2, the activity (7.3 ± 0.1 U g^{-1}) might seem moderate compared to that of $EH1_A$ (5722 ± 159 U g^{-1}). This was also proven by calculating the k_{cat}/K_M for phenyl propionate and *p*NPP, which was also found as a substrate (see Supporting Figure S9 and Table S3). On the basis of the modeled structure (Figure 4), the designed side chains were nearby but not quite in the typical orientation for the catalytic triad, which may explain why catalysis was less effective for $EH1_B$ compared to $EH1_A$. It is well-known that catalytic triads in serine ester hydrolases require a precise relative orientation of the Ser-Asp-His side chains.[11] Whatever the difference with $EH1_A$, we observed that the specific activity of $EH1_B$ is comparable to or higher than that of 93 previously reported native ester hydrolases that showed either low (≤6.8 U g^{-1}) or no (near zero) capacity to hydrolyze this substrate (about 70 esterases) (Figure 6 inset). The other reported 132 esterases were found far more active for the target substrate. Together, we can conclude that, at least for the hydrolysis of glyceryl tributyrate, the designed active site albeit being characterized by low catalytic activity, is as active as those of many natural hydrolases, but still not competitive with those of most esterases.

We found that single Ser211Ala, as well as Asp25Gln or His214Phe, substitutions in $EH1_B$ completely arrested the activity for all 96 chemically and structurally different esters examined in this study (Supporting Table S2), that included the 24 different esters hydrolyzed by $EH1_B$ (see Figure 2), and *p*NPP (see Supporting Results and Figures S9 and S10). This unambiguously confirms that they are the functional groups supporting catalysis in the new active site, which is constituted by a catalytic triad, and not a diad or a histidine. Altogether, these data demonstrated that it is plausible to use molecular modeling to find a potential extra binding pocket in a serine ester hydrolase that can be turned into a catalytic site. This site compares in terms of substrate spectra and catalytic activity with those from many other reported native ester hydrolases. It is difficult to make a statement whether the substrate spectrum of the newly introduced active site is comparable to, or is higher than that in the variants designed *de novo* by computational

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

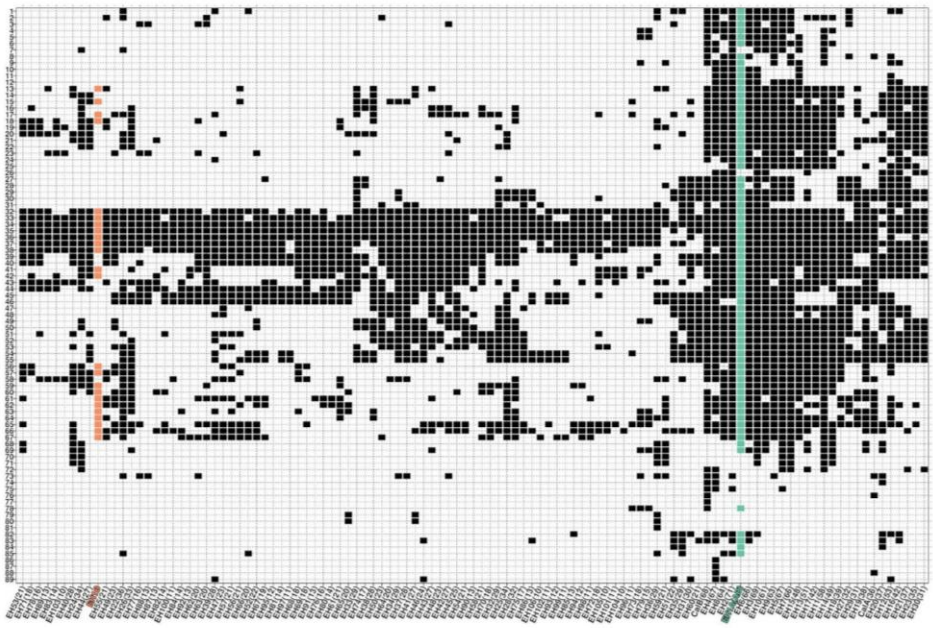


Figure 5. Clustering of the substrate range of EH1 variants within them and in comparison with those of other serine ester hydrolases. A total of 107 ester hydrolases, including the commercial preparations CalA and CalB from *Pseudozyma aphidis*, formerly *Candida antarctica*, are included. The data corresponding to each EH variant is color coded. This figure is created from data previously reported[30] and data herein obtained for EH1 variants (see Experimental Section). The list of the esters tested is shown on the left side (full name of the esters from no. 1 to 89 is given in Supporting Table S4). The ID code representing each ester-hydrolase is given at the bottom. Each hydrolase is named based on the code "EH", which means ester hydrolase, followed by an arbitrary number. The number in brackets indicates the number of esters hydrolyzed by each enzyme. The figure was created with the R language console, as described previously,[30] using a binomial table with information about the activity/inactivity (1/0) of the analyzed enzymes against the 96 substrates as a starting point.

tools previously described.[10−12,14,16] This is because these variants were only tested with a restricted set of *p*-nitrophenyl esters, substrates that are more easily hydrolyzed compared to the esters tested herein (Figure 2). It is clear though that, at least, $EH1_B$ is capable of *p*NPP hydrolysis with a k_{cat}/K_M of 16.4 M^{-1} s^{-1} (Supporting Figure S9 and Table S3), a value within the range or lower than that reported for other constructs using related substrates (k_{cat}/K_M from 18 to 309 M^{-1} s^{-1}).

Design and Characterization of a Serine Ester Hydrolase with Two Reactive Groups. An enzyme variant with the two active sites (the native and the newly identified) was created to prove that an enzyme with two hydrolytic active sites is functional. This mutant, named $EH1_{AB}$, was created by incorporating the Glu25Asp and Leu214His substitutions into the wild-type sequence. This two-active-site variant would likely employ Ser211 and Ser161 as nucleophiles (Figure 1). The corresponding gene was cloned, and the protein was expressed, purified (Supporting Figure S3 and Table S2), and characterized as described above.

The analysis of the variant $EH1_{AB}$ confirmed its ability to hydrolyze all 72 esters converted by $EH1_A$ (Figure 2). Thus, $EH1_{AB}$, as the native $EH1_A$, could be classified as an ester hydrolase with prominent substrate promiscuity (Figures 5 and 6). Because $EH1_A$, due to its promiscuous behavior, converted all esters hydrolyzed by $EH1_B$, we could not examine the effects of cooperativity for expanding the substrate spectra because of the presence of two different active sites in $EH1_{AB}$. We further observed that $EH1_{AB}$ does show overall a substrate preference similar to that of $EH1_A$ (Figure 2), although in some cases, the potential of the newly introduced active site to increasing the preference for a number of esters, such as methyl-, propyl-, and ethyl-hexanoate, was noticed. Also, $EH1_{AB}$ preferred phenyl propionate over *p*NPP (1.3-fold), whereas $EH1_A$ slightly preferred *p*NPP (1.4-fold) (Supporting Table S3). This may be a consequence of a higher preference of the second active site for phenyl propionate over *p*NPP.

The specific activity of $EH1_{AB}$ for all accepted esters was lower than that observed for $EH1_A$ but higher than that for $EH1_B$ (Figure 2). As example, using glyceryl tributyrate, its activity (597 ± 7 U g^{-1}) was 10-fold lower than that of $EH1_A$ and 81-fold higher than that of $EH1_B$. These differences were also observed when calculating the k_{cat}/K_M for phenyl propionate and *p*NPP (see Supporting Figure S9 and Table S3). The change in activity with respect $EH1_A$ could be explained by the decrease in substrate population for the main active site observed when introducing the mutations (Support-

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

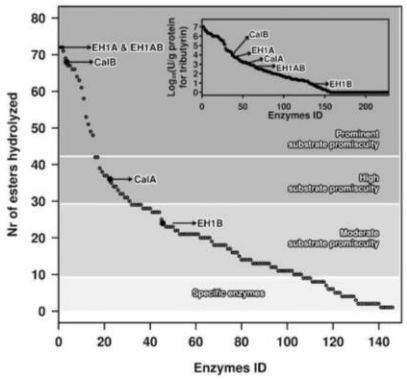


Figure 6. Comparison of the substrate range and catalytic performance of EH1 variants and other reported ester hydrolases. This figure is created from data previously reported[30] and data herein reported for EH1 variants (see Figures 2 and 5). The main figure represents the number of substrates hydrolyzed by each EH1 variant compared to that of other 146 serine ester hydrolases that included the commercial preparations CalA and CalB.[30] The inset represents the specific activity against glyceryl tributyrate of the EH1 variants and other 225 reported ester hydrolases (that included the 146 previously mentioned[30] and 79 reported elsewhere).

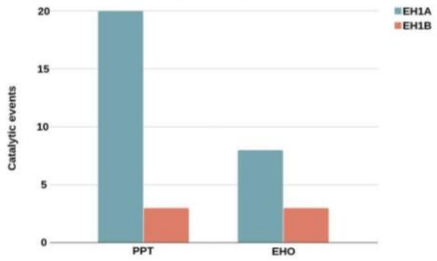


Figure 7. Number of catalytic events found in $EH1_A$ and $EH1_B$ catalytic sites along PELE simulations for ethyl hexanoate (EHO) and phenyl propionate (PPT). A catalytic event was defined when the main catalytic distances: ligand-Ser, Ser-His, and His-Asp, are lower than 3.5 Å. All simulations include the same sampling effort.

ing Figures S1a and b). Visual inspection of the models and energies (Supporting Figure S1b) shows a larger connection between the two sites, induced by changes in the hydrophobic packing (Leu24, Phe218, Leu24, and Trp217) after introducing His214. The α helix harboring the Leu214His mutation forms one face of the native active site, which slightly perturbs the main site. Whatever the catalytic activity level of the $EH1_{AB}$ variant compared to the native enzyme, its activity is significantly higher than that observed for 163 naturally occurring ester hydrolases (Figure 6, inset).

The consequences of the presence of two active sites were visible in K_M Michaelis–Menten (Supporting Figure S9 and Table S3) and pH optima (Supporting Figure S11) curves. Thus, the K_M progress curve for $EH1_{AB}$ displays a sigmoidal plot of the initial reaction rate versus substrate concentration (for both *p*NPP and phenyl propionate), rather than the hyperbolic plots observed for $EH1_A$ and $EH1_B$. This situation is typically found in allosteric enzymes, where the binding of substrates occurs in several active sites in the same enzyme molecule,[37] a situation that may most likely occurs in $EH1_{AB}$. We also observed that $EH1_{AB}$ displays an optimum pH profile distinct (narrower) to those of $EH1_A$ (which showed the broader range) and $EH1_B$ (see Supporting Results), which also differ between them (Supporting Figure S11). Subsequent investigation of the reasons that explain these differences will be needed.

We found that a combination of Ser211Ala/Ser161Ala, as well as Asp25Gln/Asp256Gln or His214Phe/His286Phe, substitutions in $EH1_{AB}$ completely arrested the activity of all 96 esters tested in this study, including the 72 esters converted by $EH1_{AB}$, as well as *p*NPP (see Supporting Results and Figures S9 and S10), unambiguously confirming their role in catalysis.

Altogether, these data unambiguously demonstrated that two different active sites can functionally and efficiently coexist in a single serine ester hydrolase. However, we did not identify a single specific advantage of this engineered ester hydrolase with two active sites, albeit we found that properties such as substrate binding and the optimum pH were altered. We hypothesized why the engineered enzyme had a lower performance. One of the mutations needed to introduce the new active site most likely slightly perturbs the native catalytic triad arrangement, which has negative consequences for the activity phenotype of the engineered mutant. Locating an extra binding pocket in positions not disturbing the original pocket and converting it into a catalytic site could solve this issue. In terms of substrate spectra, we should again highlight that $EH1_A$ already contains an active site capable of hydrolyzing a prominent number of esters, which is significantly higher than for the other 146 natural ester hydrolases, including the best commercial prototypes such as CalB (Figure 6).[30] Therefore, the probability of designing an active site capable of converting an additional set of esters compared to $EH1_A$ may be low. To unambiguously prove the advantage of introducing two active sites for expanding the substrate range, one should start with an ester hydrolase with a smaller substrate range and insert an extract active with different specificity or selectivity, which is actually being undertaken.

■ CONCLUSION

Is it possible to create enzymes with more than one active site? This idea, which has never been successfully attempted, was the starting point of the present study. Our work describes, for first time to the best of our knowledge, a neat proof-of-concept demonstration of the use of computational approaches to discover in an ester hydrolase already containing a native catalytic site, a novel binding site where a second catalytic site can be successfully introduced to generate a protein with two functional active sites for ester hydrolysis. Both active sites were structurally and catalytically different. We not only proved that introducing extra functional groups into a catalytic serine ester hydrolase is plausible, but also examined the catalytic potential of the resulting variant as compared to the target ester hydrolase and other reported native ester hydrolases. We confirmed the catalytic potential of the newly introduced active site as an individual catalytic entity or when combined with the native active site in the same polypeptide. We would like to highlight that a number of control experiments and analytics were performed to support our assertions and to ensure the

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

reported catalytic activity for the de novo designed active site was not due to a contaminant (for details see Supporting Results Section).

Can every ester hydrolase be evolved for multiple reactive groups? Possibly not, as introducing multiple catalytic environments would depend on the possibility of finding an extra binding site in an enzyme of a size large enough to bind into defined cavities. Also, if found, it may not be turned into a catalytic active site after introducing the appropriate mutations because catalysis requires amino acids at appropriate distances and angles.[11] Extending the analysis to other ester hydrolases and enzyme scaffolds will help to clarify whether the possibility to introduce extra catalytic sites is fortuitous or could be reproducible and to what extent can it happen.

The scientific and economic benefits of introducing multiple active sites into an enzyme are enormous and may revolutionize the fields of protein design and enzymology. Thus, converting binding pockets into extra catalytic active sites with different active site cavity volumes, accessible surface areas and distances, and angles between amino acids participating in the reactions compared to the native binding pockets will open novel catalytic opportunities for a given enzyme. These opportunities will come from the presence of different catalytic environments in the same polypeptide. Creating enzymes with multiple reactive groups may improve the substrate conversion range of enzymes otherwise containing only one active site, increase the competitiveness of enzymes over heterogeneous catalytic materials containing a high number of functional groups, and open application ranges with consequent reductions in production costs of other multiple enzymes. Moreover, adding new active sites to an enzyme whose original/native one cannot well be implemented for a particular application, but its overall structure has the appropriate technical requirements (scalability and robustness), may open new opportunities for a given enzyme. Such opportunities can be further explored through rational design and directed evolution focusing on residues around the new active site. Further studies in that direction would lay the foundation for streamlined and reproducible approaches for designing enzymes with multiple active sites. It would also promote new advances in a mechanistic understanding of enzymatic principles and the evolution of catalytic environments, particularly for revealing why certain areas in protein macromolecules are favored over others for incorporating active sites.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b00274.

Supporting Results Section, Supporting Figures S1−S11, and Supporting Tables S1−S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(V.G.) E-mail: victor.guallar@bsc.es.
*(M.F.) E-mail: mferrer@icp.csic.es.

ORCID

Gerard Santiago: 0000-0002-0506-3049
Peter N. Golyshin: 0000-0002-5433-0350
Víctor Guallar: 0000-0002-4580-1114
Manuel Ferrer: 0000-0003-4962-4714

Author Contributions

○G.S and M.M-M. contributed equally to this work.

Funding

This project received funding from the European Union's Horizon 2020 research and innovation program [Blue Growth: Unlocking the potential of Seas and Oceans] under Grant Agreement No. [634486] (project acronym INMARE). This research was also supported by the Grants PCIN-2014-107 (within ERA NET IB2 Grant No. ERA-IB-14-030 - MetaCat), PCIN-2017-078 (within the ERA-MarineBiotech grant Pro-Bone), BIO2014-54494-R, CTQ2016-79138-R, and BIO2017-85522-R from the Spanish Ministry of Economy, Industry and Competitiveness. P.N.G. gratefully acknowledges funding from the UK Biotechnology and Biological Sciences Research Council (Grant No. BB/M029085/1). R.B. and P.N.G. acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government. P.N.G. acknowledges the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. The authors gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF). The MALDI-TOF/TOF analysis was performed in the proteomics facility of the Spanish National Center for Biotechnology (CNB-CSIC) that belongs to ProteoRed, PRB2-ISCIII, supported by Grant PT13/0001.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.C. thanks the Spanish Ministry of Economy, Industry and Competitiveness for a Ph.D. fellowship (Grant BES-2015-073829). The author would like to acknowledge Sergio Ciordia and María C. Mena for MALDI-TOF/TOF analysis.

■ REFERENCES

(1) Dufaud, V., and Davis, M. E. (2003) Design of heterogeneous catalysts via multiple active site positioning in organic-inorganic hybrid materials. *J. Am. Chem. Soc. 125*, 9403−9413.

(2) Jing, W., and DeAngelis, P. L. (2000) Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. *Glycobiology 10*, 883−889.

(3) Frickel, E. M., Jemth, P., Widersten, M., and Mannervik, B. (2001) Yeast glyoxalase I is a monomeric enzyme with two active sites. *J. Biol. Chem. 276*, 1845−1849.

(4) Deponte, M., Sturm, N., Mittler, S., Harner, M., Mack, H., and Becker, K. (2007) Allosteric coupling of two different functional active sites in monomeric *Plasmodium falciparum* glyoxalase I. *J. Biol. Chem. 282*, 28419−28430.

(5) Khare, S. D., Kipnis, Y., Greisen, P., Jr., Takeuchi, R., Ashani, Y., Goldsmith, M., Song, Y., Gallaher, J. L., Silman, I., Leader, H., Sussman, J. L., Stoddard, B. L., Tawfik, D. S., and Baker, D. (2012) Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat. Chem. Biol. 8*, 294−300.

(6) Seelig, B., and Szostak, J. W. (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature 448*, 828−831.

(7) Toscano, M. D., Woycechowsky, K. J., and Hilvert, D. (2007) Minimalist active-site redesign: teaching old enzymes new tricks. *Angew. Chem., Int. Ed. 46*, 3212−3236.

(8) Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008)

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

De novo computational design of retro-aldol enzymes. *Science 319*, 1387−1391.
(9) Korendovych, I. V., Kulp, D. W., Wu, Y., Cheng, H., Roder, H., and DeGrado, W. F. (2011) Design of a switchable eliminase. *Proc. Natl. Acad. Sci. U. S. A. 108*, 6823−6827.
(10) Der, B. S., Edwards, D. R., and Kuhlman, B. (2012) Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry 51*, 3933−3940.
(11) Richter, F., Blomberg, R., Khare, S. D., Kiss, G., Kuzin, A. P., Smith, A. J., Gallaher, J., Pianowski, Z., Helgeson, R. C., Grjasnow, A., Xiao, R., Seetharaman, J., Su, M., Vorobiev, S., Lew, S., Forouhar, F., Kornhaber, G. J., Hunt, J. F., Montelione, G. T., Tong, L., Houk, K. N., Hilvert, D., and Baker, D. (2012) Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc. 134*, 16197−16206.
(12) Zastrow, M. L., Peacock, A. F., Stuckey, J. A., and Pecoraro, V. L. (2012) Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem. 4*, 118−123.
(13) Korendovych, I. V., and DeGrado, W. F. (2014) Catalytic efficiency of designed catalytic proteins. *Curr. Opin. Struct. Biol. 27*, 113−121.
(14) Rufo, C. M., Moroz, Y. S., Moroz, O. V., Stöhr, J., Smith, T. A., Hu, X., DeGrado, W. F., and Korendovych, I. V. (2014) Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem. 6*, 303−309.
(15) Raymond, E. A., Mack, K. L., Yoon, J. H., Moroz, O. V., Moroz, Y. S., and Korendovych, I. V. (2015) Design of an allosterically regulated retroaldolase. *Protein Sci. 24*, 561−570.
(16) Moroz, Y. S., Dunston, T. T., Makhlynets, O. V., Moroz, O. V., Wu, Y., Yoon, J. H., Olsen, A. B., McLaughlin, J. M., Mack, K. L., Gosavi, P. M., van Nuland, N. A., and Korendovych, I. V. (2015) New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. *J. Am. Chem. Soc. 137*, 14905−14911.
(17) Aranda, J., Cerqueira, N. M., Fernandes, P. A., Roca, M., Tuñon, I., and Ramos, M. J. (2014) The catalytic mechanism of carboxylesterases: a computational study. *Biochemistry 53*, 5820−5829.
(18) Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., and Golyshin, P. N. (2015) Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransform. 33*, 235−249.
(19) Zastrow, M. L., and Pecoraro, V. L. (2013) Influence of active site location on catalytic activity in de novo-designed zinc metalloenzymes. *J. Am. Chem. Soc. 135*, 5895−5903.
(20) Ross, M. R., White, A. M., Yu, F., King, J. T., Pecoraro, V. L., and Kubarych, K. J. (2015) Histidine orientation modulates the structure and dynamics of a de novo metalloenzyme active site. *J. Am. Chem. Soc. 137*, 10164−10176.
(21) Borrelli, K. W., Cossins, B., and Guallar, V. (2010) Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem. 31*, 1224−1235.
(22) Hernández-Ortega, A., Borrelli, K., Ferreira, P., Medina, M., Martínez, A. T., and Guallar, V. (2011) Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J. 436*, 341−350.
(23) Carlson, H. A., Smith, R. D., Damm-Ganamet, K. L., Stuckey, J. A., Ahmed, A., Convery, M. A., Somers, D. O., Kranz, M., Elkins, P. A., Cui, G., Peishoff, C. E., Lambert, M. H., and Dunbar, J. B., Jr. (2016) CSAR 2014: a benchmark exercise using unpublished data from pharma. *J. Chem. Inf. Model. 56*, 1063−1077.
(24) Santiago, G., de Salas, F., Lucas, M. F., Monza, E., Acebes, S., Martínez, A. T., Camarero, S., and Guallar, V. (2016) Computer-aided laccase engineering: toward biological oxidation of arylamines. *ACS Catal. 6*, 5415−5423.
(25) Lecina, D., Gilabert, J. F., and Guallar, V. (2017) Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Rep. 7*, 8466.
(26) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B 105*, 6474−6487.
(27) Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J., and Friesner, R. A. (2013) Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem. 113*, 2110−2142.
(28) Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., and Shaw, D. E. (2011) How does a drug molecule find its target binding site? *J. Am. Chem. Soc. 133*, 9181−9183.
(29) Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Rico Eguizabal, E., Guerrero, M. C., Yakunin, A. F., and Ferrer, M. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol. 79*, 3553−3562.
(30) Martínez-Martínez, M., Coscolín, C., Santiago, G., Chow, J., Stogios, P., Bargiela, R., Gertler, C., Navarro-Fernández, J., Bollinger, A., Thies, S., Méndez-García, C., Popovic, A., Brown, G., Chernikova, T. N., García-Moyano, A., Bjerga, G. E. K., Pérez-García, P., Hai, T., Del Pozo, M. V., Stokke, R., Steen, I. H., Cui, H., Xu, X., Nocek, B., Alcaide, M., Distaso, M., Mesa, V., Peláez, A. I., Sánchez, J., Buchholz, P. C. F., Pleiss, J., Fernández-Guerra, A. F., Glöckner, F. O., Golyshina, O. V., Yakimov, M. M., Savchenko, A., Jaeger, K.-E., Yakunin, A. F., Streit, W. R., Golyshin, P. N., Guallar, V., Ferrer, M., and The INMARE Consortium (2018) Determinants and prediction of esterase substrate promiscuity patterns. *ACS Chem. Biol. 13*, 225−234.
(31) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 680−685.
(32) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248−254.
(33) Janes, L. E., Löwendahl, C., and Kazlauskas, R. J. (1998) Quantitative screening of hydrolase libraries using pH indicators: Identifying active and enantioselective hydrolases. *Chem. - Eur. J. 4*, 2324−2331.
(34) Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des. 27*, 221−234.
(35) Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. J. (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput. 6*, 1509−1519.
(36) Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem. 68*, 850−858.
(37) Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry* 5th ed., W. H. Freeman, New York; Section 8.4. The Michaelis-Menten model accounts for the kinetic properties of many enzymes. Available from https://www.ncbi.nlm.nih.gov/books/NBK22430/.

DOI: 10.1021/acs.biochem.8b00274
Biochemistry XXXX, XXX, XXX−XXX

SUPPORTING INFORMATION

Rational engineering of multiple active sites in an ester hydrolase

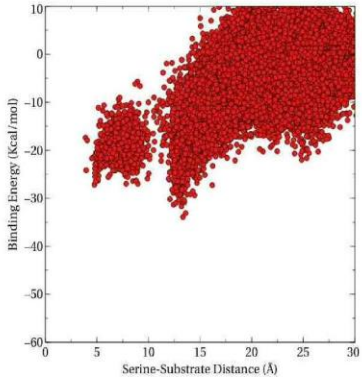
Gerard Santiago[†], Mónica Martínez-Martínez[‡], Sandra Alonso[‡], Rafael Bargiela[‡,¶], Cristina Coscolín[‡], Peter N. Golyshin[§,¥], Víctor Guallar[†,#], Manuel Ferrer[‡]

†Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
‡Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain
¶School of Chemistry, Bangor University, LL57 2UW Bangor, United Kingdom
§School of Biological Sciences, Bangor University, LL57 2UW Bangor, United Kingdom
¥Centre for Environmental Biotechnology, Bangor University, LL57 2UW, United Kingdom
#Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Table of Contents

Figure S1. PELE simulations. Substrate binding energy profile versus the catalytic distance, Ser161 (alcohol oxygen) to ligand ester carbon, for the wild type $EH1_A$ (**a**) and for the $EH1_B$ variant (Glu25Asp and Leu214His double mutant) (**b**).

a



b

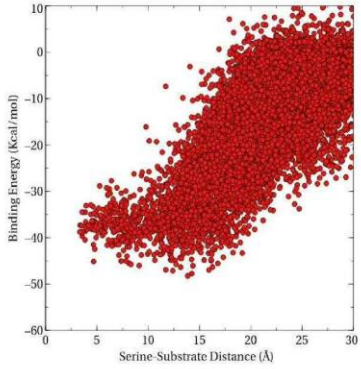


Figure S2. Root mean square deviation (RMSD) for all alpha carbon atoms along the 250ns MD simulation of the double mutant.

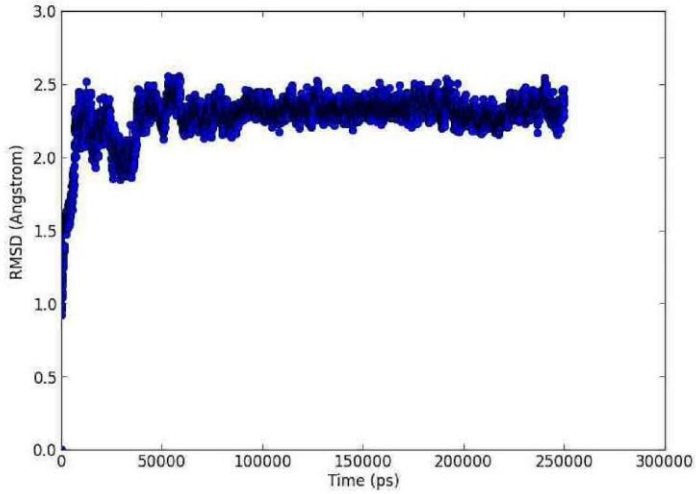
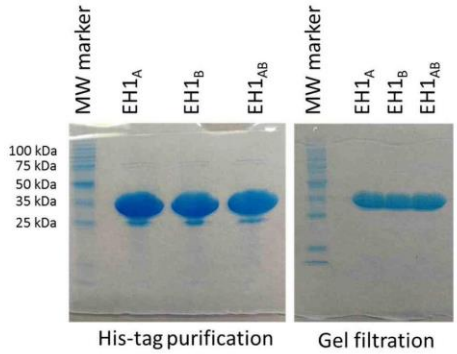
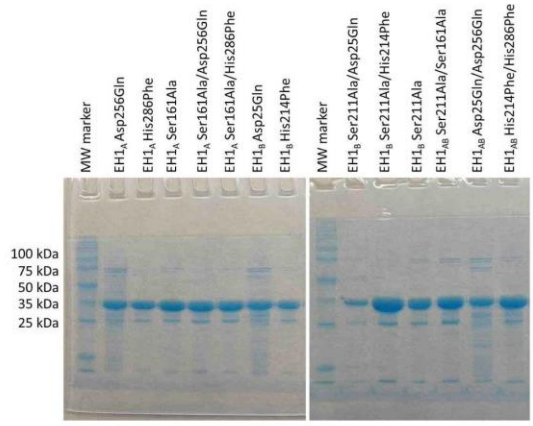


Figure S3. EH1 variants as overexpressed in the active form in *E. coli* at 16ºC and purified by His_6-tag purification and size-exclusion chromatography.

A) Coomassie-stained SDS-PAGE gel in which a total of 20 µg of $EH1_A$, $EH1_B$ and $EH1_{AB}$ purified after the His_6-tag purification step (left) and gel filtration step (right) are shown. As shown the purity of the proteins was high (>95% by densitometry) after His-tag step, and >99% after the gel filtration step, which was further confirmed by mass spectrometry.



B) Coomassie-stained SDS-PAGE gel in which a total of 10 µg of $EH1_A$, $EH1_B$ and $EH1_{AB}$ mutants purified after the His_6-tag purification step; as shown the purity of the proteins was high (>95% by densitometry).



C) Coomassie-stained SDS-PAGE gel in which a total of 10 µg of $EH1_A$, $EH1_B$ and $EH1_{AB}$ mutants purified after the gel filtration step; as shown the purity of the proteins was high (>99% by densitometry), which was further confirmed by mass spectrometry.

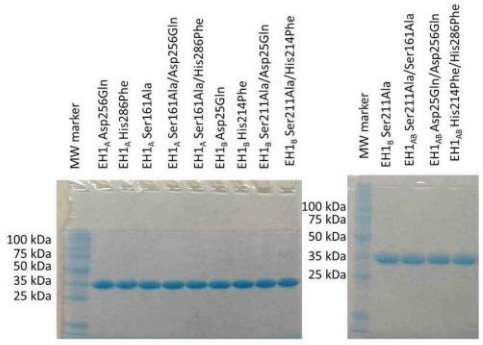


Figure S4. Catalytic activity co-eluted with enzyme peak after Superdex 75 size exclusion chromatography. The His-tagged purified protein solutions (10 mg ml^{-1}) of $EH1_A$, $EH1_B$ and $EH1_{AB}$ were subjected to size-exclusion chromatography and, after recording the absorbance at 280 nm, the collected fractions were tested for activity at 410 nm using *p*NPP, as described in Experimental Section. Y axis represents either the Abs280 nm (left axis) or the Abs410 nm (right axis).

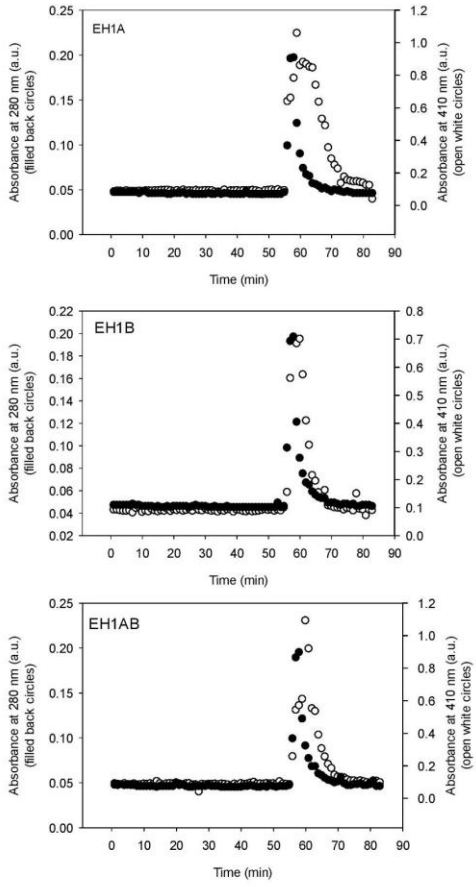
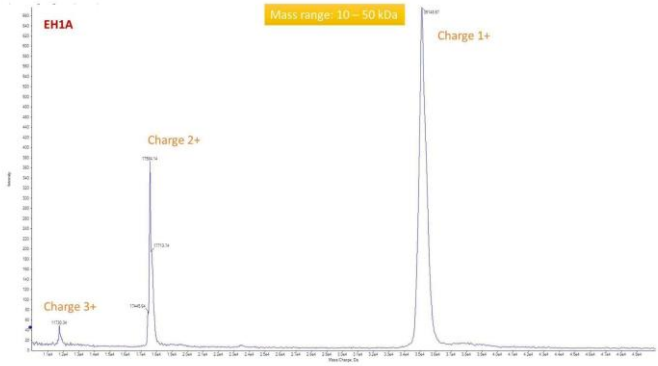


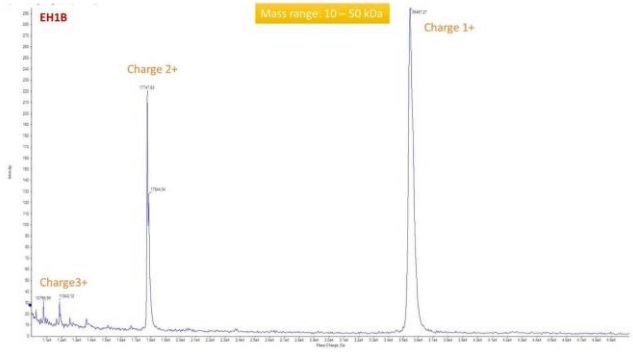
Figure S5. Estimation of molecular mass and purity of $EH1_A$, $EH1_B$ and $EH1_{AB}$ proteins by MALDI-TOF/TOF.

A) MALDI analysis of $EH1_A$ purified protein solution. The mass (m)-to-charge (z) ratios (m/z) for the mass range from 10 to 50 kDa is shown. The MALDI analysis revealed a protein with the estimated molecular mass of ~35kDa, with no apparent contaminants. In the spectra, the peaks corresponding to the double and triple charge of the precursors are indicated.



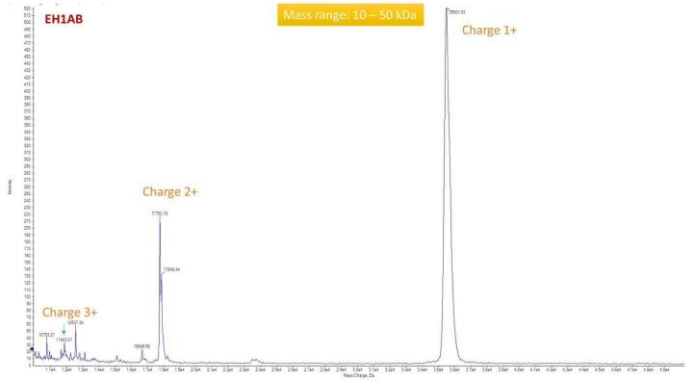
S7

B) MALDI analysis of $EH1_B$ purified protein solution. The mass (m)-to-charge (z) ratios (m/z) for the mass range from 10 to 50 kDa is shown. The MALDI analysis revealed a protein with the estimated molecular mass of ~35kDa, with no apparent contaminants. In the spectra, the peaks corresponding to the double and triple charge of the precursors are indicated. The minor peaks represent degradation products of the $EH1_B$ protein and not contaminants, as determined by peptide fingerprint (see Supporting Results).



S8

C) MALDI analysis of $EH1_{AB}$ purified protein solution. The mass (m)-to-charge (z) ratios (m/z) for the mass range from 10 to 50 kDa is shown. The MALDI analysis revealed a protein with the estimated molecular mass of ~35kDa, with no apparent contaminants. In the spectra, the peaks corresponding to the double and triple charge of the precursors are indicated. The minor peaks represent degradation products of the $EH1_{AB}$ protein and not contaminants, as determined by peptide fingerprint (see Supporting Results).



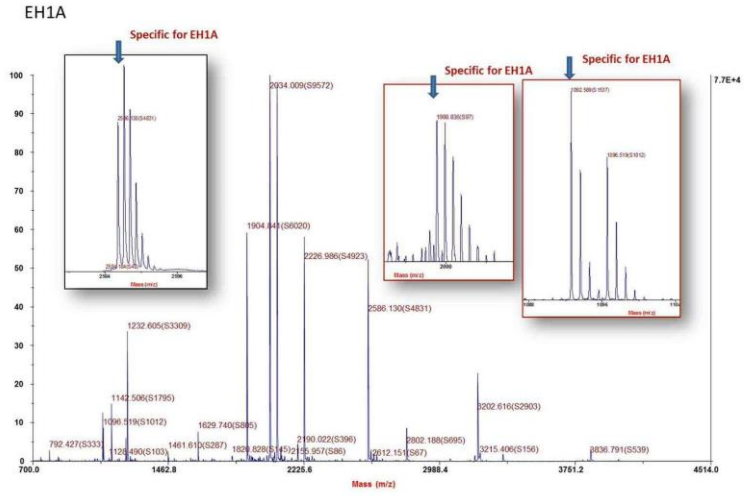
S9

Figure S6. MALDI-TOF/TOF analysis of $EH1_A$ protein solution. A) $EH1_A$ sequence in which the detected peptides are shown in red color, and the specific peptides (not found in $EH1_B$ and $EH1_{AB}$) are underlined. The amino acids constituting the native active site are with yellow background, and the amino acids representing the locations were the new active sites were introduced in $EH1_B$ and $EH1_{AB}$ are in grey background. B) MS/MS chromatogram which shows the mass (m)-to-charge (z) ratios (m/z) of identified peptides, with those being specific for $EH1_A$ being specifically indicated. C) MASCOT search results of the mass of identified peptides, with those being specific for $EH1_A$ indicated with a yellow background.

A)

MLLPETRNLLDLMDAATRGGRPRLETLPHAVGRKAVDKMSEDGEADPPEVAE
VANGGFAGPASEIRFRRYRPLGEAAGLLPTLIYYHGGGFVIGNIETHDSTCRRLA
NKSRCQVISIDYRLAPEHPFPAPIDDGIAAFRHIRDNAESFGADAARLAVGGDSA
GGAMAAVVCQACRDAGETGPAFQMLIYPATDSSRESASRVAFAEGYFLSKAL
MDWFWEAYVPEDTDLTDLRLSPLLATDFTGLPPAFVLTAGYDPLRDEGRAYA
DRLIEAGIKTTYVNYPGTIHGFFSLTRFLSQGLKANDEAAAVMGAHFGT

B)



C)

MATRIX SCIENCE **MASCOT Search Results**

Protein View: U00039

EH1A protein (Manuel Ferrer)

Database:	UsersDB_Home
Score:	875
Expect:	9.3e-086
Monoisotopic mass (M$_r$):	35811
Calculated pI:	5.26

Sequence similarity is available as an NCBI BLAST search of U00039 against nr.

Search parameters

Enzyme:	Trypsin/P; cuts C-term side of KR.
Fixed modifications:	Methylthio (C)
Variable modifications:	Oxidation (M)

Protein sequence coverage: 76%

Matched peptides shown in ***bold red***.

```
   1 MAHDDDDHVD DDDKMLLPET RNLLDLMDAA TRGGPPRLET LPNAVGRKAV
  51 DKMSEDGEAD PPEVAEVANG GFAGPASEIR FRRYRPLGEA AGLLPTLIYY
 101 HGGGFVIGNI ETHDSTCRRL ANKSRCQVIS IDYRLAPEHP FPAPIDDGIA
 151 AFRHIRDNAE SFGADAARLA VGGDSAGGAM AAVVCQACRD AGETGPAFQM
 201 LIYPATDSSR ESASRVAFAE GYFLSKALMD WFWEAYVPED TDLTDLRLSP
 251 LLATDFTGLP PAFVLTAGYD PLRDEGRAYA DRLIEAGIKT TYVNYPGTIH
 301 GFFSLTRFLS QGLKANDEAA AVMGAHFGT
```

Unformatted sequence string: 329 residues (for pasting into other applications).

Sort by ◉ residue number ○ increasing mass ○ decreasing mass
Show ◉ matched peptides only ○ predicted peptides also

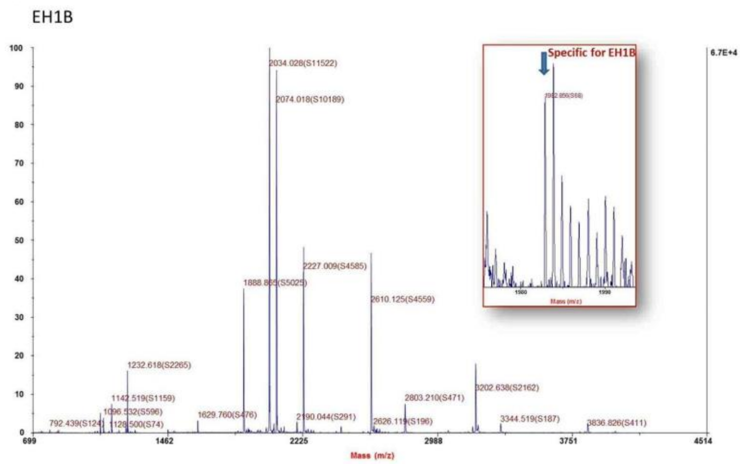
Query	Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Peptide
75	2 – 21	2440.09	2439.09	2439.13	-16.2	1		M.AHDDDDHVDDDDKMLLPETR.N
19	22 – 32	1232.61	1231.60	1231.62	-20.6	0		R.NLLDLMDAATR.G
20	22 – 32	1232.61	1231.60	1231.62	-20.6	0	68	R.NLLDLMDAATR.G
21	22 – 32	1248.60	1247.59	1247.62	-23.7	0		R.NLLDLMDAATR.G + Oxidation (M)
9	38 – 47	1092.59	1091.58	1091.61	-24.6	0		R.LETLPNAVGR.K
90	49 – 80	3215.41	3214.40	3214.47	-23.1	1		K.AVDKMSEDGEADPPEVAEVANGGFAGPASEIR.F
84	53 – 80	2802.19	2801.18	2801.24	-23.1	0		K.MSEDGEADPPEVAEVANGGFAGPASEIR.F
85	53 – 80	2802.19	2801.18	2801.24	-23.1	0	90	K.MSEDGEADPPEVAEVANGGFAGPASEIR.F
13	126 – 134	1142.51	1141.50	1141.53	-24.0	0		R.CQVISIDYR.L
53	135 – 153	2034.01	2033.00	2033.04	-17.4	0		R.LAPEHPFPAPIDDGIAAFR.H
54	135 – 153	2034.01	2033.00	2033.04	-17.4	0	118	R.LAPEHPFPAPIDDGIAAFR.H
29	154 – 168	1629.74	1628.73	1628.77	-20.4	1		R.HIRDNAESFGADAAR.L
12	157 – 168	1223.50	1222.49	1222.52	-23.5	0		R.DNAESFGADAAR.L
47	169 – 189	1998.84	1997.83	1997.86	-16.4	0		R.LAVGGDSAGGAMAAVVCQACR.D
69	190 – 210	2226.99	2225.98	2226.03	-21.5	0		R.DAGETGPAFQMLIYPATDSSR.E
70	190 – 210	2226.99	2225.98	2226.03	-21.5	0	160	R.DAGETGPAFQMLIYPATDSSR.E
71	190 – 210	2242.98	2241.98	2242.02	-19.9	0		R.DAGETGPAFQMLIYPATDSSR.E + Oxidation (M)
18	216 – 226	1231.61	1230.60	1230.63	-22.8	0		R.VAFAEGYFLSK.A
77	227 – 247	2586.13	2585.12	2585.18	-21.6	0		K.ALMDWFWEAYVPEDTDLTDLR.L
78	227 – 247	2586.13	2585.12	2585.18	-21.6	0	156	K.ALMDWFWEAYVPEDTDLTDLR.L
79	227 – 247	2602.13	2601.12	2601.17	-20.3	0		K.ALMDWFWEAYVPEDTDLTDLR.L + Oxidation (M)
88	248 – 277	3202.62	3201.61	3201.67	-19.4	1		R.LSPLLATDFTGLPPAFVLTAGYDPLRDEGR.A
89	248 – 277	3202.62	3201.61	3201.67	-19.4	1	17	R.LSPLLATDFTGLPPAFVLTAGYDPLRDEGR.A
61	290 – 307	2074.00	2072.99	2073.03	-19.3	0		K.TIYVNYPGTIHGFFSLTR.F
62	290 – 307	2074.00	2072.99	2073.03	-19.3	0	115	K.TIYVNYPGTIHGFFSLTR.F

Figure S7. MALDI-TOF/TOF analysis of $EH1_B$ protein solution. A) $EH1_B$ sequence in which the detected peptides are shown in red color, and the specific peptides (not found in $EH1_A$ and $EH1_{AB}$) are underlined. The amino acids constituting the native active site are with yellow background, and the amino acids representing the locations were the new active sites were introduced in $EH1_B$ and $EH1_{AB}$ are in grey background. B) MS/MS chromatogram which shows the mass (m)-to-charge (z) ratios (m/z) of identified peptides, with those being specific for $EH1_B$ being specifically indicated. C) MASCOT search results of the mass of identified peptides, with those being specific for $EH1_B$ indicated with a yellow background.

A)

MLLPETRNLLDLMDAATRGGRPRLDTLPHAVGRKAVDKMSEDGEADPPEVAE
VANGGFAGPASEIRFRRYRPLGEAAGLLPTLIYYHGGGFVIGNIETHDSTCRRLA
NKSRCQVISIDYRLAPEHPFPAPIDDGIAAFRHIRDNAESFGADAARLAVGGDAA
GGAMAAVVCQACRDAGETGPAFQMLIYPATDSSRESASRVAFAEGYFLSKAH
MDWFWEAYVPEDTDLTDLRLSPLLATDFTGLPPAFVLTAGYDPLRDEGRAYA
DRLIEAGIKTTYVNYPGTIHGFFSLTRFLSQGLKANDEAAAVMGAHFGT

B)



C)

MATRIX SCIENCE MASCOT Search Results

Protein View: U00040

EH1B-1 / EH1B-2 / EH1B-3 protein (Manuel Ferrer)

Database:	UsersDB_Home
Score:	869
Expect:	3.7e-085
Monoisotopic mass (M_r):	35804
Calculated pI:	5.32

Sequence similarity is available as **an NCBI BLAST search of U00040 against nr**.

Search parameters

Enzyme:	Trypsin/P; cuts C-term side of KR.
Fixed modifications:	**Methylthio (C)**
Variable modifications:	**Oxidation (M)**

Protein sequence coverage: 87%

Matched peptides shown in ***bold red***.

```
  1 MAHDDDDDVD DDDKMLLPET RMLLDLMDAA TRGGRPRLDT LPHAVGREAV
 51 DKMSEDGEAD PPEVAEVANG GFAGPASEIR FRRYRPLGEA AGLLPTLIYY
101 HGGGFVIGNI ETHDSTCRRL ANKSRCQVIS IDYRLAPEHP FPAPIDDGIA
151 AFRHIRDNAE SFGADAARLA VGGDAAGGAM AAVVCQACRD AGETGPAFQM
201 LIYPATDSSR ESASRVAFAE GYFLSKAMMD WFWEAYVPED TDLTDLRLSP
251 LLATDFTGLP PAFVLTAGYD PLRDEGRAYA DRLIEAGIKT TYVNYPGTIH
301 GFFSLTRFLS QGLKAHDEAA AVMGAMFGT
```

Unformatted sequence string: **329 residues** (for pasting into other applications).

Sort by ◉ residue number ○ increasing mass ○ decreasing mass
Show ◉ matched peptides only ○ predicted peptides also

Query	Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Peptide
51	2 – 21	2440.11	2439.10	2439.13	-10.7	1		M.AHDDDDDVDDDDKMLLPETR.M
12	22 – 32	1232.62	1231.61	1231.62	-10.0	0		R.MLLDLMDAATR.G
13	22 – 32	1232.62	1231.61	1231.62	-10.0	0	22	R.MLLDLMDAATR.G
14	22 – 32	1248.60	1247.60	1247.62	-17.1	0		R.MLLDLMDAATR.G + Oxidation (M)
5	38 – 47	1078.59	1077.58	1077.59	-13.4	0		R.LDTLPHAVGR.K
10	49 – 80	3215.43	3214.42	3214.47	-16.3	1		K.AVDKMSEDGEADPPEVAEVANGGFAGPASEIR.F
63	53 – 80	2802.21	2801.20	2801.24	-14.3	0		K.MSEDGEADPPEVAEVANGGFAGPASEIR.F
64	53 – 80	2002.21	2801.20	2801.24	-14.3	0	65	K.MSEDGEADPPEVAEVANGGFAGPASEIR.F
74	84 – 110	3836.83	3835.82	3835.80	-16.4	1		R.YRPLGEAAGLLPTLIYYHGGGFVIGNIETHDSTCR.R
8	126 – 134	1142.52	1141.51	1141.53	-12.8	0		R.CQVISIDYR.L
14	135 – 153	2034.03	2033.02	2033.04	-7.98	0		R.LAPEHPFPAPIDDGIAAFR.H
35	135 – 153	2034.03	2033.02	2033.04	-7.98	0	112	R.LAPEHPFPAPIDDGIAAFR.H
19	154 – 168	1629.76	1628.75	1628.77	-8.15	1		R.HIRDNAESFGADAAR.L
10	157 – 168	1223.52	1222.51	1222.52	-10.5	0		R.DNAESFGADAAR.L
28	169 – 189	1982.86	1981.85	1981.87	-8.79	0		R.LAVGGDAAGGAMAAVVCQACR.D
48	190 – 210	2227.01	2226.00	2226.03	-11.0	0		R.DAGETGPAFQMLIYPATDSSR.E
49	190 – 210	2227.01	2226.00	2226.03	-11.0	0	162	R.DAGETGPAFQMLIYPATDSSR.E
50	190 – 210	2243.00	2241.99	2242.02	-13.2	0		R.DAGETGPAFQMLIYPATDSSR.E + Oxidation (M)
11	216 – 226	1231.62	1230.61	1230.63	-12.5	0		R.VAFAEGYFLSK.A
56	227 – 247	2610.12	2609.12	2609.15	-13.8	0		K.AMMDWFWEAYVPEDTDLTDLR.L
57	227 – 247	2610.12	2609.12	2609.15	-13.8	0	161	K.AMMDWFWEAYVPEDTDLTDLR.L
58	227 – 247	2626.12	2625.11	2625.15	-13.9	0		K.AQMDWFWEAYVPEDTDLTDLR.L + Oxidation (M)
68	248 – 277	3202.64	3201.63	3201.67	-12.5	1		R.LSPLLATDFTGLPPAFVLTAGYDPLRDEGR.A
69	248 – 277	3202.64	3201.63	3201.67	-12.5	1	24	R.LSPLLATDFTGLPPAFVLTAGYDPLRDEGR.A
11	290 – 307	2074.02	2073.01	2073.03	-10.1	0		K.TTYVNYPGTIHGFFSLTR.F
12	290 – 307	2074.02	2073.01	2073.03	-10.1	0	112	K.TTYVNYPGTIHGFFSLTR.F
17	315 – 329	1461.63	1460.62	1460.64	-8.08	0		K.AHDEAAAVMGAMFGT.-

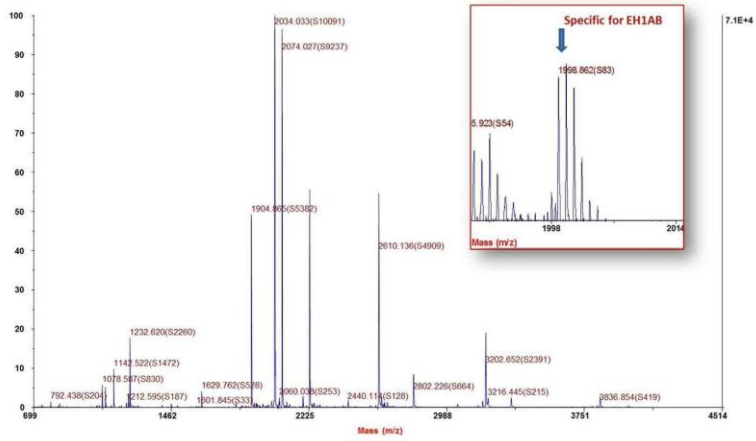
Figure S8. MALDI-TOF/TOF analysis of $EH1_{AB}$ protein solution. A) $EH1_{A>B}$ sequence in which the detected peptides are shown in red color, and the specific peptides (not found in $EH1_A$ and $EH1_B$) are underlined. The amino acids constituting the native active site are with yellow background, and the amino acids representing the locations were the new active sites were introduced in $EH1_B$ and $EH1_{AB}$ are in grey background. B) MS/MS chromatogram which shows the mass (m)-to-charge (z) ratios (m/z) of identified peptides, with those being specific for $EH1_{AB}$ being specifically indicated. C) MASCOT search results of the mass of identified peptides, with those being specific for $EH1_{AB}$ indicated with a yellow background.

A)

MLLPETRNLLDLMDAATRGGRPRLDTLPHAVGRKAVDKMSEDGEADPPEVAE
VANGGFAGPASEIRFRRYRPLGEAAGLLPTLIYYHGGGFVIGNIETHDSTCRRLA
NKSRCQVISIDYRLAPEHPFPAPIDDGIAAFRHIRDNAESFGADAARLAVGGDSA
GGAMAAVVCQACRDAGETGPAFQMLIYPATDSSRESASRVAFAEGYFLSKAH
MDWFWEAYVPEDTDLTDLRLSPLLATDFTGLPPAFVLTAGYDPLRDEGRAYA
DRLIEAGIKTTYVNYPGTIHGFFSLTRFLSQGLKANDEAAAVMGAHFGT

B)

EH1AB



C)

MASCOT Search Results

Protein View: U00041

EH1AB protein (Manuel Ferrer)

Database:	UsersDB_Home
Score:	912
Expect:	1.9e-089
Monoisotopic mass (M_r):	35820
Calculated pI:	5.32

Sequence similarity is available as **an NCBI BLAST search of U00041 against nr**.

Search parameters

Enzyme:	Trypsin/P; cuts C-term side of KR.
Fixed modifications:	**Methylthio (C)**
Variable modifications:	**Oxidation (M)**

Protein sequence coverage: 76%

Matched peptides shown in ***bold red***.

```
   1 MAHHHHHHVD DDDRMLLPET RMLLDLMDAA TRGGRPRLDT LPMAVGRKAV
  51 DRMSEDGEAD PPEVAEVAHG GFAGPASEIR FRRYRPLGEA AGLLPTLIYY
 101 HGGGFVIGNI ETHDSTCRRL ANKSRCQVIS IDYRLAPEHP FPAPIDDGIA
 151 AFRHIRDNAE SFGADAARLA VGGDSAGGAM AAVVCQACRD AGETGPAFQH
 201 LIYPATDSSR ESASPVAFAE GYFLSKAHMD WFWEAYVPED TDLTDLRLSP
 251 LLATDFTGLP PAFVLTAGYD PLRDEGRAYA DRLIEAGIKT TYVNYPGTIH
 301 GFFSLTRFLS QGLKAHDEAA AVHGAHFGT
```

Unformatted sequence string: **329 residues** (for pasting into other applications).

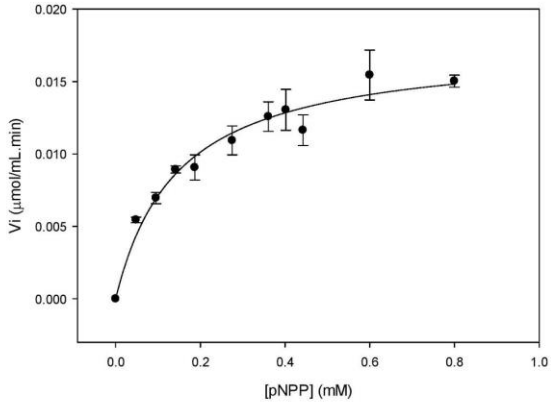
Sort by ◉ residue number ○ increasing mass ○ decreasing mass
Show ◉ matched peptides only ○ predicted peptides also

Query	Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Peptide
59	2 – 21	2440.11	2439.11	2439.13	-7.52	1		M.AHHHHHHVDDDDRMLLPETR.M
47	15 – 32	2073.02	2072.01	2072.08	-29.4	1		K.MLLPETRMLLDLMDAATR.G
12	22 – 32	1232.62	1231.61	1231.62	-8.30	0		R.MLLDLMDAATR.G
13	22 – 32	1232.62	1231.61	1231.62	-8.30	0	55	R.MLLDLMDAATR.G
14	22 – 32	1248.61	1247.60	1247.62	-14.9	0		R.MLLDLMDAATR.G + Oxidation (M)
4	38 – 47	1078.59	1077.58	1077.59	-12.4	0		R.LDTLPMAVGR.K
74	49 – 80	3215.44	3214.44	3214.47	-11.3	1		K.AVDRMSEDGEADPPEVAEVAHGGFAGPASEIR.F
68	53 – 80	2802.23	2801.22	2801.24	-9.52	0		K.MSEDGEADPPEVAEVAHGGFAGPASEIR.F
69	53 – 80	2802.23	2801.22	2801.24	-9.52	0	106	K.MSEDGEADPPEVAEVAHGGFAGPASEIR.F
7	126 – 134	1142.52	1141.51	1141.53	-9.87	0		R.CQVISIDYR.L
41	135 – 153	2034.03	2033.03	2033.04	-5.47	0		R.LAPEHPFPAPIDDGIAAFR.H
42	135 – 153	2034.03	2033.03	2033.04	-5.47	0	120	R.LAPEHPFPAPIDDGIAAFR.H
21	154 – 168	1629.76	1628.76	1628.77	-6.43	1		R.HIRDNAESFGADAAR.L
10	157 – 168	1223.52	1222.51	1222.52	-9.73	0		R.DNAESFGADAAR.L
15	169 – 189	1998.86	1997.85	1997.86	-3.52	0		R.LAVGGDSAGGAMAAVVCQACR.D
55	190 – 210	2227.02	2226.01	2226.03	-7.85	0		R.DAGETGPAFQMLIYPATDSSR.E
56	190 – 210	2227.02	2226.01	2226.03	-7.85	0	170	R.DAGETGPAFQMLIYPATDSSR.E
57	190 – 210	2243.01	2242.00	2242.02	-7.58	0		R.DAGETGPAFQMLIYPATDSSR.E + Oxidation (M)
11	216 – 226	1231.62	1230.61	1230.63	-12.5	0		R.VAFAEGYFLSK.A
61	227 – 247	2610.14	2609.13	2609.15	-9.34	0		K.AHDDWFWEAYVPEDTDLTDLR.L
62	227 – 247	2610.14	2609.13	2609.15	-9.34	0	152	K.AHDDWFWEAYVPEDTDLTDLR.L
63	227 – 247	2626.13	2625.12	2625.15	-10.0	0		K.AHDDWFWEAYVPEDTDLTDLR.L + Oxidation (M)
72	248 – 277	3202.65	3201.65	3201.67	-8.12	1		R.LSPLLATDFTGLPPAFVLTAGYDPLRDEGR.A
73	248 – 277	3202.65	3201.65	3201.67	-8.12	1	30	R.LSPLLATDFTGLPPAFVLTAGYDPLRDEGR.A
48	290 – 307	2074.03	2073.02	2073.03	-6.11	0		K.TTYVNYPGTIHGFFSLTR.F
49	290 – 307	2074.03	2073.02	2073.03	-6.11	0	96	K.TTYVNYPGTIHGFFSLTR.F
19	315 – 329	1461.64	1460.63	1460.64	-4.86	0		K.AHDEAAAVHGAHFGT.-

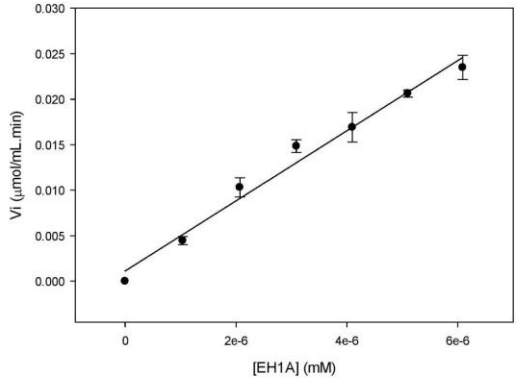
Figure S9. Experimental characterization of native and active enzyme designs. Dependence of reaction velocity on substrate concentration, and dependence of reaction velocity on enzyme concentration for *p*NPP and phenyl propionate are shown in (A) for $EH1_A$, in (B) for $EH1_B$ and in (C) for $EH1_{AB}$. The K_M and k_{cat} values are summarized in Supporting Table S1. Reactions conditions for all experiments using *p*NPP were 40 mM HEPES pH 7.0, and substrate and esterase at the indicated concentration; for K_m determinations, the amount of protein was 0.1, 16 and 1.4 μg for $EH1_A$, $EH1_B$ and $EH1_{AB}$, respectively, and for k_{cat} determinations, the amount of *p*NPP was 1 mM. Reactions conditions for all experiments using phenyl propionate were 5 mM EPPS buffer pH 8.0 with 0.45 mM Phenol Red (used as a pH indicator), and substrate and esterase at the indicated concentration; for K_m determinations, the amount of protein was 0.1, 16 and 1.4 μg for $EH1_A$, $EH1_B$ and $EH1_{AB}$, respectively, and for k_{cat} determinations, the amount of phenyl propionate was 1 mM. In all cases reactions, in triplicates, were performed at 30ºC. For K_m determinations results were fitted to simple hyperbolic Michaelis-Menten kinetics, except for the $EH1_{AB}$ where the data were fitted to sigmoidal Michaelis-Menten kinetics. D) Hydrolytic activity of $EH1_A$ mutant with Ser161Ala mutation, $EH1_B$ mutant with Ser211Ala mutation, and $EH1_{AB}$ mutant with Ser211Ala/Ser161Ala mutations. E) Hydrolytic activity of $EH1_A$ mutant with Asp256Gln, $EH1_B$ mutant with Asp25Gln mutation, and $EH1_{AB}$ mutant with Asp25Gln/Asp256Gln mutations, compared to $EH1_A$, $EH1_B$ and $EH1_{AB}$. F) Hydrolytic activity of $EH1_A$ mutant with His286Phe mutation, $EH1_B$ mutant with His214Phe mutation, and $EH1_{AB}$ mutant with His214Phe/His286Phe mutations, compared to $EH1_A$, $EH1_B$ and $EH1_{AB}$. The figures in Panels D) to F) represent the Δabsorbance/min, measured at 410 nm for the hydrolysis of *p*NPP or at 550 nm for the hydrolysis of phenyl propionate, over the assay time of each of the mutants compared to $EH1_A$, $EH1_B$ and $EH1_{AB}$. In panels D) to F), no appreciable increase of absorbance (at 410 nm; hydrolysis is monitored by the increase in absorbance due to formation of *p*-nitrophenoxide) or decrease (at 550 nm; hydrolysis is monitored by the decrease in absorbance due to the formation of acid) absorbance was detected for any of the mutants.

A) $EH1_A$

K_M determination *p*NPP: 146 ± 27 μM

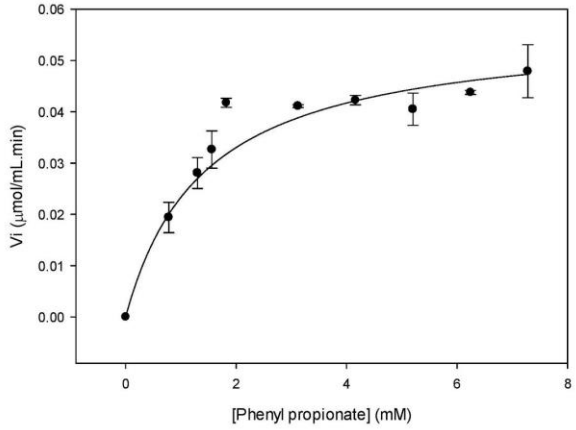


k_{cat} determination *p*NPP: 3853 ± 23 min^{-1}

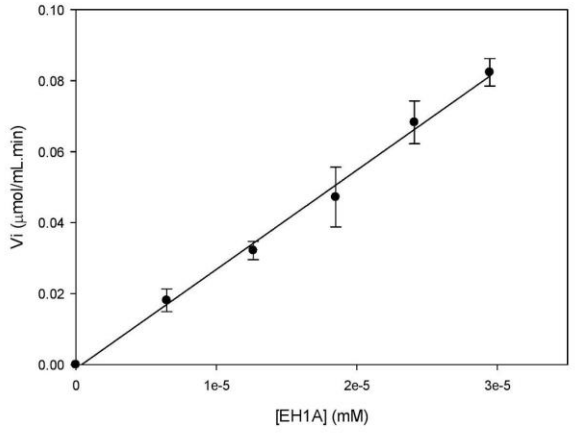


A) $EH1_A$

K_M determination phenyl propionate: 964 ± 24 μM

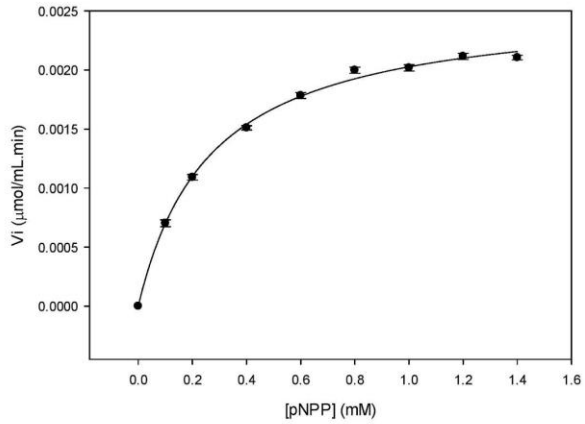


k_{cat} determination phenyl propionate: 2797± 98 min^{-1}

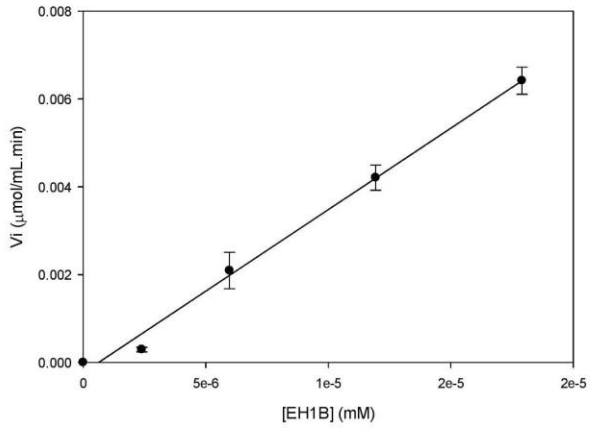


B) $EH1_B$

K_M determination *p*NPP: 267± 20 μM

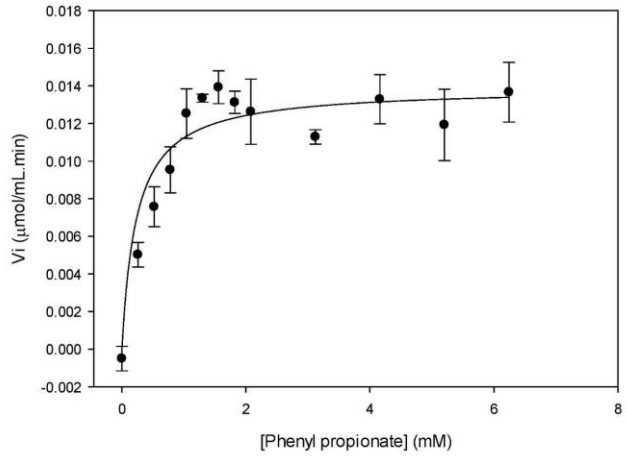


k_{cat} determination *p*NPP: 263 ± 16 x10^{-3} min^{-1}

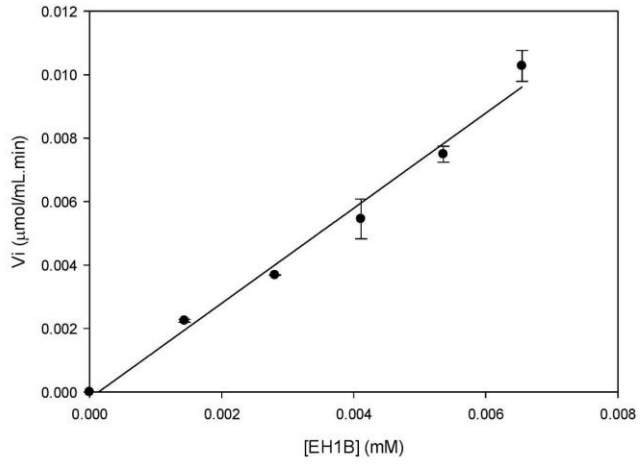


B) $EH1_B$

K_M determination phenyl propionate: 752± 20 μM



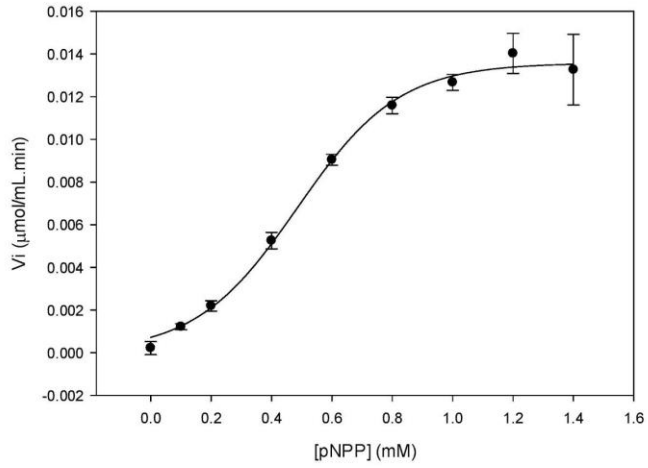
k_{cat} determination phenyl propionate: 1.49 ± 0.93 min^{-1}



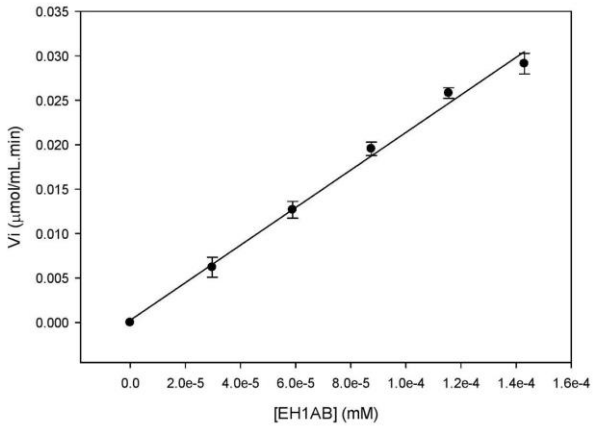
S20

C) $EH1_{AB}$

K_M determination *p*NPP: 485 ± 21 μM

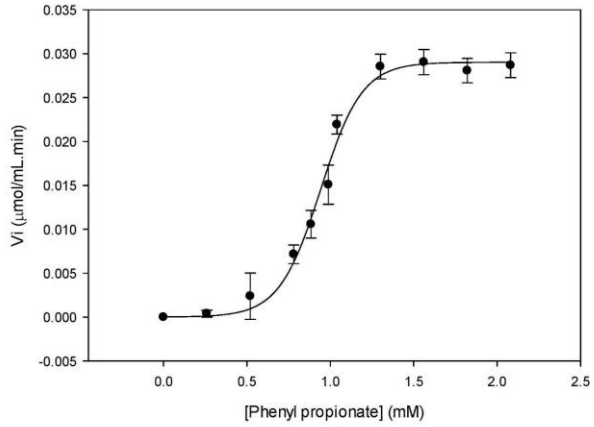


k_{cat} determination *p*NPP: 211 ± 8 min^{-1}

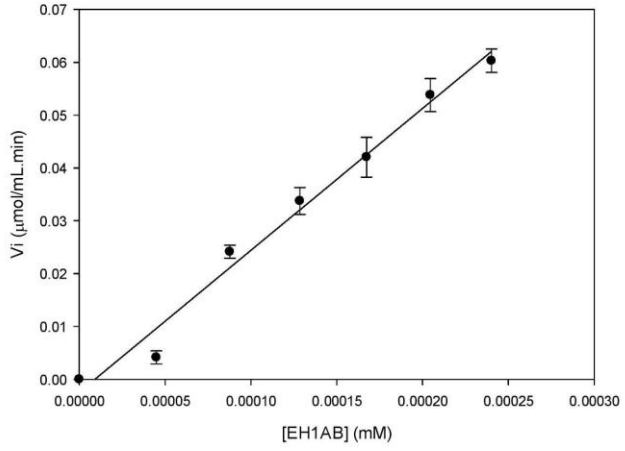


C) $EH1_{AB}$

K_M determination phenyl propionate: 942 ± 17 μM

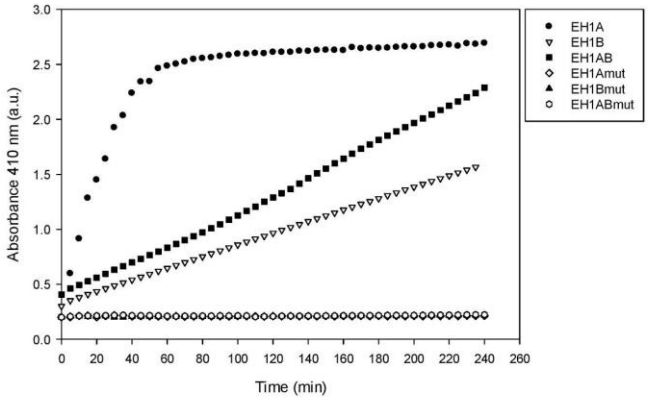


k_{cat} determination phenyl propionate: 267 ± 15 min^{-1}

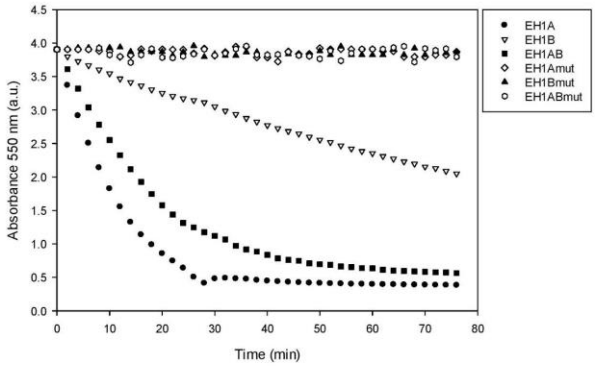


D) $EH1_A$ mutant with Ser161Ala mutation (EH1Amut), $EH1_B$ mutant with Ser211Ala mutation (EH1Bmut), and $EH1_{AB}$ mutant with Ser211Ala/Ser161Ala mutations (EH1ABmut), compared to $EH1_A$, $EH1_B$ and $EH1_{AB}$.

Abs410 nm for *p*NPP hydrolysis

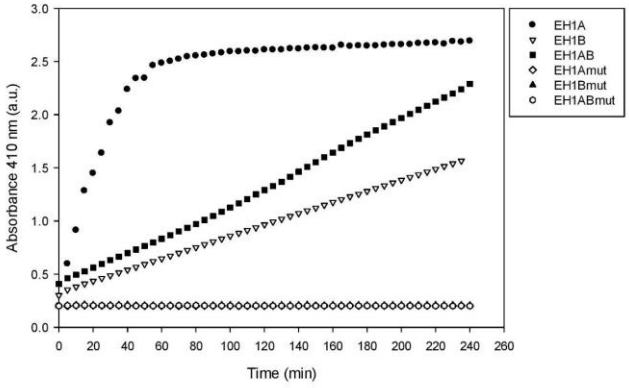


Abs550 nm for phenyl propionate hydrolysis

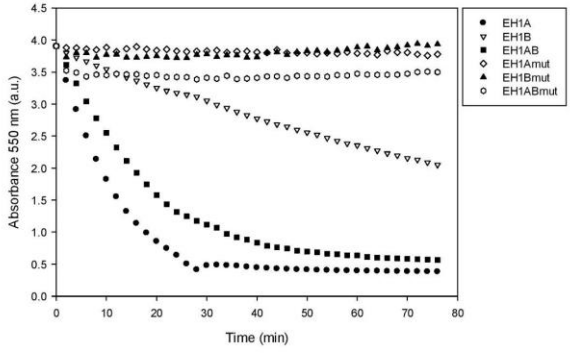


E) $EH1_A$ mutant with Asp256Gln (EH1Amut), $EH1_B$ mutant with Asp25Gln mutation (EH1Bmut), and $EH1_{AB}$ mutant with Asp25Gln/Asp256Gln mutations (EH1ABmut), compared to $EH1_A$, $EH1_B$ and $EH1_{AB}$.

Abs410 nm for *p*NPP hydrolysis

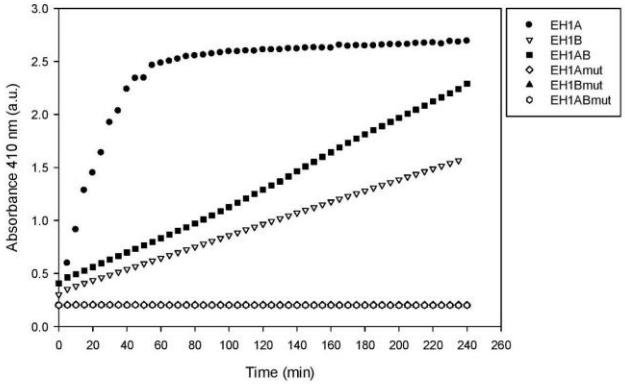


Abs550 nm for phenyl propionate hydrolysis



F) $EH1_A$ mutant with His286Phe mutation (EH1Amut), $EH1_B$ mutant with His214Phe mutation (EH1Bmut), and $EH1_{AB}$ mutant with His214Phe/His286Phe mutations (EH1ABmut), compared to $EH1_A$, $EH1_B$ and $EH1_{AB}$.

Abs410 nm for *p*NPP hydrolysis



Abs550 nm for phenyl propionate hydrolysis

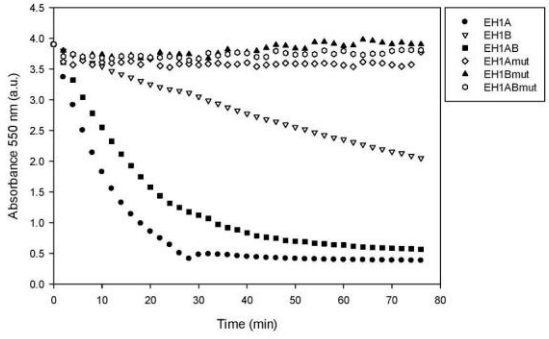
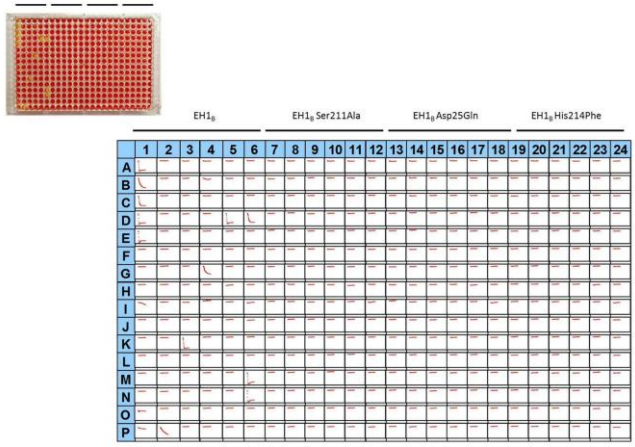


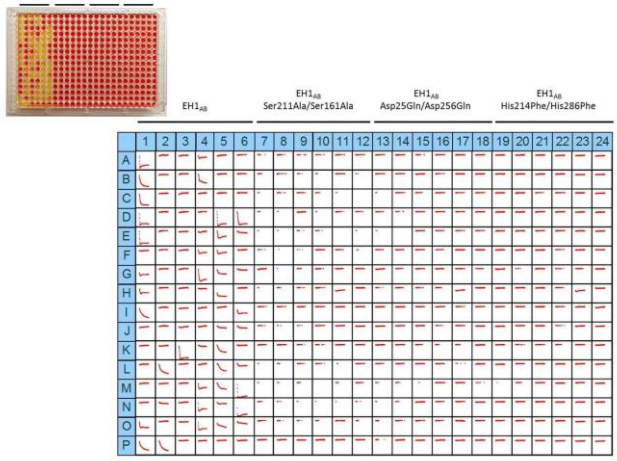
Figure S10. Effect of single and dual mutations in the hydrolytic activity of $EH1_B$ and $EH1_{AB}$. Images representing 384-well plates in which the follow up (for 7 h) hydrolysis of 96 different esters, catalyzed by $EH1_B$, $EH1_{AB}$ and their mutants ($EH1_B$ with Ser211Ala, Asp25Gln and His214Phe substitutions; and $EH1_{AB}$ with Ser211Ala/Ser161Ala, Asp25Gln/Asp256Gln and His214Phe/His286Phe substitutions), are shown. A total of 96 esters, dispensed in replicates in 384-plate, were tested (see Supporting Table S1). Reaction conditions for all experiments were as follows. Each of the 96 esters was dispensed in the 384 well plate with the help of a PRIMADIAG liquid-handling robot. A total of 2 µl of each ester stock solution (at a concentration of 25 mg ml^{-1} in acetonitrile or dimethyl sulfoxide (DMSO), was added to each well using a PRIMADIAG liquid-handling robot (EYOWN TECHNOLOGIES SL, Madrid, Spain). The 384-well plate was filled with 20 µl of 5 mM EPPS buffer, pH 8.0, containing 0.912 mM Phenol Red (used as a pH indicator) using a QFill3 microplate filler. The final ester concentration in each well was 1.14 mg ml^{-1}, and the final concentration of Phenol Red was 0.45 mM. A total of 16 and 1.4 µg $EH1_B$ and $EH1_{AB}$, respectively, or 100 µg of each mutant (all prepared in 40 mM HEPES buffer pH 7.0), was immediately added to each well using an Eppendorf Repeater M4 pipette (Eppendorf, Hamburg, Germany). The total reaction volume was 44 µl, with 4.5% (v/v) acetonitrile or DMSO in the reaction mixture. Reactions were incubated at 30 °C in a Synergy HT Multi-Mode Microplate Reader, and ester hydrolysis was measured spectrophotometrically in continuous mode at 550 nm for a total time of 24 h. If activity occurred, a yellow color is evident in the well due to acid formation because ester hydrolysis, and a decrease in absorbance at 550 nm occurs. The esters in each of the well are described in Table S1. The conditions used (ester concentration, buffer pH and concentration, and enzyme concentration) are those recommended to ensure detection of pH shift.[33] Panel A shows the follow-up hydrolysis of the 96 esters for $EH1_B$, and variants containing Ser211Ala, Asp25Gln and His214Phe substitutions. Panel B shows the follow-up hydrolysis of the 96 esters for $EH1_{AB}$, and variants containing Ser211Ala/Ser161Ala, Asp25Gln/Asp256Gln and His214Phe/His286Phe substitutions. As shown any of the mutants show appreciable activity (below detection limit) for any of the 96 ester tested.

A) Follow-up hydrolysis of the 96 esters for $EH1_B$, and variants containing Ser211Ala, Asp25Gln and His214Phe substitutions.



S27

B) Follow-up hydrolysis of the 96 esters for $EH1_{AB}$, and variants containing Ser211Ala/Ser161Ala, Asp25Gln/Asp256Gln and His214Phe/His286Phe substitutions.



S28

Figure S11. pH dependence of activity for representative enzymes. Relative activity is plotted versus pH. The plots are normalized by setting the values of each enzyme at the optimum pH to 100%. Britton and Robinson buffer (50 mM; pH 4.0-9.5) was used for the assays which were performed, in triplicates, as described in Experimental section.

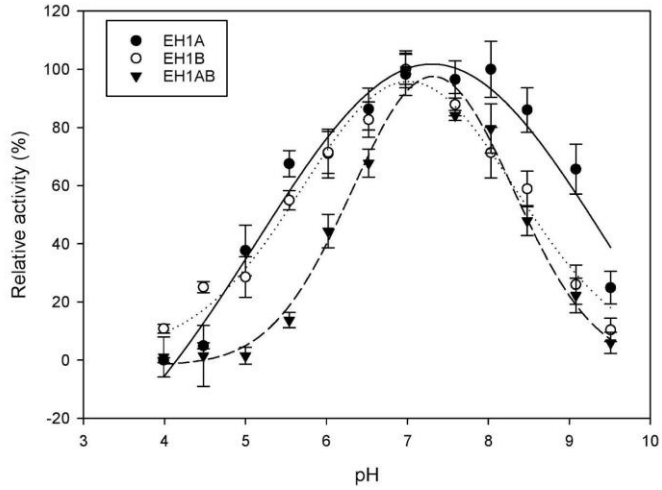


Table S1 List of the 96 esters for which hydrolytic activity was tested by the pH indicator assay. The positioning of each of the esters in the 384 well plates (four replicates) in Supporting Figure S10 is specifically shown.

Well	Well	Well	Well	Substrate esterases/lipases
A1	A7	A13	A19	1-Napthyl acetate
B1	B7	B13	B19	1-Naphthyl butyrate
C1	C7	C13	C19	Glyceryl triacetate
D1	D7	D13	D19	Glyceryl tripropionate
E1	E7	E13	E19	Glyceryl tributyrate
F1	F7	F13	F19	Glyceryl trioctanoate
G1	G7	G13	G19	Triolein (olive oil)
H1	H7	H13	H19	Hexyl acetate (caproyl acetate)
I1	I7	I13	I19	Octyl acetate (capryloyl acetate)
J1	J7	J13	J19	Dodecanoyl acetate (dodecyl acetate)
K1	K7	K13	K19	Pentadecyl acetate
L1	L7	L13	L19	Ethyl acetate
M1	M7	M13	M19	Ethyl propionate
N1	N7	N13	N19	Ethyl butyrate
O1	O7	O13	O19	Ethyl hexanoate (or ethyl caproate)
P1	P7	P13	P19	Ethyl octanoate (or ethyl caprylate)
A2	A8	A14	A20	Ethyl decanoate
B2	B8	B14	B20	Ethyl dodecanoate (ethyl laurate)
C2	C8	C14	C20	Ethyl myristate
D2	D8	D14	D20	Ethyl benzoate
E2	E8	E14	E20	(1R)-(-)-Menthyl acetate
F2	F8	F14	F20	(1*S*)-(+)-Menthyl acetate
G2	G8	G14	G20	N-Benzyl-L-proline ethyl ester
H2	H8	H14	H20	N-Benzyl-D-proline ethyl ester
I2	I8	I14	I20	Methyl (*R*)-(-)-mandelate
J2	J8	J14	J20	Methyl (*S*)-(+)-mandelate
K2	K8	K14	K20	Ethyl (R)-(+)-4-chloro-3-hydroxybutyrate
L2	L8	L14	L20	Ethyl (S)-(−)-4-chloro-3-hydroxybutyrate
M2	M8	M14	M20	(+)-Ethyl D-Lactate
N2	N8	N14	N20	(−)-Ethyl L-lactate
O2	O8	O14	O20	(+)-Methyl (*S*)-3-hydroxybutyrate
P2	P8	P14	P20	Benzoic acid, 4-formyl-, phenylmethyl ester
A3	A9	A15	A21	(-)-Methyl (R)-3-hydroxybutyrate
B3	B9	B15	B21	(1R)-(+)-Neomenthyl acetate
C3	C9	C15	C21	(1S)-(+)-Neomenthyl acetate
D3	D9	D15	D21	Propylparaben (Propyl 4-hydroxybenzoate)
E3	E9	E15	E21	Butylparaben

F3	F9	F15	F21	Methyl 3-hydroxybenzoate
G3	G9	G15	G21	Methyl 2-hydroxybenzoate (methyl salicylate)
H3	H9	H15	H21	Phthalic acid diethyl ester (Diethyl phthalate)
I3	I9	I15	I21	(−)-Methyl (*R*)-3-hydroxyvalerate
J3	J9	J15	J21	(+)-Methyl (*S*)-3-hydroxyvalerate
K3	K9	K15	K21	Benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate
L3	L9	L15	L21	Benzyl 4-hydroxybenzoate (Benzyl paraben)
M3	M9	M15	M21	Methyl benzoate
N3	N9	N15	N21	Methyl butyrate
O3	O9	O15	O21	Methyl decanoate
P3	P9	P15	P21	Methyl oleate
A4	A10	A16	A22	Methyl hexanoate (methyl caproate)
B4	B10	B16	B22	Methyl octanoate (methyl caprylate)
C4	C10	C16	C22	Methyl dodecanoate
D4	D10	D16	D22	Methyl myristate
E4	E10	E16	E22	Propyl propionate
F4	F10	F16	F22	Propyl butyrate
G4	G10	G16	G22	Propyl hexanoate (propyl caproate)
H4	H10	H16	H22	Phenethyl cinnamate
I4	I10	I16	I22	Isobutyl cinnamate
J4	J10	J16	J22	Methyl 2,5-dihydroxycinnamate
K4	K10	K16	K22	Methyl cinnamate
L4	L10	L16	L22	Methyl ferulate
M4	M10	M16	M22	Vinyl acetate
N4	N10	N16	N22	Vinyl propionate
O4	O10	O16	O22	Vinyl butyrate
P4	P10	P16	P22	Vinyl laurate
A5	A11	A17	A23	Vinyl myristate
B5	B11	B17	B23	Vinyl palmitate
C5	C11	C17	C23	Vinyl oleate
D5	D11	D17	D23	Vinyl benzoate
E5	E11	E17	E23	Vinyl crotonate
F5	F11	F17	F23	Vinyl acrylate
G5	G11	G17	G23	Geranyl acetate
H5	H11	H17	H23	3-Methyl-3-buten-1-yl acetate
I5	I11	I17	I23	Ethyl 2-ethylacetoacetate
J5	J11	J17	J23	Ethyl 2-methylacetoacetate
K5	K11	K17	K23	Ethyl 3-oxohexanoate
L5	L11	L17	L23	Ethyl acetoacetate
M5	M11	M17	M23	Ethyl propionylacetate
N5	N11	N17	N23	ϒ-Valerolactone
O5	O11	O17	O23	Methyl glycolate
P5	P11	P17	P23	D-α-Hydroxy-β,β-dimethyl-γ-butyrolactone

A6	A12	A18	A24	L-α-Hydroxy-β,β-dimethyl-γ-butyrolactone
B6	B12	B18	B24	(1R)-(-)-dimenthyl succinate
C6	C12	C18	C24	Ethyl 2-chlorobenzoate
D6	D12	D18	D24	Cyclohexyl butyrate
E6	E12	E18	E24	n-Pentyl benzoate
F6	F12	F18	F24	2,4-Dichlorophenyl 2,4-dichlorobenzoate
G6	G12	G18	G24	2,4-Dichlorobenzyl 2,4-dichlorobenzoate
H6	H12	H18	H24	Diethyl-2,6-dimethyl 4-phenyl-1,4-dihydro pyridine-3,5-dicarboxylate
I6	I12	I18	I24	(+)-Methyl D-Lactate
J6	J12	J18	J24	(-)-Methyl L-Lactate
K6	K12	K18	K24	Propyl acetate
L6	L12	L18	L24	Butyl acetate
M6	M12	M18	M24	Phenyl acetate
N6	N12	N18	N24	Phenyl propionate
O6	O12	O18	O24	Phenyl hexanoate
P6	P12	P18	P24	Trilaurate

Table S2 Purification summary.

A) Purification summary for $EH1_A$

	Total protein (g)[b]	Total activity (units)[c]	Specific activity (U/g)[c]	Yield (%)[d]	Purity (%)
Crude lysate[a]	0.311	1180	3794		-
Ni-NTA His-Bind resin	0.0155	870	56120	73.7	>95[e]
Superdex 75	0.0135	829	61400	70.3	>99[f]

B) Purification summary for $EH1_B$

	Total protein (g)[b]	Total activity (units)[c]	Specific activity (U/g)[c]	Yield (%)[d]	Purity (%)
Crude lysate[a]	0.443	0.52	1.2		-
Ni-NTA His-Bind resin	0.0227	0.36	16.0	70.0	>95[e]
Superdex 75	0.0184	0.35	18.7	65.7	>99[f]

C) Purification summary for $EH1_{AB}$

	Total protein (g)[b]	Total activity (units)[c]	Specific activity (U/g)[c]	Yield (%)[d]	Purity (%)
Crude lysate[a]	0.461	37.8	82		-
Ni-NTA His-Bind resin	0.0178	27.0	1513	71.2	>95[e]
Superdex 75	0.0165	25.4	1538	67.1	>99[f]

[a] From 1 g of wet weight *E. coli* cell pellet (from 700 ml of bacterial culture).
[b] Protein concentration determined by Bradford assay using BSA as a standard protein.
[c] Activity determined using glyceryl tri-propionate as substrate.
[d] Calculated by the recovery of total units for the hydrolysis of glyceryl tri-propionate.
[e] Calculated by densitometry in the SDS-PAGE (Supporting Figure S3).
[d] Calculated by densitometry in the SDS-PAGE (Supporting Figure S3) and MALDI analysis (Supporting Figure S5).

Table S3. Kinetic parameters of $EH1_A$, $EH1_B$ and $EH1_{AB}$. The kinetic parameters for phenyl propionate (at pH 8.0) and *p*NPP (at pH 7.0) were calculated as described in Experimental section, in triplicate at 30°C. Kinetic parameters were calculated by simple Michaelis-Menten kinetics (see Supporting Figure S9). Mean values are given with the standard deviation.

	Kinetic parameters		
	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($s^{-1}M^{-1}$)
$EH1_A$ (for *p*NPP)	146 ± 27	3853 ± 23	4.4×10^5
$EH1_B$ (for *p*NPP)	267± 20	263 ± 16 $\times 10^{-3}$	16.4
$EH1_{AB}$ (for *p*NPP)	485 ± 21	211 ± 8	7250
$EH1_A$ (for phenyl propionate)	964 ± 24	2797± 98	4.8×10^4
$EH1_B$ (for phenyl propionate)	752± 20	1.49 ± 0.93	33.0
$EH1_{AB}$ (for phenyl propionate)	942 ± 17	267 ± 15	4720

Table S4 The 89 esters listed in Figure 5.

Nr	Substrate
1	Methyl cinnamate
2	Ethyl benzoate
3	Methyl benzoate
4	Isobutyl cinnamate
5	Phenylethyl cinnamate
6	n-Pentyl benzoate
7	Dodecanoyl acetate
8	Methyl decanoate
9	Ethyl decanoate
10	Benzylparaben
11	Butylparaben
12	Propylparaben
13	Ethyl butyrate
14	Ethyl propionate
15	Vinyl acetate
16	Vinyl acrylate
17	Vinyl crotonate
18	Butyl acetate
19	Propyl acetate
20	3-Methyl-3-buten-1-yl acetate
21	Propyl butyrate
22	Propyl propionate
23	Methyl butyrate
24	Ethyl 2-chlorobenzoate
25	Phthalic acid diethyl ester
26	2,4-Dichlorobenzyl 2,4-dichlorobenzoate
27	(-)-Methyl (R)-3-hydroxybutyrate
28	(+)-Methyl (S)-3-hydroxybutyrate
29	(+)-Methyl (S)-3-hydroxyvalerate
30	(−)-Methyl (R)-3-hydroxyvalerate
31	Ɣ-Valerolactone
32	Glyceryl tributyrate
33	Glyceryl triacetate
34	Glyceryl tripropionate
35	1-Napthyl acetate
36	1-Naphthyl butyrate
37	Phenyl acetate
38	Phenyl propionate
39	(-)-Methyl L-Lactate
40	(+)-Methyl D-Lactate
41	Vinyl butyrate
42	Vinyl propionate
43	Glucose pentaacetate
44	Methyl glycolate
45	Methyl (S)-(+)-mandelate

46	Methyl (R)-(-)-mandelate
47	Ethyl 2-methylacetoacetate
48	Ethyl 2-ethylacetoacetate
49	(−)-Ethyl L-lactate
50	(+)-Ethyl D-Lactate
51	Ethyl propionylacetate
52	Ethyl acetoacetate
53	Ethyl 3-oxohexanoate
54	Ethyl (S)-(+)-4-chloro-3-hydroxybutyrate
55	Ethyl (R)-(+)-4-chloro-3-hydroxybutyrate
56	Octyl acetate
57	Hexyl acetate
58	Geranyl acetate
59	Cyclohexyl butyrate
60	Propyl hexanoate
61	Methyl octanoate
62	Methyl hexanoate
63	Ethyl octanoate
64	Ethyl hexanoate
65	Vinyl benzoate
66	Benzyl (R)-(+)-2-hydroxy-3-phenylpropionate
67	Benzoic acid, 4-formyl-, phenylmethyl ester
68	(1S)-(+)-Menthyl acetate
69	(1R)-(-)-Menthyl acetate
70	(1S)-(+)-Neomenthyl acetate
71	(1R)-(+)-Neomenthyl acetate
72	Ethyl acetate
73	Glyceryl trioctanoate
74	Ethyl dodecanoate
75	Pentadecyl acetate
76	Glyceryl trilaurate
77	Triolein
78	Vinyl laurate
79	L-Pantolactone
80	D-Pantolactone
81	N-Benzyl-L-proline ethyl ester
82	Methyl ferulate
83	Methyl 3-hydroxybenzoate
84	Methyl 2-hydroxybenzoate
85	Methyl 2,5-dihydroxycinnamate
86	Diethyl-2,6-dimethyl 4-phenyl-1,4-dihydro pyridine-3,5-dicarboxylate
87	Methyl myristate
88	Methyl dodecanoate
89	2,4-Dichlorophenyl 2,4-dichlorobenzoate

Supporting Results Section

Control experiments to demonstrate the purity and identity of proteins. A number of control experiments were performed to support our assertions: that is plausible to rational engineer a second active site in an ester hydrolase, so that to create enzymes with multiple active sites.

Each His-tagged protein, native and engineered variants, were purified at 4 °C after binding to a Ni-NTA His-Bind resin, followed by size-exclusion chromatography (see EXPERIMENTAL AND COMPUTATIONAL METHODS section), and the purified protein solutions were further analyzed to ensure identity and purity, as follows. The purified proteins were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on 12 % gels (in a Mini PROTEAN electrophoresis system (Bio-Rad)) in which the proteins were stained with Coomassie brilliant blue (Protoblue Safe, National Diagnostics, GA, USA). As shown in Supporting Figure S3 the purity of the proteins was high (>95% after Ni-NTA His-Bind resin step and >99% after gel filtration step, as determined by densitometry). The degree of purity was further confirmed by MALDI-TOF/TOF (see below). Through size-exclusion chromatography we further monitored whether the catalytic activity (measured by monitoring the activity towards *p*NPP at 410 nm) elutes in the same pattern as the enzyme (by measuring protein elution at 280 nm). As shown in Supporting Figure S4, we observed that the peak of catalytic activity is at the same position as the enzyme peak, which demonstrates that no contamination in the protein solution exists. As many esterases and proteases may have a similar molecular weight as the EH1 enzyme variants (~35 kDa) and that such possible contaminants may co-elute (both in the SDS-PAGE and the gel filtration), we further used proteomics analytical platforms to analyze protein purity and homogeneity. Briefly, mass spectrometry of whole (intact) proteins (Supporting Figure S5) was used to confirm the identity and quality of the produced and purified recombinant protein variants (purified after binding to a Ni-NTA His-Bind resin followed by size-exclusion chromatography). We first determine the molecular mass of intact proteins. A monoisotopic mass of 35811Da (for $EH1_A$ protein), 35804 Da (for $EH1_B$ protein) and 35820 Da (for $EH1_{AB}$ protein) was calculated on the basis of sequence information. The molecular mass experimentally obtained (~35 kDa) in all cases agree with the theoretical mass (Supporting Figure S5), supporting the recombinant proteins do show the expected molecular masses. MALDI intact mass analysis can also interpret the full mass spectrum without any deconvolution step and the analysis therefore gives a good indication of purity of protein samples. As shown in Supporting Figure S5, the chromatographic protein purity was higher than 99%. Finally, peptide mass fingerprinting by

MALDI-TOF/TOF was performed to detect the presence of any protein contaminant, for example, remaining contaminations from *E. coli* proteins as this bacterium was used as a host, as well as other proteins that may be introduced during the purification steps. By peptide fingerprint analysis we did not identify any protein from *E. coli* nor other protein when searching in a protein database with the sequences encoding all proteins from this bacterium in NCBI database, and the entire NCBI. Therefore, with the sensitivity allowing SDS-PAGE and MALDI, we can conclude that the purity of His-tagged protein variants was higher than 99% and that the presence of any protein contaminant can be excluded.

MALDI-TOF/TOF is a technique that can be also utilized to analyze the peptides for primary sequence confirmation and presence of mutations. For this reason to confirm that the proteins contained the introduced mutations, peptide fingerprint by MALDI-TOF/TOF was performed using the purified proteins. As shown in Supporting Figure S6 (for $EH1_A$), Figure S7 (for $EH1_B$) and Figure S8 (for $EH1_{AB}$), by MALDI-TOF/TOF we have detected peptides covering the entire protein sequences, including the specific peptides corresponding to each of the protein variants. This analysis supports that the mutations were correctly introduced and that the identity of proteins was unambiguously confirmed. Note that mutations were also confirmed in the cloned gene by DNA sequencing.

Hydrolytic capacity of $EH1_A$, $EH1_B$, $EH1_{AB}$ and their mutants. Two different tests were used to evaluate the hydrolytic activity of all hydrolase variants herein examined. The first one consisted in the hydrolysis of *p*NPP (up to 8 mM), whose hydrolysis can be monitored by measuring at 410 nm (herein used at pH 7.0) or 346 nm (regardless pH) the production of *p*-nitrophenoxide. This substrate is commonly used to evaluate the performance of ester hydrolases and mutants, because its ease of measurement and low detection limit. The second one consisted in the hydrolysis of esters others than *p*NPP, i.e. phenyl propionate, whose hydrolysis can be monitored via a pH indicator assay, at pH 8.0 (or pH 7.0). Acid is produced after ester bond cleavage by the action of the hydrolytic enzyme, which induces a color change of the pH indicator that can be measured spectrophotometrically at 550 nm. Following recommendation elsewhere,[30,33] the concentrations of the pH indicator (0.45 mM) and each of the esters (up to 6 mM) were chosen to maximize the accuracy and sensitivity. In all cases, a positive reaction was indicated by the restrictive criterion of a change greater than 6-fold above the background signal. Also, we would like to highlight that the amount of protein in the assays was significantly high, particularly for mutants, whose activity was evaluated using 100 μg pure protein and time frames up to 24 h to ensure sensitivity and detection of ester

hydrolysis.[33] Therefore, the catalytic rates of $EH1_A$, $EH1_B$, $EH1_{AB}$ and the knock-out mutants were all evaluated under conditions that ensure sensitivity and detection limit. In this study we consider no (near zero) capacity to hydrolyze an ester when the absorbance after 24 h incubation at pH 7.0 (for *p*NPP hydrolysis) or pH 8.0 (for esters other than *p*NPP) at 30ºC was 2-fold below the background signal. This was the case of all mutants for all esters tested.

3.4. PluriZyme: catalytic advantages of having two active sites in an enzyme

Gerard Santiago[2,8], Sandra Alonso[1,8], Isabel Cea-Rama[3,8], Cristina Coscolín[1], Laura Fernández-López[1], Mónica Martínez-Martínez[1] , David Almendral[1], Helena Marrero[1], Ruth Matesanz[4], Rafael Bargiela[5,6] ,Peter N. Golyshin[5,6], Julia Sanz-Aparicio[3,9,*], Manuel Ferrer[1,9, *] and Victor Guallar[2,7,9,*]

1 Institute of Catalysis, Consejo Superior de Investigaciones CientÌficas (CSIC), 28049 Madrid, Spain.

2 Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain.

3 Department of Crystallography & Structural Biology, Institute of Physical Chemistry "Rocasolano", CSIC, 28006 Madrid, Spain.

4 Spectroscopy Laboratory, Centro de Investigaciones Biológicas, CSIC, 28040 Madrid, Spain.

5 School of Natural Sciences, Bangor University, LL57 2UW Bangor, UK.

6 Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, UK.

7 Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

8 These authors contributed equally: Sandra Alonso, Gerard Santiago, Isabel Cea-Rama.

9 These authors equally coordinated the work: Julia Sanz-Aparicio, Manuel Ferrer, Victor Guallar. *e-mail: *(J.S.) Email: xjulia@iqfr.csic.es (J.S-A.); mferrer@icp.csic.es (M.F.); victor.guallar@bsc.es (V.G.).

PluriZyme: catalytic advantages of having two active sites in an enzyme

Sandra Alonso[1,8], Gerard Santiago[2,8], Isabel Cea-Rama[3,8], Cristina Coscolín[1], Laura Fernández-López[1], Mónica Martínez-Martínez[1], David Almendral[1], Helena Marrero[1], Ruth Matesanz[4], Rafael Bargiela[5,6], Peter N. Golyshin[5,6], Julia Sanz-Aparicio[3,9,*], Manuel Ferrer[1,9,*] and Víctor Guallar[2,7,9,*]

[1]Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), 28049 Madrid, Spain. [2]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain. [3]Department of Crystallography & Structural Biology, Institute of Physical Chemistry "Rocasolano", CSIC, 28006 Madrid, Spain. [4]Spectroscopy Laboratory, Centro de Investigaciones Biológicas, CSIC, 28040 Madrid, Spain. [5]School of Natural Sciences, Bangor University, LL57 2UW Bangor, UK. [6]Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, UK. [7]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain. [8]These authors contributed equally: Sandra Alonso, Gerard Santiago, Isabel Cea-Rama. [9]These authors equally coordinated the work: Julia Sanz-Aparicio, Manuel Ferrer, Víctor Guallar. *e-mail: *(J.S.) Email: xjulia@iqfr.csic.es (J.S-A.); mferrer@icp.csic.es (M.F.); victor.guallar@bsc.es (V.G.).

KEYWORDS: active site, esterase, enzyme catalysis, modeling, PELE, rational design

There are numerous enzymes with an artificial active site catalyzing desired reactions (k_{cat} as high as 5 s^{-1}). A recent advancement involved the addition of a second active site into an ester hydrolase, producing a new site that, however, could not bring in catalytic advantages. Here, we significantly challenge this approach, introducing to the artificial site a k_{cat} as high as 14 s^{-1}. Its incorporation to the "mother" enzyme intensifies ca. 5000-fold the k_{cat}/K_m, introduces stereo-specificity (%ee >99.9%), and unexpectedly, broadens the temperature for optimal activity. Structural and further engineering efforts provide definite and reproducible evidences of a *plurizyme*: a two active site enzyme where substrate binding and conversion concomitantly occur, each trigging distinct conformational changes, that are additive. Adding extra reactive sites with identical chemistry catalyzed can thus help intensifying the bio-catalytic properties of "mother" ester-hydrolases. We hypothesized that it can also help generating multi-catalytic systems with different chemistry.

The enzyme engineering field has developed considerably, opening great perspectives for future studies aiming at greener solutions in synthetic biochemistry. Directed evolution and computation-driven rational mutagenesis are fostering such developments[1-4]. Within the most common enzyme engineering actions, the tuning of specificity and improving the activity and/or stability are the most attractive. In addition, significant efforts in *de novo* active sites design are being undertaken; the introduction of catalytic sites opened new opportunities for non-catalytic protein scaffolds[5-9]. There are numerous examples of enzymes with artificial active sites computational designed, although directed evolution and protein engineering was required to boost the activity of the original design[5-9]. As such, certain artificial enzymes approach the diffusion limit while still catalyzing the desired reaction using an artificial active site[10-12], with turnover rates as high as 1-5 s^{-1}.

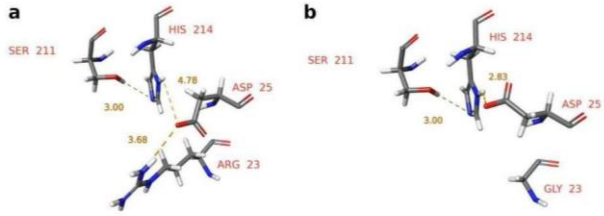
In a recent study we significantly challenged enzyme engineering by introducing, in a natural serine Ester-Hydrolase ($EH1_A$), a second active site ($EH1_B$). Using the Protein Energy Landscape Exploration (PELE) software, we identified a potential binding pocket and successfully turned it into a catalytic site by introducing a few mutations (for additional details see Note 1 in Supporting Introduction)[13]. The first EH with two separate (and active) catalytic Ser-Asp-His triads, along with their respective essential oxyanion holes[14], was thus designed ($EH1_{AB}$). However, introducing the new artificial active site in our target $EH1_A$ has not produced any benefit[13], neither at the level of catalytic efficiency nor at the level of substrate scope, selectivity and specificity. Thus, the artificial site has a max. k_{cat} of up to 0.03 s^{-1} for a restricted set of substrates[13], value which is much below that of previous artificial designs (1-5 s^{-1})[5-9], and that of the "mother" $EH1_A$ native site (ca. 47 s^{-1}), which also showed a broader substrate spectrum[13].

Still, we focus here on the *plurizyme* concept: developing enzymes with additional (artificial) actives sites where improved catalytic properties are introduced, a field where there is a lack of knowledge. To what extend can additional sites intensify the catalytic performance? Which catalytic

and structural properties can the artificial site introduce, compared to those of the native one, when both sites coexist and are identical in respect to the chemistry catalyzed? Are these properties additive, complementary or different? Can the extra reactive site be a manifold to introduce additional catalytic conversions and thus to design novel multi-catalytic systems? These questions are investigated here in order to evaluate both the academic and biotech potential of the *plurizyme* concept. First, we demonstrate that the catalytic potential of a "mother" hydrolase can be significantly boosted by adding an artificial site with an appropriated catalytic configuration. Second, we provide a definite structural evidence of the existence of a *plurizyme*, introducing a crystal structure of a serine ester-hydrolase with a native and an artificial active site, as well as a comprehensive identification of the conformation changes occurring upon substrate binding. Third, we evaluate the reproducibility of introducing an extra catalytic site in other randomly selected serine ester-hydrolase[15], which also associated with a catalytic enhancement.

Results and discussion

Using the PELE software with $EH1_B$ as our initial model, we mapped glyceryl tri-propionate dynamics, one of the preferred substrates of this site[13]. The rationale was to investigate the cumulative effects of mutations in non-active site residues close to catalytic positions; PELE suggested the R23G mutation as best potential target for active site remodeling. This Arg amino acid seems to work to destabilize the artificial catalytic triad by partially sequestering Asp25, and substitution by Gly indicated a better stability of the His-Asp interaction (Fig. 1).



	EH1AB (A)	EH1AB1 (B)
SER211-HIS214 Distance (Å)	5,2 ± 1,8	3,4 ± 1,5
HIS214-ASP25 Distance (Å)	8.4 ± 1.6	5.5 ± 1.4

Fig. 1 Two representative snapshots of $EH1_B$ (a) and $EH1_{B1}$ (b) catalytic triads from the molecular dynamic simulations, where we clearly observe the movement in Asp25 as a result of the R23G mutation. The lower panel indicates the average distances between key catalytic amino acids obtained for both species along the molecular dynamics.

In addition, introducing the smaller side chain significantly increased the effective volume of the active site due to a higher cavity size, which increased from 58 to 66 $Å^3$, a parameter which largely

correlates with substrate broadening[5]. Such an increase allowed the new site to overcome the previously established threshold (62.5 Å^3) above which EH hydrolyze more than 24 substrates (from a set of 96 tested). The R23G mutation was introduced into the $EH1_A$, $EH1_B$ and $EH1_{AB}$ variants to evaluate its effect on their catalytic performance. The variants containing the R23G mutation, which was found not to affect *per se* the activity of the original enzyme (see Note 1 in Supporting Results), will be referred to as $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$.

Catalytic performance of the remodeled artificial site

As shown in Table 1 and Fig. 2a, the R23G mutation produced an artificial esterase (EH_{B1}) with a k_{cat} as high as 820.4 ± 2.5 min^{-1}, and a substrate range of 58 chemically and structurally diverse esters (for additional details see Note 1 in Supporting Results). Those values are much above the interquartile range of most naturally occurring hydrolytic enzymes (Table 1), and approximate those of the native site ($EH1_{A1}$), one of the most active and promiscuous in terms of the substrate scope[13,15]. Also, they are above those of the best artificial enzymes obtained to date, which achieved k_{cat} of up to 1 to 5 s^{-1} in best cases[5-9,16-24].

Table 1 Substrates accepted by EH1 variants and other reported EHs, and their conversion rates.

	Number of esters converted[a]	V_{max} (U g^{-1})[b]	k_{cat} (min^{-1})[c]
EHs in the literature	21 (9 – 29)	270680 (0 – 1049)	-
$EH1_A$	79	22620 ± 160	2797 ± 18 – 0.17± 0.02
$EH1_B$	24	10.97± 0.14	1.49 ± 0.15 – 0.004 ± 0.003
$EH1_{AB}$	79	1791 ± 21	267.0 ± 1.6 – 0.02 ± 0.01
$EH1_{A1}$	79	22950 ± 173	2715 ± 2.1 – 0.07 ± 0.01
$EH1_{B1}$	58	5051 ± 51	820.4 ± 2.5 – 0.084 ± 0.001
$EH1_{AB1}$	79	29790 ± 146	3468 ± 12.6 – 0.075 ± 0.001

[a]The esters, out of 98 tested, converted by EHs herein generated and in literature (145 in total) are available in Supporting Table S1 and Santiago et al. (2018)[13], respectively. [b]V_{max} (with standard deviation of triplicates) for the hydrolysis of glyceryl tri-butyrate at pH 8.0, 30ºC and 50 mM ester. For EHs in literature (225 in total) the data, available in Santiago et al. (2018)[13] and Martínez-Martínez et al. (2018)[15], correspond to specific activities under different conditions; the values given represent the average value, with 50% of the data from 0-1049 U g^{-1} shown in brackets. [c]k_{cat} range for the more and less hydrolyzed esters converted by EHs generated in this study at pH 8.0, 30ºC and 50 mM substrates. Values (average with standard deviation of triplicates) for each ester are available in Supporting Table S1.

Although, all esters hydrolyzed by $EH1_{B1}$ were also hydrolyzed by $EH1_{A1}$ (Fig. 2a; Supporting Table S1), we observed that the remodeled site shows, compared to the native site, not only an altered substrate scope, but also an altered substrate-preference and stereo-selectivity, possibly due to differences in substrate diffusion to access both active sites and product diffusion. This is exemplified by the fact that within all 55 common substrates, 50 were better transformed by $EH1_{A1}$, while 5 were

better converted by $EH1_{B1}$: methyl 3-hydroxybenzoate (70-fold), vinyl propionate (29-fold), methyl hexanoate (9-fold), methyl (2*R*)-2-phenylpropanoate (8-fold), and vinyl crotonate (3.6-fold). Also, by the fact that $EH1_{B1}$ converts methyl 3-hydroxybenzoate (k_{cat}: 58.3 ± 0.3 min^{-1}) but not methyl 2-hydroxybenzoate, ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate (k_{cat}: 80.7 ± 4.1 min^{-1}) but not ethyl (*R*)-(+)-4-chloro-3-hydroxybutyrate, and methyl (2*R*)-2-phenylpropanoate (k_{cat}: 526.0 ± 11.8 min^{-1}) but not methyl (2*S*)-2-phenylpropanoate; opposite, all these esters were converted by $EH1A_1$ (Fig. 2a; Supporting Table S1).

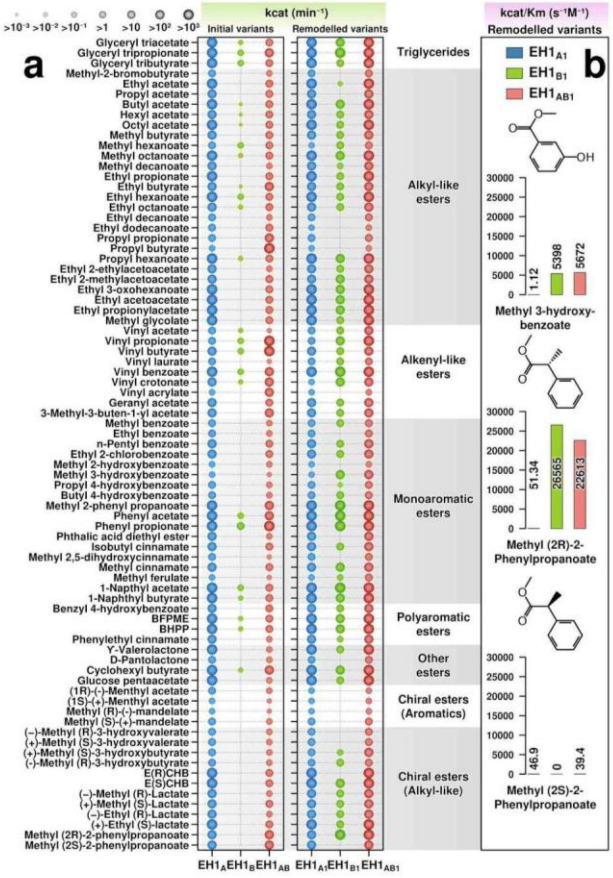


Fig. 2 The artificial site intensifies the bio-catalytic capability of the native enzyme. a, Substrate ranges and turnover number (k_{cat}) of the initial EH1 variants and the variants containing the R23G mutation. The figure was created with the R language console, using information about the k_{cat} (min^{-1}) of the analyzed enzymes against the substrates hydrolyzed as a starting point. Abbreviations: E(R)CHB and E(S)CHB, ethyl (*R* or *S*)-(−/+)-4-chloro-3-hydroxybutyrate. For other abbreviations and raw data see Supporting Table S1. **b**, k_{cat}/K_m ($s^{-1}M^{-1}$) of $EH1_{A1}$,

$EH1_{B1}$ and $EH1_{AB1}$ against two substrates preferentially hydrolyzed by $EH1_{B1}$ and which are best representative of the specificity and stereo-chemistry. k_{cat} and K_m values are given in Supporting Table S2. In all cases, the activity protocol established and used to identify the esters hydrolyzed by each variant was based on a continuous pH indicator assay at pH 8.0 and 30 °C, performed in triplicate (with the average value and standard deviation shown) and corrected for background signal (see Methods section).

Therefore, remodeling of the extra active site not only broadens its substrate spectra and increases prominently its hydrolytic rate, but also introduces stereo-preference, a feature that do not characterize the native site. This may be a direct consequence of disabling the Asp25-Arg23 interaction after R23G mutation such that Asp25, which is part of the artificial catalytic triad, can be better oriented towards the catalytic histidine (Fig. 1). Arg23 may also restrict substrate accessibility in the artificial site, and its mutation to Gly has a large effect in reducing these constraints such that the remodeled active site can accommodate larger substrates.

$EH1_{AB1}$ advantages compared to the "mother" $EH1_{A1}$ enzyme

$EH1_{AB1}$ has the ability to hydrolyze all 79 substrates that $EH1_{A1}$ and $EH1_{B1}$, combined, were able to convert (Fig. 2a). Interestingly, k_{cat} values are close to the sum of the individual values from variants containing each of the separate sites. As a consequence, we found that $EH1_{AB1}$, compared to the "mother" enzyme, increase its k_{cat} for best-hydrolyzed substrate by 1.3-fold (Table 1), and by ~2.7-fold on average for all common esters converted, with best improvement values of 74-fold (Fig. 2a; Supporting Table S1; see Note 1 Supporting Results).

Calculation of the k_{cat}/K_m for methyl 3-hydroxybenzoate and the chiral ester methyl (2*R*)-2-phenylpropanoate, which were best converted by the engineered site compared to the native one (see above), and representing substrate-specificity and stereo-selectivity features, were performed to quantify to what extend the addition of a second active site intensifies the catalytic efficiency of the "mother" hydrolase. Results revealed that the significantly higher turnover rate (k_{cat}) and affinity (K_m) for these substrates in the artificial site (Supporting Table S2) intensifies the k_{cat}/K_m of the "mother" enzyme by up to ca. ~5000-fold (Fig. 2b). Kinetic resolution of enantiomers in a racemic mixture of methyl (2*R*)-2-phenylpropanoate and methyl (2*S*)-2-phenylpropanoate was used to generate conversion and %ee values for substrates and products, and to prove to what extend the addition of a second active site intensifies enantio-selectivity. Note that, when tested separately, while the (*R*) enantiomer is converted by both the artificial and native sites, although preferentially by the first, the (*S*) is only concerted by the native one (Fig. 2b). We found that the addition of the artificial site (E-value >1000) increases the E-value of the "mother" enzyme (E-value ~1.2) by ca. 1200-fold (E-value for $EH1_{AB1}$: ~1380).

Taken together, we have provided unambiguous evidences demonstrating that adding an extra active site can thus significantly increase the catalytic efficiency of a native ester hydrolase due to the

presence of two reactive groups. At the same time, through differences in affinities and conversion rates, it can introduce new properties to the original enzyme, namely, improved stereo-selectivity, as demonstrated by using a set of randomly selected chiral esters.

The artificial site broadens the catalytic temperature window

The activity level of each variant at different temperatures was investigated (Fig. 3a) using glyceryl tri-propionate, one of the preferred substrates. $EH1_{A1}$ was strongly active (>80% rel. act.) at temperatures in the range 30-50ºC, with the activity progressively reduced outside this range. In detail, the optimum temperature (T_{opt}) was 35-45ºC, and the enzyme retained <50% activity at 8ºC and 55ºC. $EH1_{B1}$, exhibited the reverse of this trend —it was strongly inactivated at temperatures above 35ºC, and more active at lower temperatures. The T_{opt} was ~30ºC, and the enzyme retained ~85% activity at 8ºC and <25% at 40ºC. The enzyme with two reactive sites exhibited a hybrid pattern, being most active at 8-50ºC. Thus, a high activity level was observed at temperatures as low as 8ºC (~85%) and as high as 50ºC (~80%). The two-active site enzyme thus retained the psychrophilic-like phenotype introduced by the extra active site and the mesophilic-like phenotype of the native one. Denaturing temperatures (T_d) of 42.1±0.2ºC and 33.8±0.2ºC were calculated, by circular dichroism (CD) spectroscopy, for $EH1_{A1}$ and $EH1_{B1}$, respectively (Fig. 3b-d). Interestingly, the T_d of $EH1_{AB1}$ was 47.5±0.2ºC. These values are in agreement with the corresponding optimal temperatures for the activity of each protein variant (Fig. 3a).

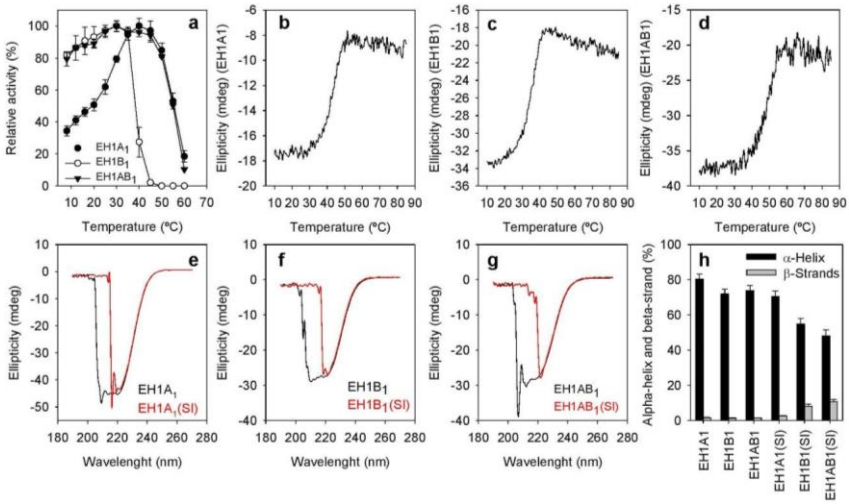


Fig. 3 Thermal and conformation changes associated to each of the protein variants. a, Temperature profiles of the purified $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ variants. The data represent the relative percentages (%) of

specific activity at pH 8.0, expressed as U mg^{-1} compared with the maximum activity when glyceryl tri-propionate was used as the substrate. All data, calculated from three independent assays ± SD, are not fitted to any model. **b-d**, Thermal denaturation curves of $EH1_{A1}$ (b), $EH1_{B1}$ (c) and $EH1_{AB1}$ (d) variants, measured by ellipticity changes at 220 nm and obtained at different temperatures and at pH 7.0. The results from the three variants (statistically similar) were averaged. **e-g**, CD analysis of $EH1_{A1}$ (e), $EH1_{B1}$ (f) and $EH1_{AB1}$ (g) recorded at 25°C in the absence (black) or upon binding (red) of the suicide inhibitor (SI) butyl 4-nitrophenyl hexylphosphonate. Shown is the ellipticity in mdeg from 190 to 270 nm. **h**, Effect of SI binding on the conformation of $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$. Shown is the α-helix and β-strand content (%) of the query protein's secondary structure, calculated from CD data in the range of 190 to 240 nm, as reported previously (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/)[27]. Reported are the average values of three measurements ± SD.

The above results demonstrate that the physicochemical properties of the serine ester-hydrolase EH1 can be modulated by the presence of multiple reactive sites, each conferring different catalytic performances to the protein scaffold, most likely due to the different active site architectures. It is plausible that the remodeled artificial site is less stable, probably because of the different configurations and positioning (larger exposure to the solvent) of the catalytic residues, which are differentially affected by thermal conditions. The higher activity at low temperatures may be associated with an easier diffusion of the substrates in the remodeled site, compared to the native site, due to its larger solvent exposure. Having said that, while the increase of the temperature for optimal activity could be directly ascribed to the presence of two active sites, the observed changes in thermo-stability cannot be explained by the introduction of a second active site per se, as this is a global phenomenon of the entire protein. Whatever the case, those changes may depend from protein to protein to be used as "mother" target.

Crystal structure of the two active sites enzyme

To unambiguously prove that both sites are capable of substrate binding, we performed structural analysis. We have obtained crystals from $EH1_{AB1}$ diffracting to 2.1 Å resolution. These crystals were soaked with suicide inhibitor butyl 4-nitrophenyl hexylphosphonate to obtain the corresponding derivative complex, with two molecules of the inhibitor bound at the catalytic Ser161 (original nucleophile) and Ser211 (artificial nucleophile) sites (Fig.5a; see Note 2 in Supporting Results). The 4-nitrophenyl phosphonate inhibitor is susceptible to nucleophilic attack by the catalytic Ser, leading to covalent modification and complete inactivation of the enzyme (see Note 3 in Supporting Results)[25,26]. The solved three-dimensional structures show high flexibility of a region (which it resembles but does not equal to a typical *lid* of lipases) containing the two N-terminal helical regions, Pro4-Gly19 and Ala30-Gly43, which give access to the active-site pocket. Conformational changes have been observed at the secondary Ser211 catalytic site, the artificial site, upon inhibitor binding, as shown in Fig. 5b, which introduce distortions in the packing arrangement within the soaked crystals explaining the decreased resolution observed. Nevertheless, the atomic interaction of the inhibitor

bound at the two sites can be depicted and is displayed in Fig. 6. Analysis of ligand binding through CD measurements further confirm this, demonstrating that substrate binding triggers distinct conformational changes in each of the sites (Fig. 3e-g, red lines), which were not observed in the absence of substrate (Fig. 3e-g, black lines).

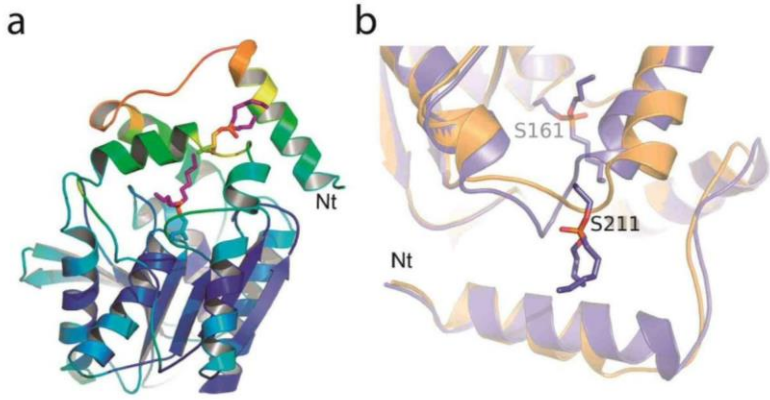


Fig. 5 Crystal structure of $EH1_{AB1}$. a, Cartoon of the soaked crystals coloured by B factors, from low (blue) to red (high), with the two bound molecules from the suicide inhibitor butyl 4-nitrophenyl hexylphosphonate, represented as magenta sticks. **b**, Detail of the comparison between the free (orange) and complexed (slate) $EH1_{AB1}$, showing the conformational changes observed in the environment of Ser211 upon inhibitor binding.

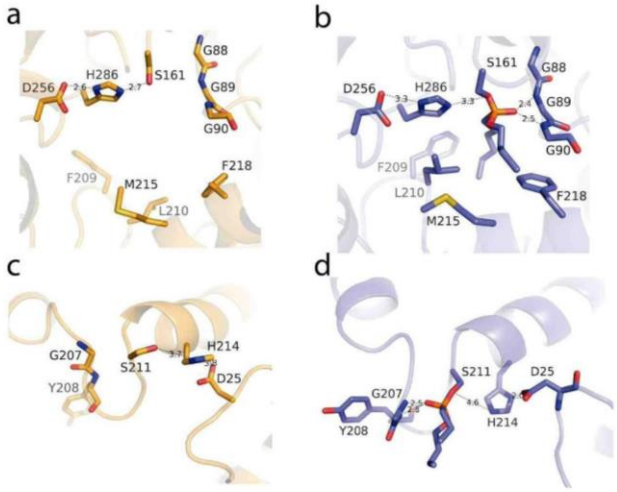


Fig. 6 Detail of the Ser161 and Ser 211 binding sites in the free (a, c) and complexed (b, d) crystals, showing relevant distances as dashed lines. For detailed superimposition see Note 2 in Supporting Results (Fig. S3).

Distinct conformational changes were further revealed by calculating the query proteins' secondary structures of α-helices and β-strands[27]. As shown in Fig. 3h, in the absence of substrate their content was similar in $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$. Upon substrate binding, the CD spectra of $EH1_{A1}$ slightly decreased in α-helices (1.13-fold) and slightly increased (1.52-fold) in the β-strand content. These differences were higher in $EH1_{B1}$, in which upon substrate binding, the α-helix content decreased by 1.3-fold and the β-strand content increased by 5.61-fold. This suggests that substrate binding triggers distinct conformational changes in each of the sites. A greater conformational change was observed in $EH1_{AB1}$ upon substrate binding, with a 1.54-fold decrease in α-helix content and a 7.12-fold increase in β-strand content. Moreover, changes were additive, as the β-strand content upon substrate binding in $EH1_{AB1}$ was approx. the sum of the values for each individual site.

The *plurizyme* approach is applicable to other ester-hydrolase

To test reproducibility of adding a second site, we chose a serine EH, referred to as $EH102_A$ (5JD3), isolated from the same habitat as $EH1_A$[15]. Its exposed native active site hydrolyzes only 16 substrates out 96 tested[11], with the V_{max} from 1038.3 ± 0.2 to 145.1 ± 22.1 U g^{-1} (Fig. 7a). PELE simulations revealed a second binding site (Fig. 7b), and following the nomenclature of EH1, two variants, $EH102_B$ and $EH102_{AB}$, were created (for additional details see Note 4 Supporting Results).

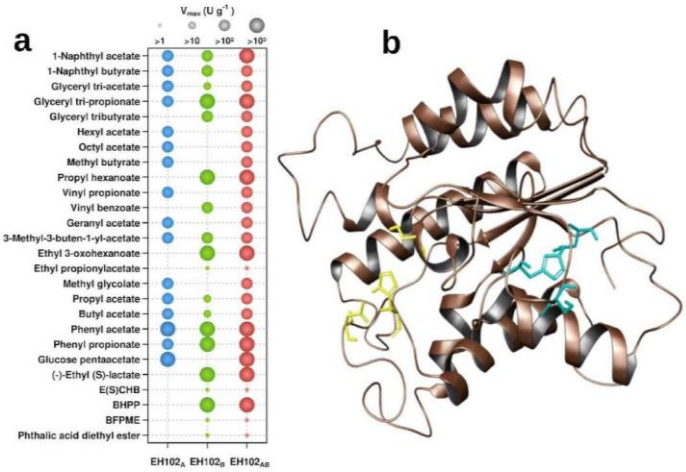


Fig. 7 The artificial site intensifies the bio-catalytic capability of the EH102 native enzyme. a, V_{max} of EH102 variants for all esters tested out of 96. The figure was created with the R language console, using abbreviations and raw data in Supporting Table S3. **b,** Relative positions of the active sites of EH102 with their catalytic triads. The original active site ($EH102_A$) is shown in yellow, with a catalytic triad formed by Ser15, Asp192 and His195 and Gly64 as oxyanion hole. The artificial active site ($EH102_B$) is shown in blue, with a second catalytic triad formed by Ser30, His34 and Asp57 and Asn43 as oxyanion hole. Note: (+)-ethyl (*R*)-

lactate and ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate are not shown as no appreciable activity was detected under our assay conditions for any of the tested hydrolase variants.

As shown in Fig. 7a, $EH102_B$ hydrolyzed 19 esters with V_{max} values ranging from 5726.5 ± 71.2 to 1.44 ± 0.70 U g^{-1}, values which are in the range of the native site. Interestingly, $EH102_B$ does show stringent stereo-chemistry for (+)-ethyl (*R*)-lactate (1227.7 ± 10.8 U g^{-1}) and ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate (6.64 ± 1.49 U g^{-1}), substrates for which $EH102_A$ shows activity below detection limit. $EH102_{AB}$, with two active sites, was able to convert 26 substrates that $EH102_A$ and $EH102_B$ converted, with V_{max} values (from 6282.3 ± 5.8 to 1.52 ± 0.05 U g^{-1}) which were close to the sum of the individual values from variants containing each of the separate sites. Compared to the original enzyme, $EH102_{AB}$ was on average for all esters converted ~2.3-fold more active, thus demonstrating again that additivity exists between the native and the artificial sites as observed in $EH1_{AB1}$. This variant also retained the capacity to efficiently hydrolyze esters initially not converted by the native enzyme. By kinetic resolution of enantiomers in a racemic mixture we found that the addition of the artificial site (E-value >1000 (+)-ethyl (*R*)-lactate and ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate) increases the E-value of the "mother" enzyme from no conversion (in $EH1102_A$) to >1000 (for $EH102_{AB}$) in both cases. As for the EH1 variants, we also observed that the introduction of the artificial site to the "mother" hydrolase expanded the temperature window for activity, with only slight differences in T_d (for additional details see Note 5 Supporting Results). In this case, the native site has a psychrophilic-like phenotype whereas the artificial site a mesophilic-like phenotype, which were both transferred to $EH102_{AB}$. Also, substrate binding was also shown to trigger distinct conformational changes in each of the sites (for additional details see Note 5 Supporting Results). Thus, through introducing a second active site with a different architecture, we expanded substrate range, slightly increased the temperature window, and converted an initially non-enantiospecific serine EH into a synthetic biocatalyst capable of converting a chiral molecule with stringent stereo-chemistry.

Conclusions

Improving the efficiency of enzymes is rapidly becoming a necessity. One could imagine attacking the problem by engineering more and more reactive sites, with the same chemistry catalyzed, into a single enzyme scaffold, a concept we name as *plurizyme* (adding the latin root *pluri*: multiplicity). Given that each catalytic site in enzymes might have different architecture, one could precisely control catalytic activities and stabilities. Several questions about this procedure were expressed in the introduction, and addressed here. We demonstrated that it is possible to add catalytically efficient artificial active sites into enzymes containing a native one, and that the two active sites can coexist in the same polypeptide without compromising the original activity. Most importantly, we have shown that adding an extra active site can not only significantly increase catalytic performance but also confer additional properties to the original enzyme. In particular, we conferred improved stereo-specificity and

expanded the temperature range in which the enzyme can be active. More importantly, through reporting the crystal structure of an enzyme with two active sites soaked with ligand and in combination with CD analysis, we demonstrate that substrate binding can concomitantly occur in two sites, each inducing different additive conformation changes. Note that quantification of conformational changes in a number of enzymes upon substrate binding as part of the enzymatic mechanisms in a single active site exist in the literature, but their quantification when two sites coexist has no precedent.

The results further demonstrated that this approach is reproducible and can be easily extended to other serine ester-hydrolase and can confer stringent stereo-specificity to a native protein initially non-enantiospecific. We should emphasise that we obtained two *plurizymes* in the two ester-hydrolases that we attempted. As such, the present study could provide a platform to improve the bio-catalytic capability and to broaden the catalytic temperature window of naturally occurring enzymes through a synergetic combination of native and artificial active sites. Creating enzymes with multiple reactive groups may have important implications in future biotechnology, to help expanding their substrate range, increasing competitiveness by boosting their catalytic performance and stereo-specificity or modifying their operational conditions. By combining natural biological diversity potential with the rational active site design, new bio-catalytic systems with improved properties for sustainable catalysis can become feasible. The final catalytic outcomes would depend on the architecture of both the native and the artificial sites, and may vary from enzyme to enzyme.

One open question is: How frequently can a catalytic triad and an oxyanion hole be introduced after the docking simulation is done in ester-hydrolases? As stated, we have proven this by just picking two random ester-hydrolases, one being among those with the broader substrate spectra, and one with a narrow substrate spectra[13], but these calculations may be needed in a broader sense to find whether this is a "general" methodology in other "mother" ester-hydrolases or other enzymes that are not naturally esterase's. Regardless, the results reported here provide unambiguous evidences demonstrating that incorporation of a second active site with the same chemistry catalyzed as the native one, is reproducible and may significantly improve the catalytic capabilities of "mother" ester hydrolases. We hypothesized it can open new catalytic opportunities, an aspect that should be studied in depth in the future by combining active sites with different chemistry, so that to create multi-catalytic systems in one.

Methods

A full description of the methods is available in Supporting Information. The atomic coordinates have been deposited in the Protein Data Bank under accession numbers 6I8F and 6I8D.

References

1. Bren, K. L. Engineered biomolecular catalysts. *J. Am. Chem. Soc.* **139**, 14331-14334 (2017).

2. Ebert, M. C. & Pelletier, J. N. Computational tools for enzyme improvement: why everyone can - and should - use them. *Curr. Opin. Chem. Biol.* **37**, 89-96 (2017).
3. Arnold, F. H. Directed evolution: Bringing new chemistry to life. *Angew Chem. Int. Ed. Engl.* **57**, 4143-4148 (2018).
4. Acebes, S. et al. Rational enzyme engineering through biophysical and biochemical modeling. *ACS Catal.* **6**, 1624-1629 (2016).
5. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448**, 828-831 (2007).
6. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387-1391 (2008).
7. Richter, F. et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **134**, 16197-16206 (2012).
8. Rufo, C. M. et al. Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem.* **6**, 303-309 (2014).
9. Moroz, Y. S. et al. New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. *J. Am. Chem. Soc.* **137**, 14905-14911 (2015).
10. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418-421 (2013).
11. Khersonsky, O. et al. Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407**, 391-412 (2011).
12. Röthlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).
13. Santiago, G. et al. Rational engineering of multiple active sites in an ester hydrolase. *Biochemistry* **57**, 2245-2255 (2018).
14. Aranda, J. et al. The catalytic mechanism of carboxylesterases: a computational study. *Biochemistry* (Mosc.) **53**, 5820-5829 (2014).
15. Martínez-Martínez, M. et al. Ferrer M. Determinants and prediction of esterase substrate promiscuity patterns. *ACS Chem. Biol.* **13**, 225-234 (2018).
16. Wang, T. et al. Rational redesign of the active site of selenosubtilisin with strongly enhanced glutathione peroxidase activity. *J. Catal.* **359**, 27-35 (2018).
17. Hoque, M. A. et al. Stepwise loop insertion strategy for active site remodeling to generate novel enzyme functions. *ACS Chem. Biol.* **12**, 1188-1193 (2017).
18. Payer, S. E. et al. A Rational active-site redesign converts a decarboxylase into a C=C hydratase: "Tethered Acetate" supports enantioselective hydration of 4-hydroxystyrenes. *ACS Catal.* **8**, 2438-2442 (2018).
19. Zastrow, M. L. & Pecoraro, V. L. Influence of active site location on catalytic activity in de novo-designed zinc metalloenzymes. *J. Am. Chem. Soc.* **135**, 5895-5903 (2013).

20. Ross, M. R. et al. Histidine orientation modulates the structure and dynamics of a de novo metalloenzyme active site. *J. Am. Chem. Soc.* **137**, 10164-10176 (2015).
21. Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **51**, 3933-3940 (2012).
22. Khare, S. D. et al. Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat. Chem. Biol.* **8**, 294-300 (2012).
23. Zastrow, M. L. et al. Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **4**, 118-123 (2012).
24. Dydio, P. et al. An artificial metalloenzyme with the kinetics of native enzymes. *Science* **354**, 102-106 (2016).
25. Zandonella, G. et al. Interactions of fluorescent triacylglycerol analogs covalently bound to the active site of a lipase from *Rhizopus oryzae*. *Eur. J. Biochem.* **262**, 63-69 (1999).
26. Tokuriki, N. et al. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.* **3**, 1257 (2012).
27. Louis-Jeune, C., Andrade-Navarro, M. A. & Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins* **80**, 374-381 (2012).

Acknowledgements

This work was funded by grant 'INMARE' from the European Union's Horizon 2020 (grant agreement no. 634486), grants PCIN-2017-078 (within the Marine Biotechnology ERA-NET), CTQ2016-79138-R, BIO2017-85522-R, and BIO2016-76601-C3-3-R from the Ministry of Science, Innovation and Universities (former Spanish Ministry of Economy, Industry and Competitiveness), and grant 201980E031 funded by CSIC. P.N.G. and R.B. acknowledge the support of the UK Biotechnology and Biological Sciences Research Council (BBSRC; grant No. BB/M029085/1), and the Centre of Environmental Biotechnology Project and the Supercomputing Wales project, which are partly funded by the European Regional Development Fund (ERDF) via the Welsh Government. The authors gratefully acknowledge financial support provided by the ERDF. C. Coscolín thanks the Spanish Ministry of Economy, Industry and Competitiveness for a PhD fellowship (Grant BES-2015-073829). Authors would like to acknowledge Sergio Ciordia and María C. Mena for MALDI-TOF/TOF analysis. We thank the staff of the European Synchrotron Radiation Facility (ESRF), Grenoble, France, for providing access and for technical assistance at beamline ID30A-1/MASSIf-1 and the Synchrotron Radiation Source at Alba (Barcelona, Spain) for assistance at BL13-XALOC beamline.

Author contributions

SA, GS and IC-R contributed equally to this work. The manuscript was written through contributions of M. Ferrer, V. Guallar and J. Sanz-Aparicio. All authors have given approval to the final version of the manuscript. SA, CC, MM-M, HM and PNG contributed to site-directed mutagenesis, and protein expression, purification and characterization. RB contributed to biochemical data analysis. GG and VG conducted the PELE simulations and molecular dynamics. IC-R and JS-A did the crystallization and X-ray structure determination. RM and SA contributed to secondary structure analysis by circular dichroism. MF and VG conceived the work. MF wrote the initial draft of the manuscript which further written through contributions of MF, VG and JS-A.

Competing interests

The authors declare no competing financial interest.

Additional information

Supplementary information in available for this paper.

Materials and correspondence should be sent to M.F or V.G

SUPPORTING INFORMATION

Plurizyme: catalytic advantages of having two active sites in an enzyme

Sandra Alonso[1,8], Gerard Santiago[2,8], Isabel Cea-Rama[3,8], Cristina Coscolín[1], Laura Fernández-López[1], Mónica Martínez-Martínez[1], David Almendral[1], Helena Marrero[1], Ruth Matesanz[4], Rafael Bargiela[5,6], Peter N. Golyshin[5,6], Julia Sanz-Aparicio[3,9,*], Manuel Ferrer[1,9, *] and Víctor Guallar[2,7,9,*]

[1]Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), 28049 Madrid, Spain. [2]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain. [3]Department of Crystallography & Structural Biology, Institute of Physical Chemistry "Rocasolano", CSIC, 28006 Madrid, Spain. [4]Spectroscopy Laboratory, Centro de Investigaciones Biológicas, CSIC, 28040 Madrid, Spain. [5]School of Natural Sciences, Bangor University, LL57 2UW Bangor, UK. [6]Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, UK. [7]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain. [8]These authors contributed equally: Sandra Alonso, Gerard Santiago, Isabel Cea-Rama. [9]These authors equally coordinated the work: Julia Sanz-Aparicio, Manuel Ferrer, Víctor Guallar. *e-mail: *(J.S.) Email: xjulia@iqfr.csic.es (J.S-A.); mferrer@icp.csic.es (M.F.); victor.guallar@bsc.es (V.G.).

Supporting Introduction

Note 1

Summary of the residues constituting the native and artificial sites in EH1 hydrolase

The native active site of the target $EH1_A$ hydrolase is formed by a catalytic triad formed by Ser161, Asp256 and His286 and an oxyanion hole formed by Gly88, Gly89 and Gly90 (PDB code 5JD4). The artificial active site is formed by a catalytic triad formed by Ser211, Asp25 and His214 and an oxyanion hole formed by Gly207, Tyr208 and Phe209. The artificial $EH1_B$ hydrolase variant was created by introducing Glu25Asp, Leu214His and Ser161Ala substitutions into the wild-type sequence ($EH1_A$), so that this variant would employ a new catalytic triad (Ser211, Asp25 and His214) with Ser211 as the nucleophile and a new oxyanion hole (Gly207, Tyr208 and Phe209). Note that the Ser161Ala mutation abolished the activity of the native site (Ser161 being the nucleophile). The artificial $EH1_{AB}$ variant was created by incorporating the Glu25Asp and Leu214His substitutions into the wild-type sequence ($EH1_A$). This variant would employ two active sites, one formed by the Ser161, Asp256 and His286 triad

and the Gly88, Gly89 and Gly90 oxyanion hole, and a second formed by Ser211, Asp25 and His214 triad and Gly207, Tyr208 and Phe209 oxyanion hole. The design and creation of the $EH1_B$ and $EH1_{AB}$ variants and unambiguously confirmation that Ser161, Asp256, and His286 (conforming the native sites) and Ser211, Asp25, His214 (conforming the artificial site) are the functional groups supporting catalysis, which is constituted by a catalytic triad and not a diad or by a histidine or an aspartic residue, have been extensively described by Santiago et al. (2018)[1].

The $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ variants were created by adding R23G mutations to the $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ variants. Mutations Ser161Ala, Asp256Gln, and His286Phe (conforming the native site) and Ser211Ala, Asp25Gln, His214Phe (conforming the artificial site) were herein performed in $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$. The corresponding mutants were produced and tested as described by Santiago et al. (2018)[1]. We also found that those mutations (individually and in combination) fully arrested the activity for the $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ variants containing R23G mutation.

Supporting Results

Note 1:

Determination of specific activity, V_{max}, K_m, k_{cat} and k_{cat}/K_m were evaluated using an indirect method based on a pH-indicator assay in which the hydrolysis of an ester is followed colorimetrically. The method and conditions have been extensively described elsewhere[1,2].

R23G mutation does not have any side effect on $EH1_A$

No statistically significant differences were found in terms of V_{max} for glyceryl tributyrate and k_{cat} values for all the 79 esters being converted by both $EH1_A$ and $EH1_{A1}$ (Fig. 1, Table 1, Supporting Table S1). Thus, R23G mutation in the remodeled site did not cause any local negative side effect in the catalytic performance of the original enzyme.

R23G mutation improved the turnover number and substrate range of the artificial site

As shown in Table 1 and Fig. 2, the R23G mutation caused a significant increase in the substrate range of the artificial active site, as $EH1_{B1}$ hydrolyzed 58 esters, whereas $EH1_B$ was only able to hydrolyze 24 esters. It also caused a significant increase in k_{cat} of more than 500-fold for best-hydrolyzed ester (Table 1), and an average improvement of up to ~1600-fold (k_{cat} range from ~22000- to 15-fold) when all common esters converted were considered (Supporting Table S1). Esters hydrolyzed by $EH1_{B1}$ but not by $EH1_B$ include short to large alkyl and alkenyl-like fatty acid esters (12 in total), 9 mono-aromatic esters, 2 poly-aromatic esters, 7 chiral esters and 3 other esters, to cite some. As stated, this mutant was designed aiming to improve promiscuity by increasing its effective volume[1]. The experimental data reinforces this correlation, as the remodeled artificial site was not only capable of accepting a 2.4-fold higher number of substrates compared to the nonremodeled site but was also capable of accepting very large and voluminous substrates (Fig. 2; Supporting Table S1), such as methyl decanoate, benzylparaben and phenylethyl cinnamate (k_{cat}: up to 4.27 ± 0.12 min^{-1}). Analysis of the specificity for alkyl esters revealed that $EH1_{B1}$ was also capable of accepting esters with alkyl chains from acetate to laurate and alcohol substituents from methyl to octyl (Fig. 2).

$EH1_{B1}$ does show also a substrate range and V_{max} for the common ester hydrolase substrate glyceryl tributyrate above the interquartile range of most native EHs, a situation contrasting with that of the original $EH1_B$ design, which was ranked among the less active EHs (Table 1). The remodeling also approximated the performance of the artificial and native sites, as $EH1_{B1}$ was only 3.3-fold less active (k_{cat}) than $EH1_{A1}$ for best-hydrolyzed substrate (Table 1), and ~19-fold on average (range from ~134- to 1.3-fold) for the 55 common esters converted (Supporting Table S1). In five cases, the catalytic activity of $EH1_{B1}$ exceeded that of $EH1_{A1}$: methyl 3-hydroxybenzoate (70-fold), vinyl propionate (29-fold), methyl hexanoate (9-fold), methyl (2R)-2-phenylpropanoate (8-fold), and vinyl crotonate (3.6-fold). Without R23G mutation the differences, in k_{cat}, were ~1900-fold for the best substrates and ~11200-fold on average (range from ~111000- to 4.3-fold) for all common esters converted, and none of the esters was better hydrolyzed by the artificial site.

$EH1_{AB1}$ is better performing than $EH1_{AB}$

The k_{cat} of $EH1_{AB1}$ for best-hydrolyzed substrate was ~13-fold higher than that of $EH1_{AB}$ (Table 1), and ~113-fold on average (range from ~2050- to 3.2-fold, depending on the ester) for all common esters converted (Supporting Table S1). This is a direct consequence of the increased catalytic performance of the remodeled extra site in $EH1_{AB1}$.

Note 2:

Crystallization and X-ray structure determination of $EH1_{AB1}$

Stereo views of the electron density map at the Ser161 and Ser211 catalytic sites in the free and complexed crystals are provided in Supporting Fig. S2 and S3.

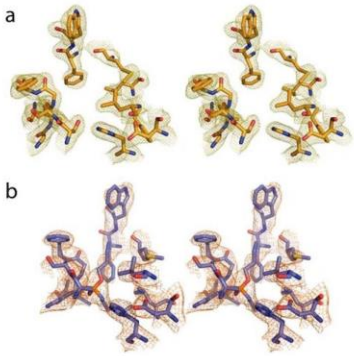


Fig. S1 Stereo view of the electron density map at the Ser161 catalytic site in the free (a) and complexed (b) crystals. The 2Fo-Fc map is contoured a 1.0 σ.

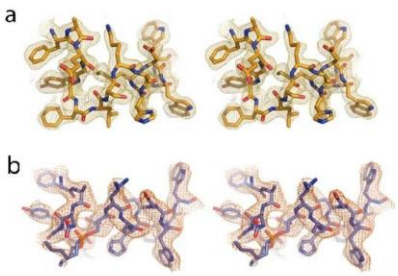


Fig. S2 Stereo view of the electron density map at helix Phe204-Trp219, containing the Ser 211/His214 motif, in the free (a) and complexed (b) crystals. The 2Fo-Fc map is contoured a 0.9 σ.

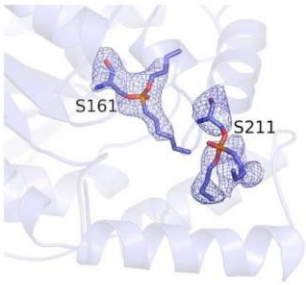
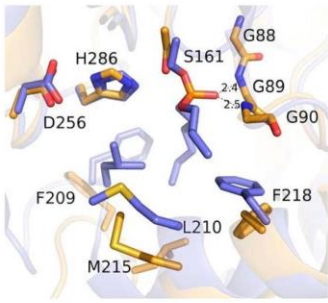


Fig S3. Polder omit maps calculated at the Ser161 and Ser211 catalytic sites, displayed at 3.0 and 2.7 s cutoff, respectively.

a



b

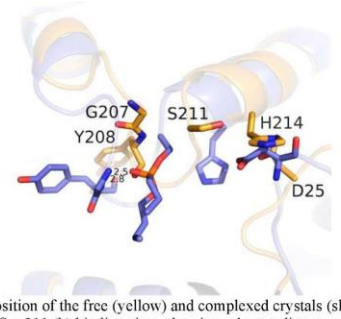


Fig. S4 Superimposition of the free (yellow) and complexed crystals (slate) illustrating the Ser161 (a) and Ser 211 (b) binding sites, showing relevant distances as dashed lines.

Note 3:

Suicide inhibitor as substrates for EH1

As substrate for soaking and binding tests we used the so-called suicide inhibitors, namely, methyl, butyl and octyl 4-nitrophenyl hexylphosphonate. The 4-nitrophenyl phosphonate inhibitors are susceptible to nucleophilic attack by the catalytic Ser of the active site of lipases and related enzymes. This leads to stable covalent modification of the serine residue in the active site and complete inactivation of the enzyme[3,4]. According to our inhibition experiments, treatment with butyl 4-nitrophenyl hexylphosphonate (which was found as a substrate because of the release of *p*-nitrophenol by all 3 EH1 variants; not shown) resulted in inactivation of >99% for $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ when tested with glyceryl tripropionate (Supporting Fig. S5) in less than 20 sec. This substrate can thus be used for crystallographic structural analysis (soaking) and to investigate conformational changes of substrate-free and substrate-bound protein variants because of irreversible bounding to the corresponding nucleophiles.

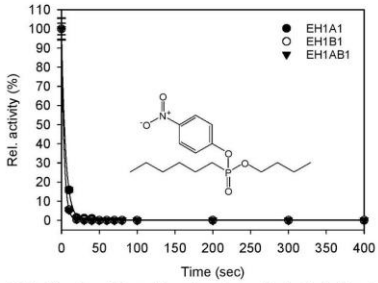


Fig. S5 Effect of butyl 4-nitrophenyl hexylphosphonate on the hydrolysis rate of glyceryl tripropionate by $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$. Enzyme variants were incubated with the inhibitor at 25°C as described in Experimental section and the enzyme activities were measured spectrophotometrically over time. The relative activities are shown as percentage of control (time 0). Reactions were performed in triplicates with standard deviations indicated. The structure of butyl 4-nitrophenyl hexylphosphonate is shown.

Note 4:

Reproducibility tests for introducing an extra reactive group to other ester hydrolases

To design a second catalytic triad we applied the same protocol using the PELE software to locate potential pockets where target substrates can be accommodated[4]. Glyceryl tripropionate was used as a target compound, as it was found as an ester commonly hydrolyzed by most ester hydrolases[1], including $EH102_A$ (Supporting Table S3). PELE simulations revealed a second binding site located ~23 Å from the native catalytic position at Ser15 (Supporting Fig. S6A) that already contains a histidine residue (His34). Thus, we computationally designed additional mutations, adding Asp and Ser residues to build a proper catalytic triad, taking special care concerning catalytic distances, and the resulting effective volume. We found that the Leu57Asp and Leu30Ser double mutant revealed good enzyme-substrate interaction energies (Supporting Fig. S6B) and a suitable catalytic position for the glyceryl tripropionate substrate (Supporting Fig. S6C). In addition, our results showed that residue Asn43 acts as a potential oxyanion hole, a key element in ester hydrolase catalysis (Supporting Fig. S6C). Moreover, extensive MD simulations indicate proper stabilization of the double mutant.

By using site-directed mutagenesis, we created two variants, referred to as $EH102_B$ and $EH102_{AB}$. $EH102_B$ introduced Leu57Asp, Leu30Ser and Ser15Ala substitutions, so that this variant would employ only the new catalytic triad; the second variant, $EH102_{AB}$, carried both the native and the artificial sites.

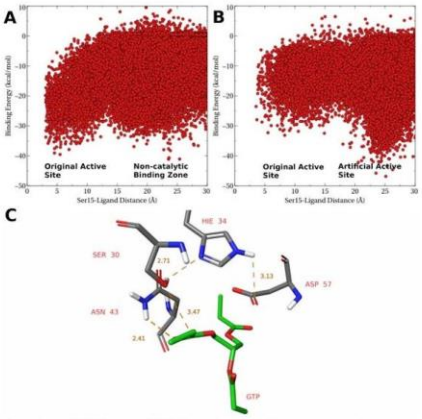


Fig. S6 PELE exploration of EH102. **a**, PELE exploration for glyceryl tripropionate diffusion in $EH102_A$. **b**, New catalytic triad interaction energies profile for $EH102_{AB}$. **c**, Glyceryl tripropionate binding mode as obtained with PELE.

The participation of Ser15 (in EH_{102A} and EH_{102AB}) and Ser30 (in EH_{102B} and EH_{102AB}) as the catalytic nucleophiles was confirmed by using the suicide inhibitors methyl, butyl and octyl 4-nitrophenyl hexylphosphonate. According to our inhibition experiments, treatment with the butyl derivative resulted in total inactivation of $EH102_A$, $EH102_B$ and $EH102_{AB}$ activity when tested with glyceryl tripropionate (Supporting Fig. S7). Methyl and octyl derivatives were also used as substrates and bound irreversibly but at lower rates (not shown). This unambiguously confirms that Ser15 and Ser30 are the functional nucleophilic groups supporting catalysis.

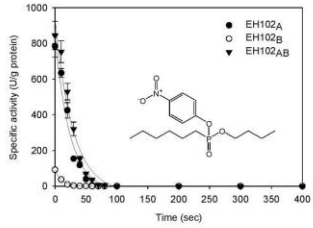


Fig. S7 Effect of butyl 4-nitrophenyl hexylphosphonate on the hydrolysis rate of glyceryl tripropionate by $EH102_A$, $EH102_B$ and $EH102_{AB}$. Enzyme variants were incubated with the inhibitor at 25ºC as described in Experimental section and the enzyme activities were measured spectrophotometrically over time. The relative activities are shown as percentage of control (time 0). Reactions were performed in triplicates with standard deviations indicated. The structure of butyl 4-nitrophenyl hexylphosphonate is shown.

Note 5:

The artificial site broadens the catalytic temperature window of EH102 and induced different conformation changes upon substrate binding

Using glyceryl tripropionate, we found that $EH102_A$ was strongly active (>70% rel. act.) at temperatures in the range 4-40ºC, with the activity progressively reduced at higher temperatures (Supporting Fig. S8a). The T_{opt} was ~12ºC°C. $EH102_B$, exhibited the reverse of this trend —it was less active at temperatures lower than 12ºC and more active at temperatures from 45-65ºC (Supporting Fig. S8a). The T_{opt} was ~25ºC, and the enzyme retained >70% activity at 16-55ºC. The enzyme with two reactive sites exhibited a hybrid pattern, retaining >70% rel. act. at temperatures from 4 to 55ºC, still retaining 56% of its activity at 65ºC (Supporting Fig. S8a). The two-active site enzyme thus retained the psychrophilic-like phenotype of the native site and the mesophilic-like phenotype introduced by the extra active site. Denaturing temperatures (T_d) of 52.83±0.95ºC, 53.11±0.18ºC, and 54.30±0.2.80ºC were calculated, by circular dichroism (CD) spectroscopy, for $EH102_A$, $EH102_B$, and $EH102_{AB}$, respectively (Supporting Fig. S8b-d).

As for the EH1 variants, distinct conformational changes were further revealed by circular dichroism in the native and artificial active site in EH102 scaffold (Supporting Fig. S8e-f). As shown, in the absence of substrate their CD spectra were similar in $EH102_A$, $EH102_B$ and $EH102_{AB}$. Upon substrate binding (butyl 4-nitrophenyl hexylphosphonate), the CD spectra change, the level of which varies among the 3 variants.

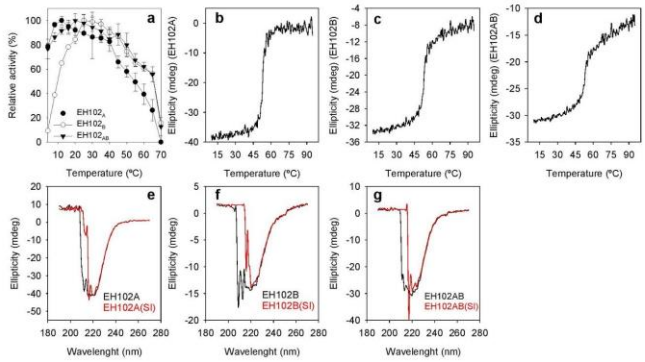


Fig. S8 Thermal and conformation changes associated to each of the EH102 protein variants. a, Temperature profiles of the purified $EH102_A$, $EH102_B$ and $EH102_{AB}$ variants. The data represent the relative percentages (%) of specific activity at pH 8.0, expressed as U mg^{-1} compared with the maximum activity when glyceryl tripropionate was used as the substrate. All data, calculated from three independent assays ± SD, are not fitted to any model. **b-d**, Thermal denaturation curves of $EH102_A$ (b), $EH102_B$ (c) and $EH102_{AB}$ (d) variants, measured by ellipticity changes at 220 nm and obtained at different temperatures and at pH 7.0. The results from the three variants (statistically similar) were averaged. **e-g**, CD analysis of $EH102_A$ (e), $EH102_B$ (f) and $EH1_{AB}$ (g) recorded at 25°C in the absence (black) or upon binding (red) of the suicide inhibitor (SI) butyl 4-nitrophenyl hexylphosphonate. Shown is the ellipticity in mdeg from 190 to 270 nm.

Note 6:

Protein purity and confirmation of mutations

Purity (>99%) was confirmed by SDS-polyacrylamide gel electrophoresis (Supporting Fig. S9) on 12% gels and Coomassie brilliant blue staining, as determined by densitometry.

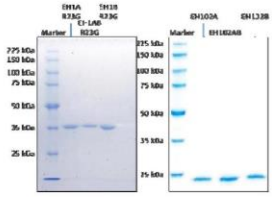


Fig. S9 Protein purity as determined by SDS-PAGE. A Coomassie brilliant blue (Protoblue Safe, National Diagnostics, GA, USA) stained SDS-PAGE gel in which a total of 20 µg of proteins purified after the His6-tag purification step followed by gel filtration step, are shown for EH1 (left panel) and EH102 (right panel) variants. Purity of the proteins was high (>99% by densitometry), which was further confirmed by mass spectrometry. The theoretical molecular mass of the native $EH1_A$ and $EH102_A$ proteins is 33936.23 and 24328.83 Da, respectively. SDS-PAGE was performed in a Mini PROTEAN electrophoresis system (Bio-Rad).

MALDI mass spectrometry of purified proteins was used to confirm the purity of the purified recombinant protein variants. We found that the molecular mass experimentally obtained (~35 kDa) in all cases by MALDI agree with the theoretical monoisotopic mass of intact proteins calculated on the basis of sequence information, supporting the recombinant proteins do show the expected molecular masses (Supporting Fig. S10). MALDI intact mass analysis can also interpret the full mass spectrum without any deconvolution step and the analysis therefore gives a good indication of purity of protein samples. As shown in Supporting Fig. S8, the chromatographic protein purity was higher than 99%.

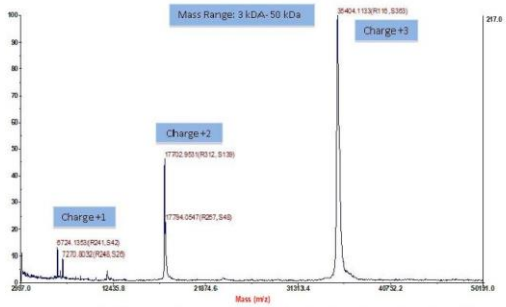


Fig. S10A. Analysis of $EH1_{A1}$ purified protein solution by MALDI-TOF/TOF. The mass (m)-to-charge (z) ratios (*m/z*) for the mass range from 3 to 50 kDa is shown. The MALDI analysis revealed a protein with the estimated molecular mass of ~35 kDa, with no apparent

contaminants. In the spectra, the peaks corresponding to the double and triple charge of the precursors are indicated.

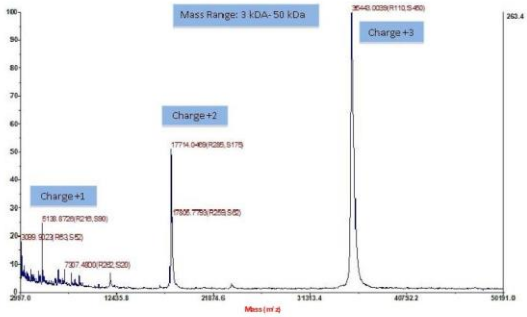


Fig. S10B. Analysis of $EH1_{B1}$ purified protein solution by MALDI-TOF/TOF. The mass (m)-to-charge (z) ratios (*m/z*) for the mass range from 3 to 50 kDa is shown. The MALDI analysis revealed a protein with the estimated molecular mass of ~35 kDa, with no apparent contaminants. In the spectra, the peaks corresponding to the double and triple charge of the precursors are indicated.

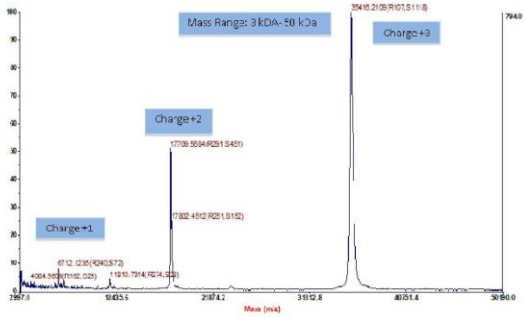


Fig. S10C. Analysis of $EH1_{AB1}$ purified protein solution by MALDI-TOF/TOF. The mass (m)-to-charge (z) ratios (*m/z*) for the mass range from 3 to 50 kDa is shown. The MALDI analysis revealed a protein with the estimated molecular mass of ~35 kDa, with no apparent contaminants. In the spectra, the peaks corresponding to the double and triple charge of the precursors are indicated.

MALDI-TOF/TOF is a technique that can be also utilized to analyze the peptides for primary sequence confirmation and presence of mutations. For this reason, to confirm that the proteins contained the introduced mutations, peptide fingerprint by MALDI-TOF/TOF was performed using the purified proteins, and mutations where unambiguously confirmed in $EH1_{A1}$ (Supporting Fig. S11), $EH1_{B1}$ (Supporting Fig. S12) and $EH1_{AB1}$ (Supporting Fig. S13). We also used DNA sequencing to also ensure that the mutations were correctly introduced.

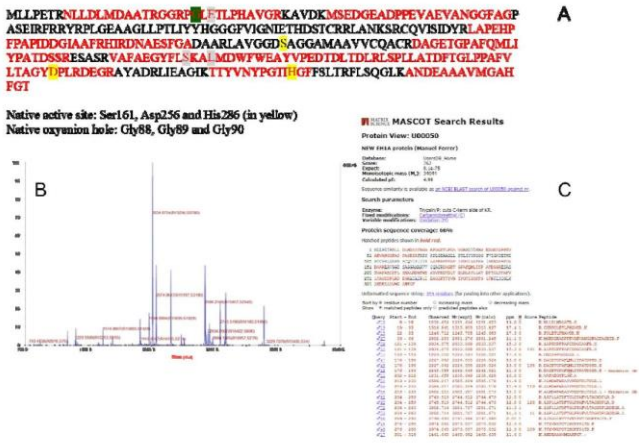


Fig. S11. MALDI-TOF/TOF analysis of $EH1_{A1}$ protein solution. (A) $EH1_{A1}$ sequence in which the detected peptides are shown in red color. In yellow are the residues conforming the native catalytic triad. In gray color the original residues that were needed to further accommodate the artificial active site, are indicated. In dark green color the residue which has been found key for improving the catalytic performance of the artificial site is indicated; this residue (number 23) was originally an Arg in the native $EH1_A$ protein, but was substituted by Gly in $EH1_{A1}$. (B) MS/MS chromatogram which shows the mass (m)-to-charge (z) ratios (*m/z*) of identified peptides, with those being specific for $EH1_{A1}$ being specifically indicated. (C) MASCOT search results of the mass of identified peptides.

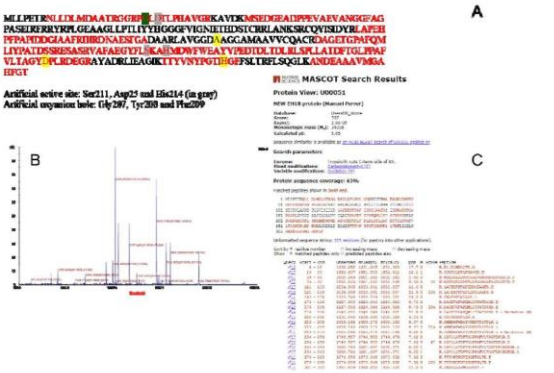


Fig. S12. MALDI-TOF/TOF analysis of $EH1_{B1}$ protein solution. A) $EH1_{B1}$ sequence in which the detected peptides are shown in red color. In yellow are the residues conforming the native catalytic triad. In gray color the original residues that were needed to further accommodate the artificial active site, are indicated. In dark green color the residue which has been found key for improving the catalytic performance of the artificial site is indicated; this residue (number 23) was originally an Arg in the native $EH1_A$ protein, but was substituted by Gly in $EH1_{A1}$. B) MS/MS chromatogram which shows the mass (m)-to-charge (z) ratios (m/z) of identified peptides, with those being specific for $EH1_{B1}$ being specifically indicated. C) MASCOT search results of the mass of identified peptides.

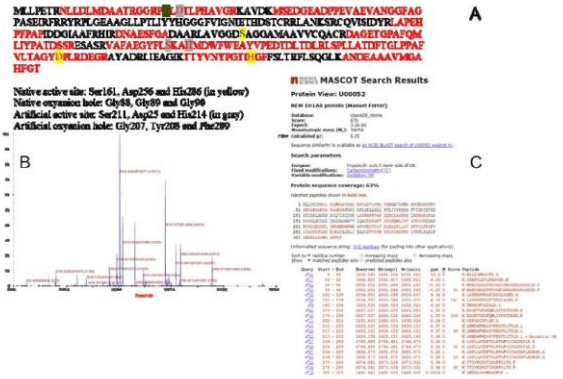


Fig. S13. MALDI-TOF/TOF analysis of $EH1_{AB1}$ protein solution. A) $EH1_{AB1}$ sequence in which the detected peptides are shown in red color. In yellow are the residues conforming the native catalytic triad. In gray color, the residues constituting the artificial active site are indicated. In

dark green, the residue which has been found key for improving the catalytic performance of the artificial site is indicated. B) MS/MS chromatogram which shows the mass (m)-to-charge (*z*) ratios (*m/z*) of identified peptides, with those being specific for $EH1_{AB1}$ being specifically indicated. C) MASCOT search results of the mass of identified peptides.

Supporting Methods

Chemicals, oligonucleotides, source of enzyme, strains, and medium. The sources of all chemicals (of the purest grade available), oligonucleotides for DNA amplification and serine ester hydrolases $EH1_A$, $EH1_B$, $EH1_{AB}$, and $EH102_A$ (available in the expression vector pET46 Ek/LIC plasmid in *Escherichia coli* BL21 as a host) used in the present study, were reported[1,4].

Site-directed mutagenesis. Mutagenic PCR was developed using the QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies, Cheadle, UK), as described previously[1]. The forward primers used to generate the $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ variants with the R23G mutation are as follows: AGG GGC GGC CGG CCC GGG CTG GAG ACC CTG CCG or AGG GGC GGC CGG CCC GGG CTG GAT ACC CTG CCG. The forward primers used to generate the $EH102_B$ and $EH102_{AB}$ variants are as follows: Leu57Asp, CGA ACG GCG GCT GCG Gga CGT CAA CCA GGG CAA C; Leu30Ser, GAT CGG CGG CGA GGG Gag CTT CAA CGC CCA TGG C; Ser15Ala, CCC GCG TCC GTG ATC GcG TCG CCG ATG ATG ACC. To obtain variants containing Ser161Ala, Asp256Gln, and His286Phe (conforming the native sites) and Ser211Ala, Asp25Gln, His214Phe mutations, individually or in combination, the pET46 Ek/LIC plasmid containing $EH1_{A1}$, $EH1_{B1}$, or $EH1_{AB1}$ DNA inserts the primers Ser211Ala Fwd (GCC GAA GGC TAC TTC CTC GCC AAG GCG CAC ATG GAC TGG), Ser161Ala Fwd (GTG GGC GGC GAT GCG GCG GGC GGC G), Asp25Gln Fwd (CGG CCC GGG CTG GCG ACC CTG CCG CAT GC), Asp256Gln Fwd (ACC GCC GGC TAC CAA CCG CTG CGC GAC G), His214Phe Fwd (TTC CTC AGC AAG GCG TTC ATG GAC TGG TTC TGG G) and His286Phe Fwd (T CCC GGC ACC ATC TTC GGC TTC TTC TCG), were used. Mutagenic PCR conditions were as described[1].

Gene expression and protein purification. Expression (performed in Luria-Bertani medium with shaking at 250 rpm and 16ºC overnight) and purification (at 4ºC by binding to a Ni-NTA His-Bind resin, followed by size-exclusion chromatography) of the EH1 (Supporting Table S4) and EH102 (Supporting Table S5) protein variants, His-tagged at the N-terminus, were performed as previously described[4]. Purified proteins were stored at -20ºC. Purity (>99%) was confirmed by SDS-polyacrylamide gel electrophoresis on 12% gels and Coomassie brilliant blue staining, along with matrix-assisted laser desorption/ionization-time-of-flight/time-of-flight analysis which was also used to confirm the mutations introduced, as previously described (see Note 6 in Supporting Results)[1]. The protein concentration was determined spectrophotometrically (at 280 nm) according to the corresponding amino acid sequence (www.expasy.org/tools/protparam.html) using a Take3 plate compatible with a Synergy HT Multi-Mode Microplate Reader.

Enzyme assays. Specific activity (units mg^{-1}), V_{max}, k_{cat} and K_m determinations for ester substrates were assayed using a pH indicator (phenol red; $\varepsilon_{550\ nm} = 8450\ M^{-1}\ cm^{-1}$) assay at 550 nm in 384-well plates as previously described[1], although with minor changes. Briefly, for K_m determinations the amount of protein was 0.2 µg, and 50 mM of each ester were used to ensure substrate saturation while determining k_{cat} values. If otherwise not indicated, reactions were performed, in triplicates and corrected for non-enzymatic transformation, at 30ºC and in 5 mM N-(2-hydroxyethyl)piperazine-N'-(3-propanesulfonic acid (EPPS) buffer. Temperatures between 8 and 65 °C were tested for determination of the conditions under which each protein variant displayed maximal activity, using α-naphthyl acetate and glyceryl tripropionate.

We used the suicide inhibitors methyl, butyl and octyl 4-nitrophenyl hexylphosphonate (Eurocodis Bioscience GmbH, Wien, Austria) to test for the ability to inactivate all ester hydrolases reported in this study. For this experiment, 7.5 mM proteins were incubated with 15 mM (for $EH1_{A1}$, $EH1_{B1}$, $EH102_A$, and $EH102_B$) or 30 mM (for $EH1_{AB1}$ and $EH102_{AB}$) inhibitors (from a 0.3 M-stock solution in dimethylsulfoxide). Incubation was performed in 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer at pH 7.0, with a total volume of 200 µL. The progress of inhibition at 25ºC was followed by taking aliquots (10 µL) at

different time intervals and measuring the remaining enzyme activity using glyceryl tripropionate[4]. In all cases, reactions were performed in triplicate. The capacity of the proteins to perform hydrolysis of the suicide inhibitor was also followed at 405 nm based on the release of *p*-nitrophenol (ε_{pNP} at 405 nm = 14280 M^{-1} cm^{-1}).

Circular dichroism to estimate the protein secondary structure and thermal denaturation. Circular dichroism (CD) spectra were acquired between 190 and 270 nm with a Jasco J-720 spectropolarimeter equipped with a Peltier temperature controller, employing a 0.1 mm cell at 25°C. Spectra were analyzed, and denaturation temperatures were determined at 220 nm between 10 and 85ºC at a rate of 30ºC per hour, as reported previously[6,7]. In all cases, a protein concentration of 0.5 mg ml^{-1} was used for CD analyses. We used the suicide inhibitor butyl 4-nitrophenyl hexylphosphonate (Eurocodis Bioscience GmbH, Wien, Austria) to evaluate the conformational changes induced in $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ as well as $EH102_A$, $EH102_B$ and $EH102_{AB}$. For this analysis, the CD spectra of each protein (0.13 mM) were recorded at 25ºC between 190 and 270 nm in a 0.1 mm cell. A final volume of protein solution of 200 µL was used. Then, 0.2 µL of the inhibitor stock solution in dimethylsulfoxide (0.26 M) was added for $EH1_{A1}$ and $EH1_{B1}$, and 0.4 µL was added for $EH1_{AB1}$, which corresponds to a protein:inhibitor ratio of 1:2 for $EH1_{A1}$ and $EH1_{B1}$ and 1:4 for $EH1_{AB1}$. Considering that EH1 is a dimer, a ratio of 1:1 per site and per monomer was thus used. A total of 0.8 mg/ml of $EH102_A$, $EH102_B$ and $EH102_{AB}$ were used, with the corresponding amount of inhibitor. . After 5 min incubation at 25ºC, the CD spectra were acquired. Estimates of the protein secondary structure were performed from the CD data as reported elsewhere[5].

PELE simulations. We modeled the binding mode of glyceryl tripropionate with $EH1_A$ and $EH102_A$ structures (PDB codes: 5JD4 and 5JD3), as described previously[4]. Crystal structures were taken from the PDB codes mentioned, and the protonation of the titratable residues was approximated using Protein Preparation Wizard (PROPKA) and a visual inspection. At pH 8, the value at which reactions were evaluated in this study, the catalytic triad residues form a serine-histidine and histidine-aspartic hydrogen-bonding inner network. The chosen ligand (glyceryl tripropionate) parameters were extracted from full quantum mechanics optimizations using Jaguar from Schrodinger. We used PELE software to explore the protein space for noncatalytic binding sites and to sample the binding modes of the ligand. PELE is a Monte Carlo algorithm based on a sequence of perturbations and relaxations and a Metropolis acceptance test.[8] In the first step, the ligand is subjected to random rotations and translations, while the protein is perturbed based on the anisotropic network model (ANM). The maximum allowed translation for the ligand perturbation was 1.5 Å, and the maximum rotation was 20°. During the protein perturbation, all atoms were displaced by a maximum of 0.5 Å by moving the α-carbons following a random linear combination of the six lowest eigenvectors obtained in the ANM model. The relaxation step included the repositioning of all amino acid side chains within 6 Å of the ligand and the five side chains with the highest energy increases during the previous ANM step. The relaxation stage ended with a truncated Newton minimization using the OPLS all-atom force field and an implicit surface-generalized Born continuum solvent. The new proposed minima were then accepted or rejected based on a Metropolis test. The substrate binding plots contained all accepted conformations for three 12-h simulations performed using 200 processors.

Molecular dynamics. We performed 250-ns molecular dynamics with GROMACS to ensure enzymatic stability[9]. After appropriate system preparation, as was previously described[4], an orthorhombic box (minimum distance of 10 Å) was introduced. Then, equilibration was performed, followed by NPT simulation at 300 K and 1 atm with the OPLS-2005 force field. Temperature regulation was controlled by a Berendsen thermostat.

Crystallization and X-ray structure determination of $EH1_{AB1}$. Initial crystallization conditions were explored by high-throughput techniques with a NanoDrop robot (Innovadyne Tecnhologies Inc.), using different protein concentrations in 40 mM HEPES pH 7, 150 mM NaCl, and commercial screens: Index (Hampton Research), JBScreen PACT++ (Jena Bioscience) and JCSG+ Suite (Qiagen). Further optimizations were carried out and thin bar-shape crystals of $EH1_{AB1}$ were grown using 1 µL of protein (30 mg/mL) and 0.5 µL of precipitant solution (45% PEG P400, 0.1 M BIS-TRIS pH 6.5). The complex with the suicide

inhibitor, a derivative of butyl 4-nitrophenyl hexylphosphonate, was obtained by soaking crystals in precipitant solution supplemented with 10 mM butyl 4-nitrophenyl hexylphosphonate for 3 hours.

For data collection, crystals were transferred to cryoprotectant solutions consisting of mother liquor plus 20-25 % (v/v) glycerol, before being cooled in liquid nitrogen. Diffraction data were collected using synchrotron radiation on the XALOC beamline at ALBA (Cerdanyola del Vallés, Spain) and the ID30A-1/MASSIf-1 beamline at ESRF (Grenoble, France). Diffraction images were processed with XDS[10] and merged using AIMLESS[11] from the CCP4 package[12]. Both crystals were indexed in the $P2_12_12_1$ space group, with two molecules in the asymmetric unit and 55% solvent content within the unit cell. The data-collection statistics are given in Supporting Table S6. The structures of $EH1_{AB1}$ and the $EH1_{AB1}$-butyl hexylphosphonate complex were solved by Molecular Replacement with MOLREP[13] using the coordinates of the native protein (PDB code 5JD4). Crystallographic refinement was performed using the program REFMAC[14] within the CCP4 suite with flat bulk-solvent correction, maximum likelihood target features and local non-crystallographic symmetry (NCS). Two regions, 19-33 and 203-218 building the secondary catalytic site, were excluded from the NCS groups in the ligand-free structure. Free R-factor was calculated using a subset of 5% randomly selected structure-factor amplitudes that were excluded from automated refinement. At the later stages, ligands were manually built into the electron density maps with Coot[15] and water molecules were included in the model, which, combined with more rounds of restrained refinement, reached the R factors listed in Supporting Table S6. For butyl 4-nitrophenyl hexylphosphonate inhibitor derivative, not present in the Protein Data Bank, a model was built using MacPyMOLX11Hybrid (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The model was used to automatically generate coordinates and molecular topologies with eLBOW[16] suitable for REFMAC refinement. The figures were generated with PyMOL. The crystallographic statistics of $EH1_{AB1}$ are listed in Supporting Table S6.

Gas chromatography (GC) analysis for determination of chiral selectivity. Enantioselectivity was evaluated (at 30ºC) by using kinetic resolution of methyl (2*R*)-2-phenylpropanoate and methyl (2*S*)-2-phenylpropanoate, or (+)-ethyl (*R*)-lactate and (−)-ethyl (*S*)-lactate, or ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate and ethyl (*R*)-(+)-4-chloro-3-hydroxybutyrate. Briefly, 2 µL of racemic ester (from a stock solution of 200 mg/mL in acetonitrile) were added to 96 µL of 40 mM HEPES buffer, pH 7.0. Then, 2 µL of enzyme solution (from a stock solution of 10.0 mg/ml in 40 mM HEPES buffer, pH 7.0) were added. After 60 min (for $EH1_{B1}$ and $EH1_{AB1}$) or 24 h (for $EH1_{A1}$), reactions with racemic mixtures of methyl (2*R*)-2-phenylpropanoate and methyl (2*S*)-2-phenylpropanoate, were stopped by adding 1800 µL HPLC-grade methanol and the reaction products analysed by GC (for details see GC Chromatograms at the end of the Supporting Information) as described Coscolín et al. (2019)[17]. For EH102 variants, reaction time was fixed to 60 min.

Supporting References

1. Santiago, G. et al. Rational engineering of multiple active sites in an ester hydrolase. *Biochemistry* **57**, 2245-2255 (2018).
2. Martínez-Martínez, M. et al. Determinants and prediction of esterase substrate promiscuity patterns. *ACS Chem. Biol.* **13**, 225-234 (2018).
3. Zandonella, G. et al. Interactions of fluorescent triacylglycerol analogs covalently bound to the active site of a lipase from *Rhizopus oryzae*. *Eur. J. Biochem.* **262**, 63-69 (1999).
4. Tokuriki, N. et al. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.* **2012**, *3*, 1257.
5. Martínez-Martínez, M. et al. Biochemical Diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol.* **79**, 3553-3562 (2013).
6. Ferrer, M. et al. A purple acidophilic di-ferric DNA ligase from *Ferroplasma*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8878-8883 (2008).
7. Alcaide M. et al. Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol.* **17**, 332-345 (2015).

8. Lecina, D., Gilabert, J.F. & Guallar, V. Adaptive simulations, towards interactive protein-ligand modeling. *Sci Rep*. **7**, 8466 (2017).
9. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
10. Kabsch, W. XDS. *Acta Cryst.* **D66**, 125-132 (2010).
11. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Cryst* **D69**, 1204-1214 (2013).
12. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Cryst.* **D67**, 235-242 (2011).
13. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025 (1997).
14. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D535**, 240-255 (1997).
15. Emsley, P. et al. Features and development of Coot. *Acta Cryst.* **D60**, 2126-2132 (2004).
16. Moriarty, N.W., Grosse-Kunstleve, R.W. & Adams, P.D. Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Cryst.* **D65**, 1074-1080 (2009).
17. Coscolín, C. Bioprospecting reveals class III ω-transaminases converting bulky ketones and environmentally relevant polyamines. *Appl. Environ. Microbiol*. 85, e02404-18 (2019).

Table S1. k_{cat} (min^{-1}) against a set of structurally different esters. The assays were performed at 30 °C and pH 8.0, as triplicates, with the average value and standard deviation shown. K_m values for all variants and esters tested range from 0.08±0.01 to 40.19±0.20 mM (see below). For this reason, 50 mM of each ester were used to ensure substrate saturation while determining k_{cat} values.

	k_{cat} (min^{-1})[1]					
Ester[1]	**$EH1_A$**	**$EH1_B$**	**$EH1_{AB}$**	**$EH1_{A1}$**	**$EH1_{B1}$**	**$EH1_{AB1}$**
Glyceryl triacetate	1115.98±13.27	0.051±0.014	5.12±0.9	1110.6±14.08	45.6±0.31	1195.47±6.75
Glyceryl tripropionate	2247.34±6.49	0.547±0.034	21.11±1.35	2205.4±18.67	221.85±3.63	2528.98±19.76
Glyceryl tributyrate	440.83±2.3	0.214±0.008	8.19±0.98	458.12±2.91	5.84±0.22	465.31±6.77
Methyl-2-bromobutyrate	9.14±1.79	0	0.12±0.05	9.23±0.16	0	9.39±6.97
Ethyl acetate	209.23±7.64	0	3.22±0.14	206.63±3.33	1.01±0.08	217.9±2.56
Propyl acetate	13.67±0.05	0	2.84±0.07	15.81±0.2	0	15.89±0.73
Butyl acetate	465.68±3.35	0.004±0.003	6.86±0.02	453.52±6.38	86.5±9.96	551.04±1.78
Hexyl acetate	34.11±0.45	0.007±0.003	6.88±0.06	33.47±0.49	9.39±0.04	48.14±0.49
Octyl acetate	147.04±9.45	0.005±0.001	3.31±0.06	152.92±1.16	6.53±0.22	147.15±0.78
Methyl butyrate	27.34±0.03	0	8.03±0.15	29.94±0.06	5.06±0.07	41.65±0.82
Methyl hexanoate	1.03±0.09	0.135±0.021	1.22±0.04	0.94±0.07	8.0±0.11	8.02±0.7
Methyl octanoate	107.22±4.15	0.01±0.002	4.02±0.14	109.55±2.2	14.84±0.09	123.02±1.23
Methyl decanoate	11.12±0.36	0	0.58±0.09	12.01±0.21	0.2±0.01	12.42±0.19
Ethyl propionate	207.93±4.51	0	3.9±0.35	201.91±3.2	1.65±0.08	207.67±2.16
Ethyl butyrate	82.13±3.48	0.007±0.002	13.78±0.28	83.52±0.13	2.69±0.09	86.24±1.7
Ethyl hexanoate	184.08±1.45	0.113±0.059	4.01±0.07	180.37±2.58	8.59±0.33	194.28±1.39
Ethyl octanoate	61.66±1.22	0.041±0.018	1.75±0.06	56.07±0.14	6.24±0.27	69.39±2.65
Ethyl decanoate	1.44±0.18	0	0.14±0.02	1.33±0.24	0	1.42±0.09
Ethyl dodecanoate	1.25±0.22	0	1.44±0.03	0.96±0.02	0	0.97±0.08
Propyl propionate	1.96±0.07	0	35.22±0.3	1.24±0.02	0	1.25±0.12
Propyl butyrate	0.48±0.03	0	115.46±0.74	0.44±0.08	0	0.47±0.14
Propyl hexanoate	154.73±4.6	0.075±0.03	5.27±0.13	156.36±1.98	10.23±0.66	176.18±1.27
Ethyl 2-ethylacetoacetate	13.2±1.24	0	0.52±0.02	13.46±0.27	9.76±0.13	31.73±0.7
Ethyl 2-methylacetoacetate	85.21±1.06	0	6.61±0.07	81.37±1.64	11.87±0.16	95.01±3.0
Ethyl 3-oxohexanoate	345.98±4.16	0	8.54±0.11	343.57±6.07	15.61±0.07	378.83±3.3
Ethyl acetoacetate	329.59±3.05	0	5.51±0.03	320.23±6.75	11.63±0.16	339.11±2.93
Ethyl propionylacetate	494.67±3.77	0	4.77±0.02	468.37±8.71	12.25±0.16	490.45±3.59
Methyl glycolate	470.83±2.03	0	2.02±0.04	470.32±7.44	5.29±0.25	519.58±4.46
Vinyl acetate	8.36±0.02	0.07±0.03	0.03±0.01	8.84±0.21	2.43±0.03	11.77±0.68
Vinyl propionate	3.36±0.22	0.786±0.03	155.4±0.38	3.14±0.62	91.34±2.21	100.81±1.46
Vinyl butyrate	11.12±0.16	0.342±0.089	177.76±0.16	11.83±0.24	5.21±0.19	23.91±0.45
Vinyl laurate	4.57±0.09	0	0.03±0.01	4.82±0.01	3.58±0.05	8.45±0.25
Vinyl benzoate	448.52±6.77	0.112±0.035	63.91±0.54	488.1±9.67	109.26±3.04	599.35±3.61
Vinyl crotonate	18.55±0.55	0.087±0.012	15.9±0.13	0.96	66.88±2.15	85.14±0.66
Vinyl acrylate	0.17±0.02	0	22.97±0.3	0.15±0.02	0	0.19±0.04
Geranyl acetate	55.38±2.74	0	3.53±0.09	59.43±	6.57±0.88	67.42±1.18
3-Methyl-3-buten-1-yl acetate	28.91±0.1	0	10.18±0.29	29.02±0.66	5.55±0.07	41.7±1.11
Methyl benzoate	3.43±0.4	0	0.2±0.03	3.29±0.05	1.14±0.15	4.54±0.21
Ethyl benzoate	3.43±0.18	0	0.16±0.02	3.43±0.05	0	3.91±0.26
n-Pentyl benzoate	39.29±0.76	0	1.45±0.36	36.21±0.79	4.67±0.62	41.07±0.56
Ethyl 2-chlorobenzoate	15.1±0.2	0	1.84±0.1	15.24±0.28	2.47±0.03	21.33±0.81

Methyl 2-hydroxybenzoate	0.3±0.21	0	0.2±0.01	0.41±0.1	0	0.38±0.01
Methyl 3-hydroxybenzoate	0.66±0.04	0	0.09±0.01	0.83±0.02	58.25±0.26	61.26±0.47
Propyl 4-hydroxybenzoate	2.47±0.08	0	0.28±0.03	2.45±0.03	0.08±0.04	2.59±0.52
Butyl 4-hydroxybenzoate	3.65±0.1	0	0.51±0.03	3.14±0.15	1.49±0.01	4.90±0.89
Phenyl acetate	1309.47±18.63	0.331±0.087	81.09±1.57	1312.1±15.37	775.74±3.15	2152.74±11.68
Phenyl propionate	2797.12±17.94	1.490±0.151	267±1.56	2715.09	820.44±5.48	3467.5±12.58
Phthalic acid diethyl ester	13.25±0.15	0	0.36±0.59	14.35±0.22	0	14.11±0.38
Isobutyl cinnamate	51.07±1.27	0	1.79±0.19	51.5±1.19	5.71±0.08	61.26±0.57
Methyl 2,5-dihydroxycinnamate	1.92±0.02	0	0.07±0.03	1.66±0.03	0	1.66±0.12
Methyl cinnamate	61.01±0.25	0	3.48±0.01	64.12	10.59±0.14	74.84±1.91
Methyl ferulate	0.86±0.01	0	0.07±0.01	0.86±0.01	0.6±0.08	1.38±0.12
1-Napthyl acetate	1015.98±13.08	0.143±0.046	7.61±0.88	1044.9±13.4	181.7±8.02	1272.19±11.22
1-Naphthyl butyrate	234.91±6.37	0.069±0.005	3.38±0.21	216.39±3.92	60.81±2.4	283.34±1.73
Benzyl 4-hydroxybenzoate	76.69±1.29	0	2.13±0.04	79.29±1.61	4.27±0.12	83.77±1.61
BFPME[2]	316.51±3.4	0.02±0.012	2.7±0.02	313.86±6.74	11.6±0.71	328.05±3.19
BHPP[2]	1276.92±13.68	0.073±0.024	8.78±1.01	1250.4±15.42	95.63±5.1	1368.09±7.77
Phenylethyl cinnamate	13.56±0.04	0	0.27±0.09	12.97±0.32	0.32±0.04	14.82±0.23
ϒ-Valerolactone	343.79±2.9	0	0.5±0.04	380.4±5.93	3.45±0.81	398.54±2.77
D-Pantolactone	9.08±0.03	0	0.9±0.04	8.89±0.19	0	8.89±0.26
Cyclohexyl butyrate	416.57±1.22	0.084±0.001	50.82±0.86	414.99±4.57	41.12±0.33	460.87±1.58
Glucose pentaacetate	794.08±4.04	0	0.45±0.01	816.64±9.2	6.08±0.15	916.02±4.43
(1*R*)-(-)-Menthyl acetate	1.03±0.1	0	0.18±0.01	0.96±0.01	0	0.97±0.07
(1*S*)-(+)-Menthyl acetate	0.59±0.14	0	0.02±0.01	0.62±0.08	0	0.61±0.06
Methyl (*R*)-(-)-mandelate	0.51±0.03	0	0.02±0.01	0.49±0.06	0	0.39±0.1
Methyl (*S*)-(+)-mandelate	1.51±0.24	0	0.49±0.01	1.48±0.12	0	1.54±0.42
(−)-Methyl (*R*)-3-hydroxyvalerate	9.05±0.26	0	0.45±0.01	9.88±0.12	0	9.75±0.53
(+)-Methyl (*S*)-3-hydroxyvalerate	6.26±0.01	0	0.25±0.06	6.67±0.13	0	6.95±0.4
(+)-Methyl (*S*)-3-hydroxybutyrate	9.84±0.24	0	0.2±0.02	9.81±0.17	0.69±0.24	12.1±0.32
(-)-Methyl (*R*)-3-hydroxybutyrate	5.41±0.22	0	0.3±0.03	5.34±0.08	1.03±0.14	7.05±0.38
E(*R*)CHB[2]	105.95±0.03	0	2.31±0.06	104.88±1.79	0	105.91±4.66
E(*S*)CHB[2]	108.7±1.9	0	4.72±0.16	106.14±2.23	80.69±4.05	187.83±2.78
(−)-Methyl (*R*)-Lactate	26.09±2.15	0	0.64±0.11	28.46±0.38	3.25±0.04	29.45±6.9
(+)-Methyl (*S*)-Lactate	62.72±0.56	0	2.39±0.01	68.57±0.11	4.54±0.06	78.39±2.61
(−)-Ethyl (*R*)-Lactate	24.38±0.58	0	0.7±0.01	27.32±0.48	2.45±0.08	36.12±2.75
(+)-Ethyl (*S*)-lactate	130.83±2.59	0	2.76±0.09	128.92±2.91	9.13±0.3	139.26±3.21
Methyl (2*R*)-2-phenylpropanoate	61.0±2.9	0	63.2±0.13	65.0±0.6	526.00±11.8	597.00±21
Methyl (2*S*)-2-phenylpropanoate	59.2±2.5	0	59.8±0.09	63.0±1.9	0	64.0±0.1

[1] Note: the number of esters found as potential substrates by Santiago et al. (2018) for $EH1_A$ was 72[1]. However, two of the esters found to not be hydrolyzed in the previous study were substituted by two new esters found to be converted, namely, methyl-2-bromobutyrate instead of triolein and methyl 2-phenyl propanoate instead of glyceryl tri-laurate. Additionally, ethyl acetate, ethyl benzoate and D-pantolactone, for which conversion was detected below the detection limit of the conditions used by Santiago et al. (2018)[1], were hydrolyzed in the conditions used in this study, which correspond to conditions at which substrate saturation exists.

[2] Abbreviation as follows: Benzoic acid, 4-formyl-, phenylmethyl ester [BFPME]; Benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate [BHPP]; Ethyl (*R*)-(+)-4-chloro-3-hydroxybutyrate [E(R)CHB]; Ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate [E(S)CHB]

Table S1 cont. K_m (mM) against a set of structurally different esters. The assays were performed at 30 °C and pH 8.0, as triplicates, with the average value and standard deviation shown. For K_m determinations the amount of protein was 0.2 µg, and substrate concentrations varied from 0 to 100 mM. Only the K_m (mM) for the $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ are shown.

	K_m (mM)[2]		
Ester[1]	**$EH1_{A1}$**	**$EH1_{B1}$**	**$EH1_{AB1}$**
Glyceryl triacetate	0.251±0.05	0.360±0.056	0.259±0.059
Glyceryl tripropionate	0.272±0.05	0.725±0.040	0.332±0.032
Glyceryl tributyrate	0.510±0.17	8.00±0.51	0.547±0.047
Methyl-2-bromobutyrate	5.53±0.03	-	9.20±0.082
Ethyl acetate	0.592±0.033	7.30±0.15	0.610±0.061
Propyl acetate	3.23±0.45	-	3.86±0.08
Butyl acetate	0.495±0.056	0.310±0.025	0.393±0.029
Hexyl acetate	0.405±0.069	1.41±0.04	1.82±0.08
Octyl acetate	1.05±0.10	0.620±0.075	1.30±0.03
Methyl butyrate	1.67±0.08	1.24±0.03	2.84±0.02
Methyl hexanoate	4.92±0.09	0.685±0.018	5.02±0.10
Methyl octanoate	0.930±0.110	0.225±0.037	1.05±0.10
Methyl decanoate	5.20±0.07	8.28±0.06	6.50±0.05
Ethyl propionate	7.03±0.08	5.55±0.06	6.76±0.07
Ethyl butyrate	8.86±0.22	9.28±0.09	9.70±0.07
Ethyl hexanoate	4.70±0.07	32.80±0.12	5.54±0.05
Ethyl octanoate	7.12±0.05	5.45±0.071	8.78±0.07
Ethyl decanoate	7.45±0.08	-	7.64±0.06
Ethyl dodecanoate	9.05±0.08	-	8.89±0.08
Propyl propionate	8.07±0.15	-	8.53±0.05
Propyl butyrate	5.52±0.05	-	5.19±0.12
Propyl hexanoate	0.512±0.057	1.16±0.04	0.747±0.045
Ethyl 2-ethylacetoacetate	0.420±0.031	0.493±0.150	0.850±0.058
Ethyl 2-methylacetoacetate	0.322±0.076	0.690±0.088	0.608±0.061
Ethyl 3-oxohexanoate	0.732±0.056	0.710±0.136	0.741±0.041
Ethyl acetoacetate	0.712±0.064	0.838±0.076	0.765±0.077
Ethyl propionylacetate	1.03±0.07	0.940±0.063	1.43±0.04
Methyl glycolate	0.602±0.145	0.740±0.075	0.658±0.046
Vinyl acetate	0.887±0.087	0.268±0.078	1.03±0.03
Vinyl propionate	0.927±0.098	0.475±0.082	0.725±0.073
Vinyl butyrate	0.835±0.192	0.215±0.045	0.948±0.048
Vinyl laurate	9.57±0.04	1.04±0.11	10.87±0.08
Vinyl benzoate	0.743±0.057	0.253±0.102	0.767±0.077
Vinyl crotonate	9.15±0.17	0.548±0.067	0.550±0.055
Vinyl acrylate	9.13±0.19	-	9.46±0.04
Geranyl acetate	0.592±0.056	0.290±0.040	0.480±0.048
3-Methyl-3-buten-1-yl acetate	0.582±0.065	0.490±0.092	0.953±0.050
Methyl benzoate	6.67±0.03	0.740±0.101	6.56±0.05
Ethyl benzoate	5.13±0.03	-	5.68±0.07
n-Pentyl benzoate	35.17±0.77	0.970±0.025	35.69±0.17
Ethyl 2-chlorobenzoate	16.00±0.04	1.38±0.05	16.72±0.07
Methyl 2-hydroxybenzoate	8.45±0.09	-	8.50±0.05

Methyl 3-hydroxybenzoate	12.33±0.56	0.180±0.020	0.19±0.01
Propyl 4-hydroxybenzoate	12.85±0.54	0.783±0.079	10.59±0.05
Butyl 4-hydroxybenzoate	15.85±0.04	0.958±0.080	16.55±0.05
Phenyl acetate	0.08±0.01	1.05±0.02	0.445±0.045
Phenyl propionate	0.965±0.03	0.703±0.122	0.998±0.100
Phthalic acid diethyl ester	1.12±0.08	-	1.06±0.06
Isobutyl cinnamate	2.52±0.05	1.04±0.10	2.70±0.071
Methyl 2.5-dihydroxycinnamate	4.50±0.062	-	4.63±0.06
Methyl cinnamate	3.241±0.03	1.41±0.06	4.23±0.02
Methyl ferulate	35.37±0.05	5.33±0.31	40.19±0.20
1-Napthyl acetate	0.422±0.050	0.775±0.053	0.530±0.013
1-Naphthyl butyrate	0.307±0.044	0.520±0.093	0.350±0.015
Benzyl 4-hydroxybenzoate	0.492±0.062	0.573±0.027	0.497±0.020
BFPME[1]	1.19±0.09	0.570±0.053	1.17±0.02
BHPP[1]	0.532±0.063	0.503±0.031	0.560±0.026
Phenylethyl cinnamate	17.32±0.42	38.30±3.93	20.21±0.02
ϒ-Valerolactone	1.02±0.04	8.00±0.245	1.15±0.01
D-Pantolactone	9.15±0.03	-	9.17±0.02
Cyclohexyl butyrate	1.21±0.03	0.497±0.096	1.20±0.02
Glucose pentaacetate	0.55±0.03	21.0±0.092	0.600±0.010
(1*R*)-(-)-Menthyl acetate	6.55±0.03	-	6.38±0.04
(1*S*)-(+)-Menthyl acetate	6.75±0.03	-	6.24±0.02
Methyl (*R*)-(-)-mandelate	5.22±0.07	-	5.19±0.01
Methyl (*S*)-(+)-mandelate	6.60±0.49	-	6.15±0.01
(−)-Methyl (*R*)-3-hydroxyvalerate	0.625±0.082	-	0.593±0.049
(+)-Methyl (*S*)-3-hydroxyvalerate	0.463±0.069	-	0.418±0.119
(+)-Methyl (*S*)-3-hydroxybutyrate	0.714±0.066	4.38±0.153	4.93±0.09
(-)-Methyl (*R*)-3-hydroxybutyrate	0.930±0.088	5.03±0.014	5.54±0.06
E(*R*)CHB[2]	4.71±0.063	-	4.59±0.05
E(*S*)CHB[2]	7.50±0.086	0.610±0.039	0.758±0.076
(−)-Methyl (*R*)-Lactate	2.25±0.089	0.620±0.075	0.637±0.064
(+)-Methyl (*S*)-Lactate	0.798±0.180	0.353±0.05	0.680±0.068
(−)-Ethyl (*R*)-Lactate	0.637±0.066	0.453±0.018	0.747±0.075
(+)-Ethyl (S)-lactate	0.690±0.049	0.473±0.037	1.62±0.16
Methyl (2*R*)-2-phenylpropanoate	21.10±1.80	0.33±0.02	0.44±0.03
Methyl (2*S*)-2-phenylpropanoate	22.40±3.20	-	27.10±2.20

[1]Abbreviation as follows: Benzoic acid, 4-formyl-, phenylmethyl ester [BFPME]; Benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate [BHPP]; Ethyl (*R*)-(+)-4-chloro-3-hydroxybutyrate [E(R)CHB]; Ethyl (*S*)-(−)-4-chloro-3-hydroxybutyrate [E(S)CHB].

[2]Fits for K_m determinations (for each ester and enzyme variant) are provided in separate Supplementary files (file named K_M determinations fits). The figures represent the normalized/relative µmol/minmL, calculated as described in Supporting Methods, for each of the ester concentrations. Measurements were in triplicates with average value represented. The figure was created with the R language console with Michaelis-Menten fit. K_m values were calculated by simple Michaelis-Menten kinetics as described[1].

Table S2. Kinetic parameters of $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$, for two substrates preferentially hydrolyzed by $EH1_{B1}$. The assays were performed at 30 °C and pH 8.0, as triplicates, with the average value and standard deviation shown. K_m values were calculated by simple Michaelis-Menten kinetics as described[1]. Fits for K_m determinations are shown below the table. The figures represent the µmol/minmL (in triplicates with average value represented), calculated as described in Supporting Methods, for each of the ester concentrations.

Substrates	Methyl 3-hydroxybenzoate	
	K_m (mM)	k_{cat} (min^{-1})
$EH1_{A1}$	12.33±0.56	0.83±0.02
$EH1_{B1}$	0.18±0.02	58.3±0.3
$EH1_{AB1}$	0.19±0.01	61.26±0.47
	Methyl (2*R*)-2-phenylpropanoate	
$EH1_{A1}$	21.1±1.8	65.0±0.6
$EH1_{B1}$	0.33±0.02	526.0±11.8
$EH1_{AB1}$	0.44±0.03	597.0±21.0
	Methyl (2*S*)-2-phenylpropanoate	
$EH1_{A1}$	22.4±3.2	63.0±1.9
$EH1_{B1}$	-	-
$EH1_{AB1}$	27.1±2.2	64.0±1.6

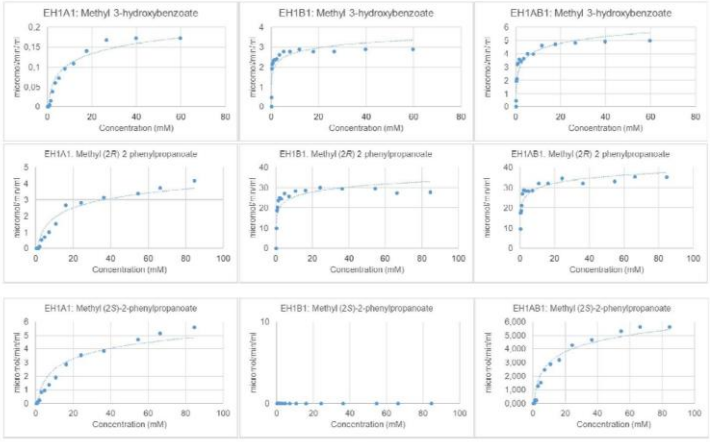


Table S3. Substrate spectra of EH102 variants. The chemically and structurally distinct esters out 96 tested[2], for which activity was detected, are listed on the left side of the table. The V_{max} for each substrate when tested at pH 8.0, 30ºC and 50 mM ester is shown according to each variant, including the corresponding standard deviation (triplicates).

	V_{max} (U g^{-1})		
Substrates[1]	**$EH102_A$**	**$EH102_B$**	**$EH102_{AB}$**
1-Naphthyl acetate	951.8±11.9	917.4±13.0	1832.3±45.3

1-Naphthyl butyrate	688.3±6.6	132.8±5.0	846.9±7.5
Glyceryl tri-acetate	784.7±3.3	92.0±3.7	884.6±10.5
Glyceryl tripropionate	896.2±13.0	2120.5±9.2	3089.4±59
Glyceryl tributyrate	0	346.1±9.4	350.3±2.5
Hexyl acetate	225.2±5.1	0	224.7±12.0
Octyl acetate	145.1±22.1	0	141.6±5.3
Methyl butyrate	193.7±1.2	0	200.0±1.9
Propyl hexanoate	0	1878.9±7.8	1885.1±1.5
Vinyl propionate	783.7±2.5	0	794.9±4.3
Vinyl benzoate	0	859.1±9.4	855.1±4.3
Geranyl acetate	168.0±1.6	0	167.2±2.8
3-Methyl-3-buten-1-yl-acetate	256.4±7.0	417.2±11.7	661.6±3.7
Ethyl 3-oxohexanoate	0	1430.8±12.6	1470.4±2.7
Ethyl propionylacetate	0	7.68±1.11	7.58±0.55
Methyl glycolate	565.6±3.0	0	560.0±9.1
Propyl acetate	621.5±6.2	69.0±2.5	707.8±4.9
Butyl acetate	656.9±3.9	11.4±1.8	672.4±8.5
Phenyl acetate	1001.5±8.4	4147.3±21.2	5171.8±1.5
Phenyl propionate	544.0±2.7	5726.5±71.2	6282.8±5.8
Glucose pentaacetate	1038.3±11.2	0	1073.6±2.9
(−)-Ethyl (*S*)-lactate	0	1227.7±10.8	1236.7±0.6
Ethyl (S)-(−)-4-chloro-3-hydroxybutyrate [E(S)CHB]		6.64±1.49	6.41±0.12
Benzyl (R)-(+)-2-hydroxy-3-phenylpropionate [BHPP]	0	2147.8±5.7	2193.3±6.4
Benzoic acid, 4-formyl-, phenylmethyl ester [BFPME]	0	6.58±0.38	6.54±0.84
Phthalic acid diethyl ester	0	1.44±0.70	1.52±0.05

[1]Note: the number of esters found as potential substrates by Martínez-Martínez et al. (2018) for $EH102_A$ was 11[2]. Additionally, hexyl acetate, octyl acetate, vinyl propionate, 3-methyl-3-buten-1-yl-acetate, methyl glycolate, for which conversion was detected below the detection limit of the conditions used by Martínez-Martínez et al. (2018)[2], were hydrolyzed in the conditions used in this study, which correspond to conditions at which substrate saturation exists.

Table S4 Purification summary for $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ proteins.

Sample	Variant	Total protein (mg)[c]	Total activity (units)[d]	Specific activity (U/mg)[d]	Purity (%)[e]
Crude lysate[a]	$EH1_{A1}$	1467	1763	1.20	-
Purified protein[b]	$EH1_{A1}$	25	1642	65.68	>99
Crude lysate[a]	$EH1_{B1}$	1143	138	0.12	-
Purified protein[b]	$EH1_{B1}$	11	127	11.55	>99
Crude lysate[a]	$EH1_{AB1}$	1360	2084	1.53	-
Purified protein[b]	$EH1_{AB1}$	24	1875	78.13	>99

[a] After Ni-NTA His-Bind resin and gel filtration.
[b] From 4 g of wet weight *E. coli* cell pellet (from 2000 ml of bacterial culture).
[c] Protein concentration determined by Bradford assay using BSA as a standard protein.
[d] Activity determined using glyceryl tripropionate as substrate[1,2].
[e] Calculated by densitometry in the SDS-PAGE.

Table S5 Purification summary for $EH102_A$, $EH102_B$ and $EH102_{AB}$ proteins.

Sample	Variant	Total protein (mg)[c]	Total activity (units)[d]	Specific activity (U/g)[d]	Purity (%)[e]
Crude lysate[a]	$EH102_A$	322	59330	184.3	-

Purified protein[b]	$EH102_A$	59	52860	895.9	>99
Crude lysate[a]	$EH102_B$	317	760	2.4	-
Purified protein[b]	$EH102_B$	41	490	12.0	>99
Crude lysate[a]	$EH102_{AB}$	454	93300	205.6	-
Purified protein[b]	$EH102_{AB}$	91	88300	970.3	>99

563 [a] After Ni-NTA His-Bind resin and gel filtration.
564 [b] From 4 g of wet weight *E. coli* cell pellet (from 2000 ml of bacterial culture).
565 [c] Protein concentration determined by Bradford assay using BSA as a standard protein.
566 [d] Activity determined using glyceryl tripropionate as substrate[1,2].
567 [e] Calculated by densitometry in the SDS-PAGE.
568

569 **Table S6.** Crystallographic statistics of $EH1_{AB1}$.

570

Values in parentheses are for the high resolution shell		
Crystal data	$EH1_{AB1}$	$EH1_{AB1}$-complex
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters		
a (Å)	82.23	85.25
b (Å)	86.04	88.16
c (Å)	98.17	98.50
Data collection		
Beamline	ESRF-ID30A-1	XALOC (ALBA)
Temperature (K)	100	100
Wavelength (Å)	0.966000	0.979240
Resolution (Å)	43.02-2.11	49.25-3.20
Data processing		
Total reflections	237360 (19288)	49457 (8489)
Unique reflections	40731 (3301)	12184 (2184)
Multiplicity	5.8 (5.8)	4.1 (3.9)
Completeness (%)	99.9 (100.0)	96.5 (96.2)
Mean I/σ (I)	10.6 (3.1)	4.5 (2.2)
R_{merge}[†] (%)	12.4 (64.1)	21.8 (52.5)
R_{pim}[††] (%)	5.7 (29.6)	11.9 (29.8)
Molecules per ASU	2	2
Refinement		
R_{work} / R_{free}[†††] (%)	16.0/19.8	22.7/27.6
Nº of atoms/average B	5245/30.81	4891/41.65
Macromolecule	4770/29.71	4747/41.32
Ligands	144/49.89	103/68.06
Solvent	331/38.26	41/13.40
Ramachandran plot (%)		
Favoured	94.3	93.9
Outliers	0.5	0.3
RMS deviations		
Bonds (Å)	0.007	0.005
Angles (°)	1.192	1.057
PDB accession codes	6I8F	6I8D

571 [†] $R_{merge} = \sum_{hkl} \sum_i | I_i(hkl) - [I(hkl)] | / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement

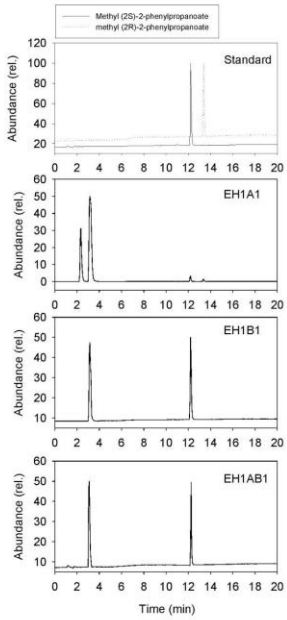
of reflection hkl and [I(hkl)] is the weighted mean of all measurements.

$^{\dagger\dagger}R_{pim} = \sum_{hkl} [1/(N - 1)]\ 1/2 \sum_i | I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$, where N is the redundancy for the hkl reflection.

$^{\dagger\dagger\dagger}R_{work} / R_{free} = \sum_{hkl} | Fo - Fc | / \sum_{hkl} | Fo |$, where Fc is the calculated and Fo is the observed structure factor amplitude of reflection hkl for the working / free (5%) set, respectively.

GC Chromatograms

The GC chromatograms for $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$ chiral determination using racemic mixture of methyl (2*R*)-2-phenylpropanoate and methyl (2*S*)-2-phenylpropanoate are shown below:



GC chromatograms. The figures represent the abundance of methyl (2*R*)-2-phenylpropanoate and methyl (2*S*)-2-phenylpropanoate (standards shown in the upper panel) and their reaction products in the presence of $EH1_{A1}$, $EH1_{B1}$ and $EH1_{AB1}$. Reactions were performed as detailed in Experimental section, and reaction products analyzed by GC. Retention time for the (*R*)- and (*S*)-acids 2.34 and 3.14 min, respectively.

Chapter 4: Discussion

In **Chapter 3: Results** a very accurate and complete description of different enzyme-related studies has been presented, with both computational and experimental work. In an attempt to facilitate comprehension, the work has been summarized in the current chapter.

4.1. Determinants and Prediction of Esterase Substrate Promiscuity Patterns

Promiscuity is an enzymatic characteristic which defines the substrate scope of a biocatalyst and it has a potential industrial interest. The prediction of substrate promiscuity from sequence and structural data remains to be elucidated, existing only experimental alternatives which are expensive in terms of costs and resources and time-consuming. To address this gap, we have used the substrate specificity of a set of 145 diverse microbial esterases against a library of 96 substrates.

To rationalize enzymatic substrate ranges, we performed ligand migration studies with PELE (Borrelli et al., 2005) in order to study the behavior of two (crystallized) esterases with very different substrate promiscuity values (5JD4: 77 substrates and 5JD3: 10 substrates), present in the data set and isolated from the same environment. Substrate migration simulations with PELE (**Fig. 4.1**) reveals two major facts: a significantly better binding profile and a reduction in the solvent accessible surface area (SASA) for 5JD4. These differences in enzyme-ligand interaction can be explained by the catalytic triad structural environment: 5JD4 has a wide buried active site, while, 5JD3 has a surface-exposed catalytic triad.

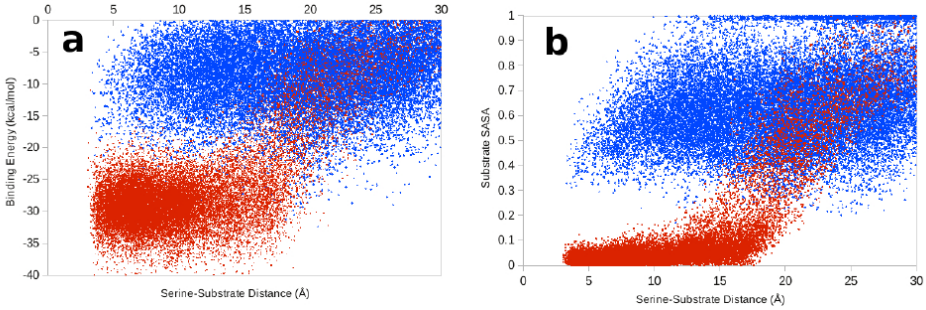


Figure 4.1: Substrate migration with PELE for 5JD4 (red) and 5JD3 (blue). Panel A plots binding energy vs. serine-substrate distance and panel B substrate SASA vs. serine-substrate distance.

These results and the previous crystal inspection drive us to the following first assumptions about promiscuity:

1) Promiscuous enzymes require "large" buried hidden cavities.
2) Solvent exposure of key residues leads to low promiscuity values.

The previous two assumptions can be translated into two major concepts, cavity area and catalytic residues SASA. In order to look for a correlation between these two values and enzymatic promiscuity, we started evaluating the crystal structures present in the set (10 crystals). Cavity volume was calculated using Fpocket (Le Guilloux, Schmidtke, & Tuffery, 2009) and catalytic residues SASA with GetArea web-server (Fraczkiewicz & Braun, 1998). Importantly, with the ratio of these two values (cavity volume/SASA) it is possible to classify between promiscuous and non-promiscuous enzymes with high accuracy (100 %): when this value (called Effective Volume) is higher than 62.5 $Å^3$ the enzyme hydrolyze 20 or more different substrates, considering the enzyme promiscuous (**Fig. 4.2**).

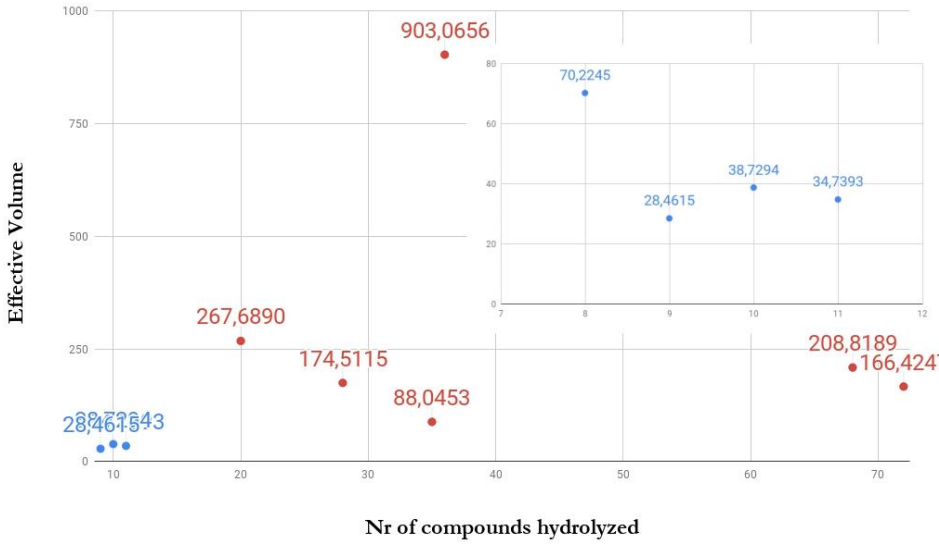


Figure 4.2: Cavity volume divided SASA ratio for dataset crystals (promiscuous in red non-promiscuous in blue). The inner plot is a zoom of the non-promiscuous region.

With this information, the analysis was extended to the remaining 135 esterases without a crystal structure. Homology modeling was used for structure prediction, followed by cavity volume and SASA calculation, and a general workflow was established (**Fig. 4.3**). Again the combination of both parameters classifies promiscuous enzymes with high accuracy (96%) (**Fig. 4.4**).

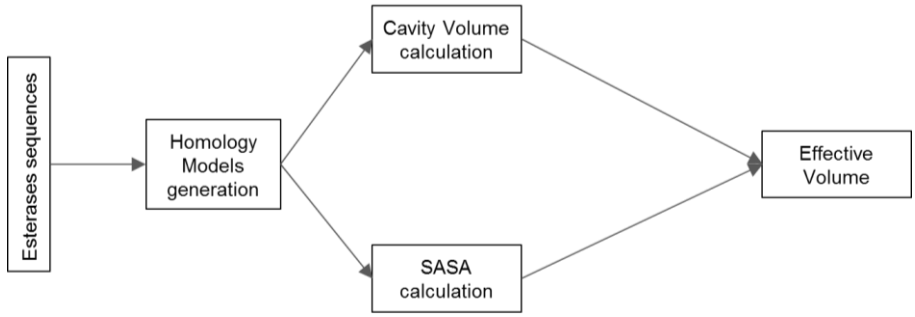


Figure 4.3: Effective Volume calculation workflow.

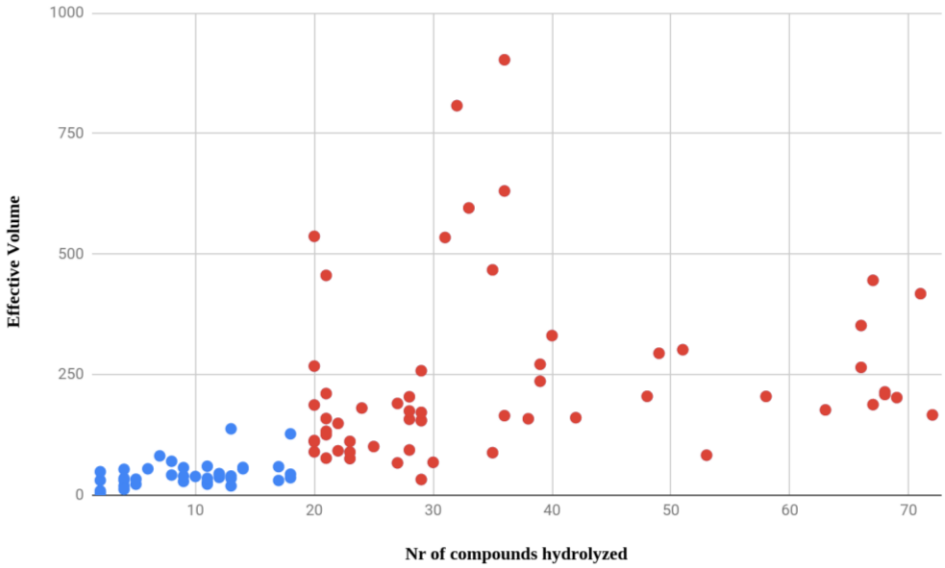


Figure 4.4: Effective volume relation with the number of compounds hydrolyzed for all dataset esterases (promiscuous in red non-promiscuous in blue).

The analysis was extended to another enzymatic family, haloacid dehalogenase phosphatases (HAD-phosphatases) C2 cap members, with similar success ratios (**Fig. 4.5**). Although a threshold could not be unambiguously established, sequences with the top 10 effective volumes belong to moderate-high to high promiscuity enzymes.

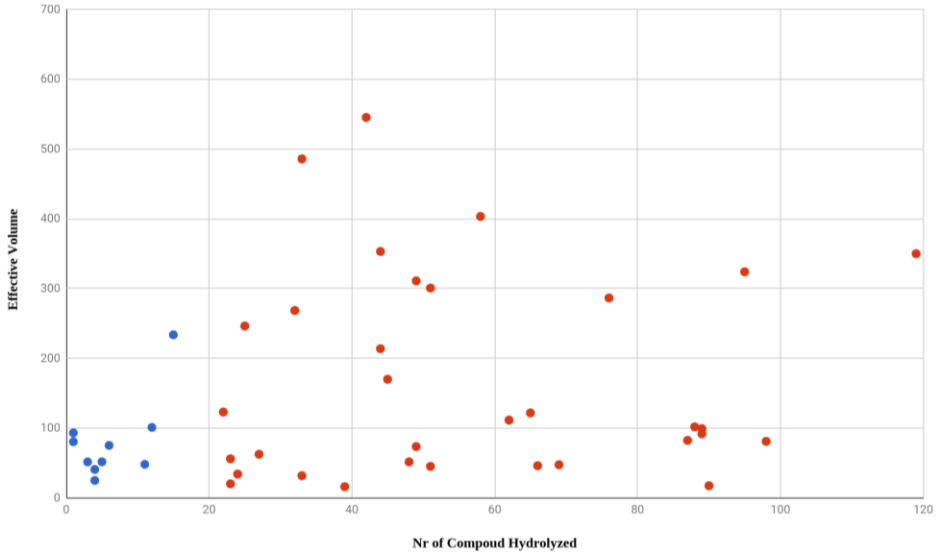


Figure 4.5: Effective volume for HAD-phosphatases dataset (promiscuous in red non-promiscuous in blue).

4.2. Computer-Aided Laccase Engineering: Toward Biological Oxidation of Arylamines

In Chapter 3.2, we introduced the rational design of laccase for arylamines oxidation. Those compounds, like aniline, have many industrial applications (paints, isolating layers, etc.) and the oxidation relies on harmful chemical compounds. Laccases are multicopper enzymes capable to oxidize a broad variety of molecules (including aniline) with soft-acids as the buffer, but the production of conductive polyaniline (PANI) remains a major issue for those enzymes.

The production of conductive polyaniline requires the presence of strong oxidation agents and a very low pH (around 3), where aniline monomers are protonated (**Fig. 4.6**). Laccases have activity in acidic pH, but the tests with wild-type enzymes (like PM1L) and chimeric variants (like 7D5) hold a really

low activity over aniline, making the biocatalysis not a suitable option for industrial purposes.

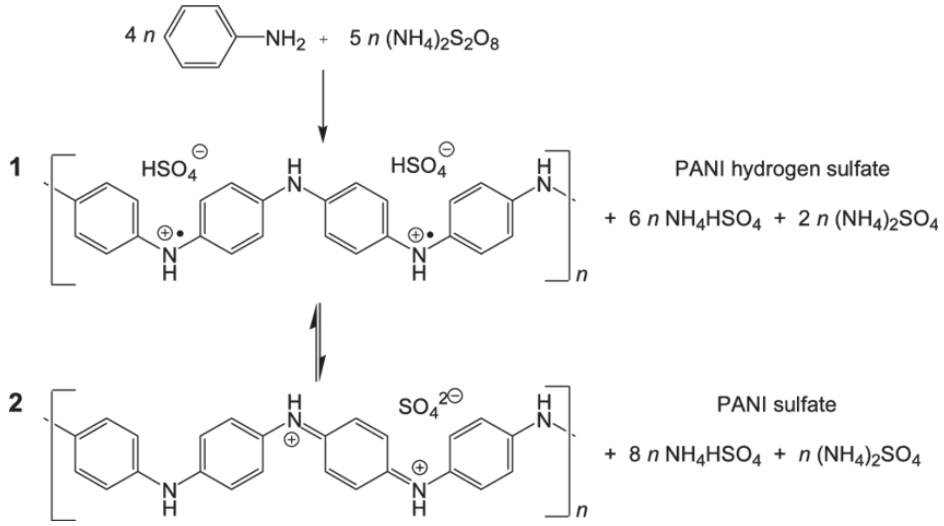


Figure 4.6: Chemical conductive polyaniline production (Mentus, Cirić-Marjanović, Trchová, & Stejskal, 2009).

Combining classical mechanics, PELE, and quantum methods (QM/MM), we developed a new 7D5 variant doubling PANI production. At pH 3 aniline monomers are protonated, thus the ligand diffusion in PM1L and 7D5 significantly difficult due to the presence of the T1-copper, also positive charge. Although, the introduction of two mutations (N236D and N207S) modifies the binding site electrostatic environment to the ligand special needs, favoring the electron transfer of the substrate. This was also translated into improved interactions energies, reduced donor-acceptor distance and SASA (**Fig. 4.7**). QM/MM calculations (Qsite (Murphy, Philipp, & Friesner, 2000; Philipp & Friesner, 1999)) revealed improved spin densities for the mutant variant compared to 7D5 and PM1L (**Table 4.1**).

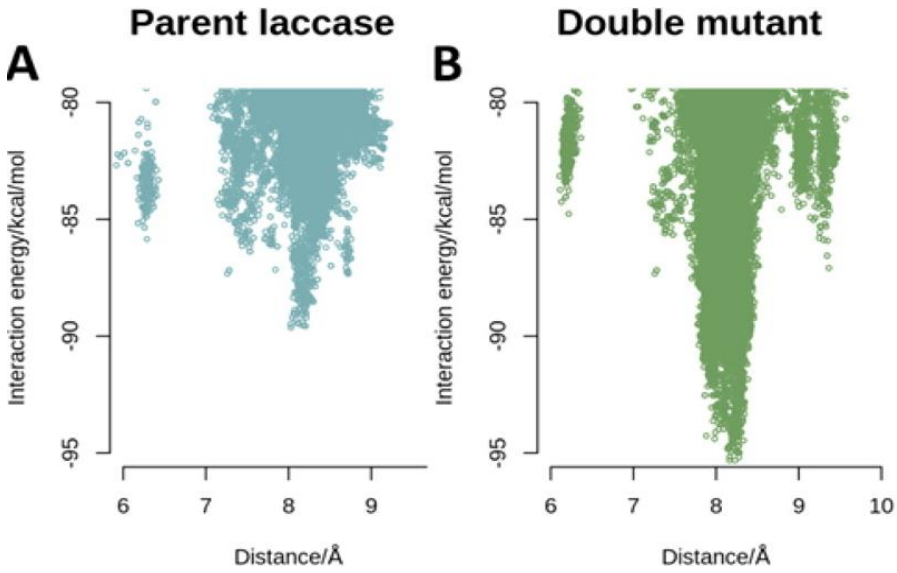


Figure 4.7: Binding energy profiles with aniline monomer for 7D5 (panel A) and the double mutant (DM) (panel B).

Table 4.1: Spin density and kinetic values for arylamines oxidation for 7D5 and DM.

System	SP (%)	K_{cat} (s^{-1})	K_m (mM)
7D5 + ANL	12	10.1	28
DM + ANL	26	22.6	59.3
7D5 + DMPD	39	459	1.7
DM + DMPD	68	741	1.2
7D5 + PPD	32	14.7	3.7
DM + PPD	46	20.8	3.5
7D5 + ABTS	23	291	0.0042
DM + ABTS	40	570	0.01

Experimental validation was successful, the new mutant had an enhanced activity over aniline and other arylamines, like Dimethyl-4-phenylenediamine (DPMD) or p-Phenylenediamine (PPD), and over a classical laccase redox mediators like 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS), following “in silico” cross-validation of these molecules lead to improved spin densities, as was previously observed for aniline monomers (**Table 4.1**). The computational hypothesis is that a better ligand placement in the binding site induced by the electrostatic environment modification caused by the introduced mutations is leading to the enhanced activity. As observed in **Figure 4.8**, the aniline monomer placement towards the T1-copper is far improved in DM variant. Moreover, the improvement of aniline oxidation leads to two times increasing PANI production (**Fig. 4.9**).

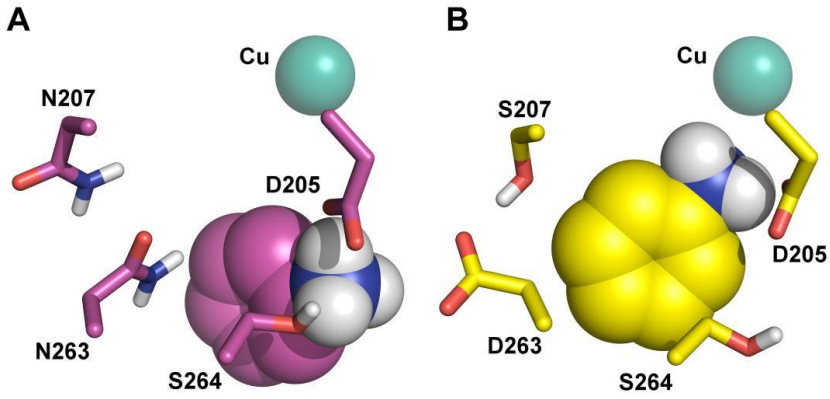


Figure 4.8: Aniline monomer placement in 7D5 (panel A) and DM (panel B) active site.

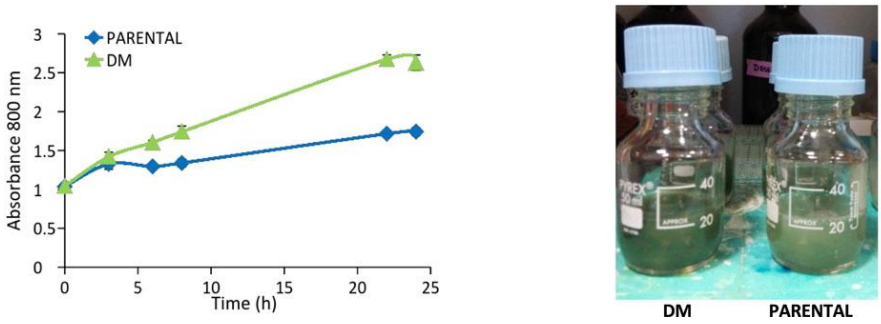


Figure 4.9: 7D5 and DM polyaniline production comparison.

4.3. Rational Engineering of Multiple Active Sites in an Ester Hydrolase

In previous chapters, we have discussed the advantages of computational modeling in enzyme engineering. After our laccase design, we challenge our methods with more complex ideas. Previous investigations in "de novo" design have proved possible the creation of artificial active sites in non-catalytic protein scaffolds (Kiss et al., 2013). Instead of this already tested approach, in this project, a disruptive idea (with few similar successful attempts, where incomplete enzyme family motif were functionalized through rational design (Shu et al., 2016)) was developed: the introduction of an artificial active site inside a currently functional enzyme to enhanced catalytic performance due to two fully functional active sites (**Fig. 4.10**).

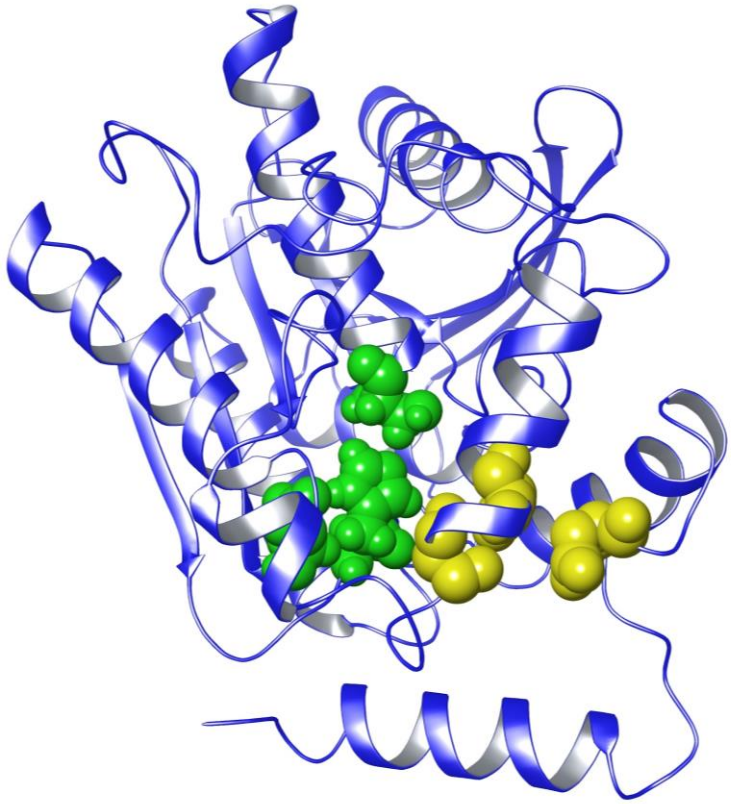


Figure 4.10: Two active site concept from a structural point of view.

As mentioned in **Chapter 1**, serine-hydrolases are extensively used enzymes for different industrial purposes. The catalysis is based in three amino acids, the catalytic triad, and requires water as the only cofactor. We hypothesized that this kind of catalytic site could be replicated in other protein areas, producing an artificial active site inside the existent enzymatic scaffold.

In previous research (**Chapter 3.1**), we worked with a huge variety of esterases and from this dataset, we selected one with a solved crystal structure and industrial potential, following these two criteria EH1 was selected (PDB code: 5JD4). EH1 is a high activity serine-esterase with a broad substrate scope.

By applying PELE we explored the complete protein surface space looking for non-catalytic binding regions where a general esterase ligand presents a good interaction profile. Applying general modeling software of visualization, these regions can be functionalized and turned into proper active sites (**Fig. 4.11**) introducing two mutations (E25D and L214H).

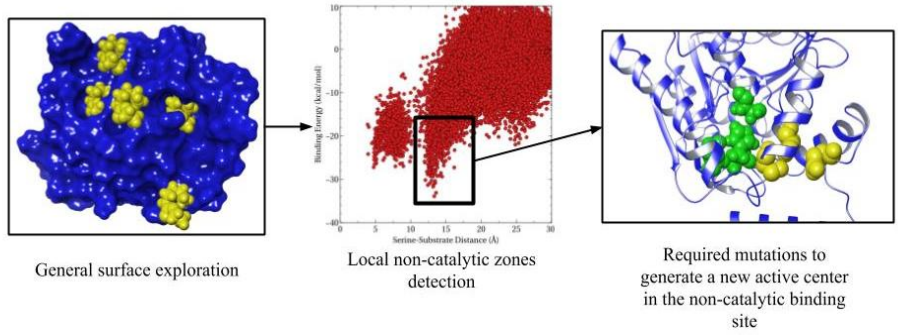


Figure 4.11: General workflow to generate PluriZymes.

Further substrate migration reevaluation, proved this new active site capable of holding catalytic positions (**Fig 4.12**) mimicking the binding modes present in wild-type serine-hydrolases. Finally, the experimental validation has shown activity over 24 different esters in the engineered active site.

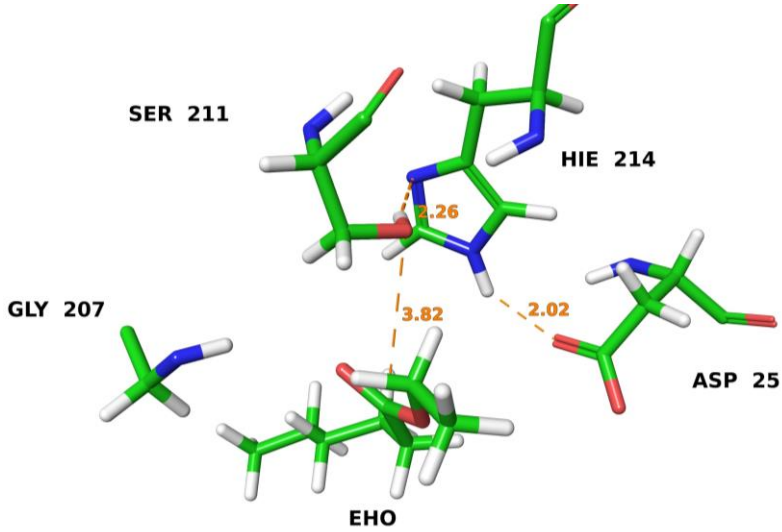


Figure 4.12: Artificial active site interaction with an ester molecule.

Table 4.2: Kinetic values for the native EH1 and the engineered forms, with only the artificial active site active (**$EH1_B$**) or both, activated (**$EH1_{AB}$**).

Enzyme variant	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($s^{-1}M^{-1}$)
$EH1_A$ (for *p*NPP)	146 ± 27	3853 ± 23	4.4×10^5
$EH1_B$ (for *p*NPP)	267± 20	$263 \pm 16 \times 10^{-3}$	16.4
$EH1_{AB}$ (for *p*NPP)	485 ± 21	211 ± 8	7250
$EH1_A$ (for phenylpropionate)	964 ± 24	2797± 98	4.8×10^4
$EH1_B$ (for phenylpropionate)	752± 20	1.49 ± 0.93	33.0
$EH1_{AB}$ (for phenylpropionate)	942 ± 17	267 ± 15	4720

Despite the success, the overall catalytic activity was really low (**Table 4.2**) and with a negative effect when both sites were working at the same time. Probably, this undesired effect is due to the proximity of both active sites.

4.4. PluriZyme: catalytic advantages of having two active sites in an enzyme

In the previous sub-chapter 4.3, we have discussed the design of an enzyme with multiple active sites and, despite the initial success, the final enzyme didn't introduce any catalytic advantage. Thus, new strategies were designed in order to improve the performance of the engineered active site:

1) Improving the first design (**Chapters 3.3 and 4.3**) by computer-aided enzyme engineering.
2) Moving the engineered active site as far as possible of the original one in another esterase model.

Looking into the previous design molecular dynamics output, a potential mutation spot was identified. Arginine 23 (**see Fig. 4.13A**) is really close to the engineered catalytic triad, reducing the effective volume and kidnapping the acid residue. The position was mutated for all possible short amino acids and the resulting enzyme-ligand interaction evaluated.

Mutation R23G was the most promising because it produced the best binding energies, opening the available reaction space, increasing substrate promiscuity (**Chapter 3.1**) and avoiding all interaction between residues 23 and 25 (**Fig. 4.13**). Further, 250 ns molecular dynamics revealed an improvement of distances between the three catalytic residues (**Table 4.3**) caused by the introduced R23G mutation.

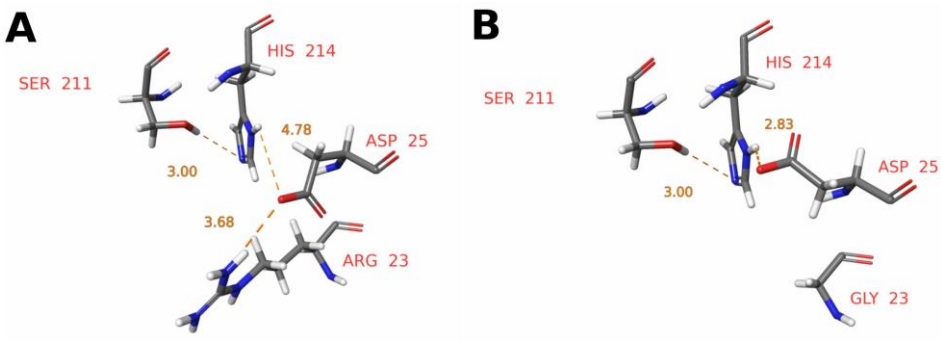


Figure 4.13: Two representative snapshots of EH1B (**A**) and EH1B1 (**B**) catalytic triads from the molecular dynamics simulations, where we clearly observe the movement in Asp 25 as a result of the R23G mutation.

Table 4.3: The lower panel indicates the average distances between key catalytic amino acids obtained for both species along 250 ns molecular dynamics for the original PluriZyme (AB) or the one including R23G mutation (AB1).

	$EH1_{AB}$	$EH1_{AB1}$
Ser 211-His 214 Distance (Å)	5.2 ± 1.8	3.4 ± 1.5
Ser 211-His 214 Distance (Å)	8.4 ± 1.6	5.5 ± 1.4

In silico results were experimentally tested, the new enzyme variant was far more efficient than the wild-type, enhancing the activity of the artificial catalytic triad 5000-fold, maintaining wild-type substrate scope, and gaining additivity when both active sites are working at the same time. Furthermore, the obtention of the crystal structure, with and without ligand (PDB codes: 6I8D, 6I8F), proved the performance of the designed enzyme.

Moreover, the design was extended to another ester hydrolase, EH102 (PDB code: 5JD3), proving the reproducibility of the multiple active site addition. To avoid competition between active sites, the engineered one was placed as far as possible of the original. Following the same protocol applied in Chapter 3.3, a non-catalytic binding site was identified, activated introducing two mutations (the required histidine was already there) and retested aiming for catalytic positions in the engineered place (**Fig. 4.14**).

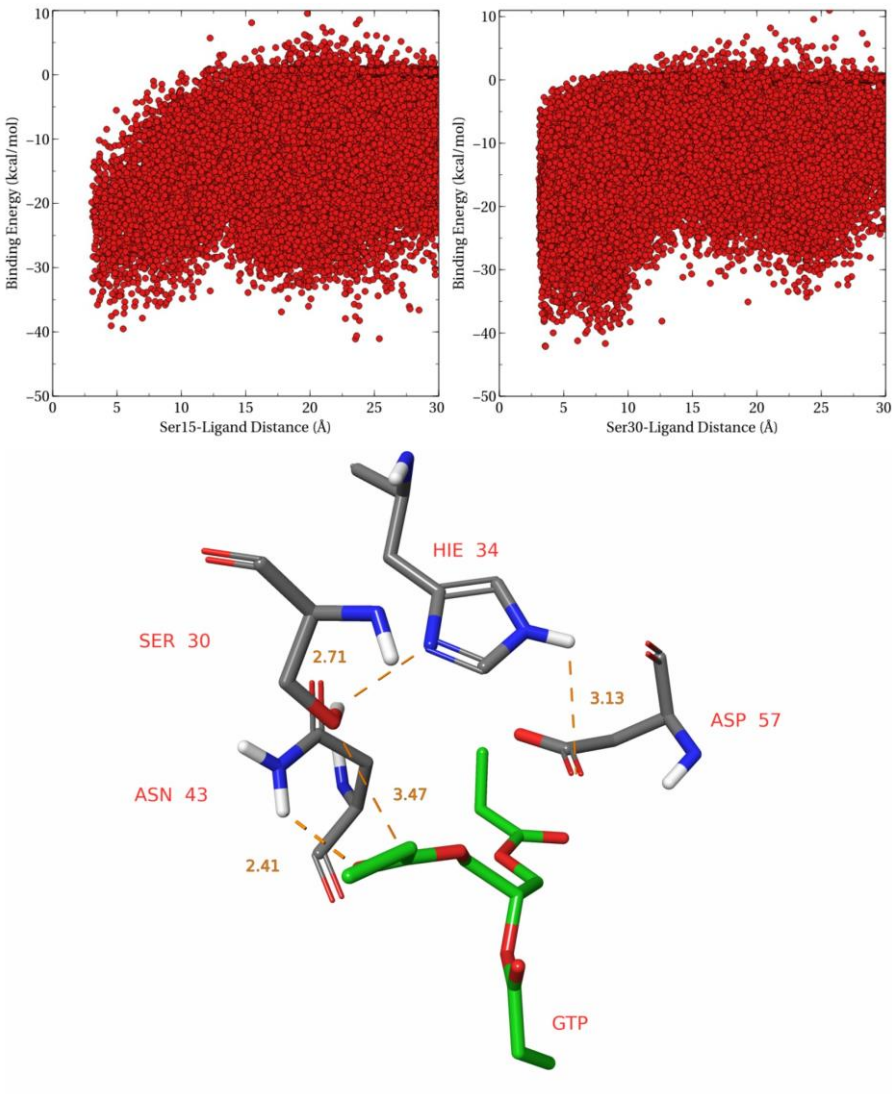


Figure 4.13: EH102 PluriZyme computational design steps. First (top left) looking for a non-catalytic active site and then create (down) and evaluate (top right) the artificial catalytic triad.

The following experimental validation showed activity for the engineered active site over 19 substrates, expanding total enzyme substrate scope up to 26 different esters, obtaining activity over 7 new substrates. Moreover, enhancing catalytic performance through additivity when both active sites are working at the same

time was also obtained in this ester hydrolase. As an interesting effect, the new enzyme gains S-enantioselectivity, exclusively hydrolyzing ethyl-S-lactate.

Chapter 5: Conclusions

In conclusion, this thesis has been driven to the fulfilment of previously presented aims (**Chapter 2**). Applying state-of-the-art computational modeling methods, our results successfully represent the importance of "in silico" research in enzyme engineering efforts and enzymology studies in general. The increasing computational power available and the effectiveness of the provided results make computer-aided enzyme design a promising tool for industrial applications.

Besides, the main conclusions of this thesis are:

1) **Elucidate the basic principles of enzymatic promiscuity in esterases**: The study of esterases structural parameters has successfully lead to the **effective volume** definition. This new structural parameter results from the combination of active site cavity volume and the solvent accessible surface area (SASA) of the catalytic residues. The effective volume allows a very accurate (> 96%) substrate promiscuity separation in serine hydrolases.

   Furthermore, the calculation has been extended to another enzymatic family, HAD-phosphatases, with similar qualitative results to separate between promiscuous and non-promiscuous enzymes.

2) **Computer-aided rational enzyme design**: we "in silico" engineered a laccase to obtain an enhanced mutant variant. The resulting tailored enzyme was "in vitro" tested and validated. The laccase variant obtained was more active for different types of arylamines and reaction mediators, specifically a two fold enhanced production of conductive polyaniline (a high-value polymer of costly and harmful production) was observed.

3) **Designing of novel enzymes**: During the thesis, we successfully developed, for the first time, an esterase **PluriZyme**. These new unique enzymes hold two fully functional catalytic triads. A computational protocol was set up and tested with an esterase. The resulting variant was “in vitro” tested and lead to a functional PluriZyme, with two working catalytic triads.

   After the first attempt, the original PluriZyme was improved with computer-aided rational design. A single point mutation was “in silico” tested and the resulting variant was really promising. The final PluriZyme appears to be additive (enhancing catalytic performance several thousands of times), enantioselective and with broader optimal working temperature. Moreover, the concept was reproduced in another esterase with very promising results.

References

Akoh, C. C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H., & Shaw, J.-F. (2004). GDSL family of serine esterases/lipases. *Progress in Lipid Research*, *43*(6), 534–552.

Alcalde, M. (2015). Engineering the ligninolytic enzyme consortium. *Trends in Biotechnology*, *33*(3), 155–162.

Aranda, C., Municoy, M., Guallar, V., Kiebist, J., Scheibner, K., Ullrich, R., … Gutiérrez, A. (2019). Selective synthesis of 4-hydroxyisophorone and 4-ketoisophorone by fungal peroxygenases. *Catalysis Science & Technology*. https://doi.org/10.1039/c8cy02114g

Aranda, J., N M F S, Fernandes, P. A., Roca, M., Tuñon, I., & Ramos, M. J. (2014). The Catalytic Mechanism of Carboxylesterases: A Computational Study. *Biochemistry*, *53*(36), 5820–5829.

Armendáriz-Ruiz, M., Rodríguez-González, J. A., Camacho-Ruíz, R. M., & Mateos-Díaz, J. C. (2018). Carbohydrate Esterases: An Overview. *Methods in Molecular Biology* , *1835*, 39–68.

Arnold, F. H. (2018). Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie* , *57*(16), 4143–4148.

Arnold, F. H., & Volkov, A. A. (1999). Directed evolution of biocatalysts. *Current Opinion in Chemical Biology*, *3*(1), 54–59.

Asther, M., Haon, M., Roussos, S., Record, E., Delattre, M., Lesage-Meessen, L., … Asther, M. (2002). Feruloyl esterase from Aspergillus niger. *Process Biochemistry*. https://doi.org/10.1016/s0032-9592(02)00196-6

Austin, H. P., Allen, M. D., Donohoe, B. S., Rorrer, N. A., Kearns, F. L., Silveira, R. L., … Beckham, G. T. (2018). Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), E4350–E4357.

Baumann, M., Hauer, B. H., & Bornscheuer, U. T. (2000). Rapid screening of hydrolases for the enantioselective conversion of “difficult-to-resolve” substrates. *Tetrahedron: Asymmetry*. https://doi.org/10.1016/s0957-4166(00)00465-1

Beloqui, A., de María, P. D., Golyshin, P. N., & Ferrer, M. (2008). Recent trends in industrial microbiology. *Current Opinion in Microbiology*, *11*(3), 240–248.

Blank, L. M., Ebert, B. E., Buehler, K., & Bühler, B. (2010). Redox biocatalysis and metabolism: molecular mechanisms and metabolic network analysis. *Antioxidants & Redox Signaling*, *13*(3), 349–394.

Bolon, D. N., & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(25), 14274–14279.

Bornscheuer, U. T. (2002a). Methods to increase enantioselectivity of lipases and esterases. *Current Opinion in Biotechnology*. https://doi.org/10.1016/s0958-1669(02)00350-6

Bornscheuer, U. T. (2002b). Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiology Reviews*. https://doi.org/10.1111/j.1574-6976.2002.tb00599.x

Bornscheuer, U. T., Bottcher, D., & Schmidt, M. (2009). Protein Engineering of Carboxyl Esterases by Rational Design and Directed Evolution. *Protein & Peptide Letters*. https://doi.org/10.2174/092986609789071216

Bornscheuer, U. T., & Kazlauskas, R. J. (2005). Hydrolases in Organic Synthesis. https://doi.org/10.1002/3527607544

Borrelli, K. W., Vitalis, A., Alcantara, R., & Guallar, V. (2005). PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of Chemical Theory and Computation*, *1*(6), 1304–1311.

Bourbonnais, R., & Paice, M. G. (1990). Oxidation of non-phenolic substrates. An expanded role for laccase in lignin biodegradation. *FEBS Letters*, *267*(1), 99–102.

Brady, L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E. J., Dodson, G. G., Tolley, S. P., … Thim, L. (1990). A SERINE PROTEASE TRIAD FORMS THE CATALYTIC CENTRE OF A TRIACYLGLYCEROL LIPASE. https://doi.org/10.2210/pdb1tgl/pdb

Bridges, B. A., & Woodgate, R. (1985). Mutagenic repair in Escherichia coli: products of the recA gene and of the umuD and umuC genes act at different steps in UV-induced mutagenesis.

Proceedings of the National Academy of Sciences of the United States of America, *82*(12), 4193–4197.

Carboni-Oerlemans, C., Domínguez de María, P., Tuin, B., Bargeman, G., van der Meer, A., & van Gemert, R. (2006). Hydrolase-catalysed synthesis of peroxycarboxylic acids: Biocatalytic promiscuity for practical applications. *Journal of Biotechnology*, *126*(2), 140–151.

Carro, J., Amengual-Rigo, P., Sancho, F., Medina, M., Guallar, V., Ferreira, P., & Martínez, A. T. (2018). Multiple implications of an active site phenylalanine in the catalysis of aryl-alcohol oxidase. *Scientific Reports*, *8*(1), 8121.

Catherine, H., Penninckx, M., & Frédéric, D. (2016). Product formation from phenolic compounds removal by laccases: A review. *Environmental Technology & Innovation*. https://doi.org/10.1016/j.eti.2016.04.001

Chapman, J., Ismail, A., & Dinu, C. (2018). Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks. *Catalysts*, *8*(6), 238.

Cheeseman, J. D., Tocilj, A., Park, S., Schrag, J. D., & Kazlauskas, R. J. (2004). Structure of an aryl esterase fromPseudomonas fluorescens. *Acta Crystallographica Section D Biological Crystallography*. https://doi.org/10.1107/s0907444904010522

Chen, C.-S., & Sih, C. J. (1989). General Aspects and Optimization of Enantioselective Biocatalysis in Organic Solvents: The Use of Lipases[New Synthetic Methods(76)]. *Angewandte Chemie International Edition in English*. https://doi.org/10.1002/anie.198906951

Chen, R. (1999). A general strategy for enzyme engineering. *Trends in Biotechnology*. https://doi.org/10.1016/s0167-7799(99)01324-4

Chen, R., Greer, A., & Dean, A. M. (1996). Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.93.22.12171

Chen, Y., Stemple, B., Kumar, M., & Wei, N. (2016). Cell Surface Display Fungal Laccase as a Renewable Biocatalyst for Degradation of Persistent Micropollutants Bisphenol A and Sulfamethoxazole. *Environmental Science & Technology*.

https://doi.org/10.1021/acs.est.6b01641

Chen, Z., Durão, P., Silva, C. S., Pereira, M. M., Todorovic, S., Hildebrandt, P., … Martins, L. O. (2010). The role of Glu498 in the dioxygen reactivity of CotA-laccase from Bacillus subtilis. *Dalton Transactions* , *39*(11), 2875–2882.

Choi, J.-M., Han, S.-S., & Kim, H.-S. (2015). Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnology Advances*. https://doi.org/10.1016/j.biotechadv.2015.02.014

Christakopoulos, P., Tzalas, B., Mamma, D., Stamatis, H., Liadakis, G. N., Tzia, C., … Macris, B. J. (1998). Production of an esterase from Fusarium oxysporum catalysing transesterification reactions in organic solvents. *Process Biochemistry*. https://doi.org/10.1016/s0032-9592(98)00039-9

Christopher, L. P., Yao, B., & Ji, Y. (2014). Lignin Biodegradation with Laccase-Mediator Systems. *Frontiers in Energy Research*. https://doi.org/10.3389/fenrg.2014.00012

Clark, D. P., & Pazdernik, N. J. (2016). Protein Engineering. *Biotechnology*. https://doi.org/10.1016/b978-0-12-385015-7.00011-9

Claus, H. (2003). Laccases and their occurrence in prokaryotes. *Archives of Microbiology*, *179*(3), 145–150.

Claus, H. (2004). Laccases: structure, reactions, distribution. *Micron* , *35*(1-2), 93–96.

Ding, H., Wu, Y., Zou, B., Lou, Q., Zhang, W., Zhong, J., … Dai, G. (2016). Simultaneous removal and degradation characteristics of sulfonamide, tetracycline, and quinolone antibiotics by laccase-mediated oxidation coupled with soil adsorption. *Journal of Hazardous Materials*. https://doi.org/10.1016/j.jhazmat.2015.12.062

Dodson, G. G., Guy Dodson, G., Lawson, D. M., & Winkler, F. K. (1992). Structural and evolutionary relationships in lipase mechanism and activation. *Faraday Discussions*. https://doi.org/10.1039/fd9929300095

Ducros, V., Brzozowski, A. M., Wilson, K. S., Brown, S. H., Ostergaard, P., Schneider, P., … Davies, G. J. (1998). Crystal structure of the type-2 Cu depleted laccase from Coprinus

cinereus at 2.2 A resolution. *Nature Structural Biology*, *5*(4), 310–316.

Durão, P., Chen, Z., Silva, C. S., Soares, C. M., Pereira, M. M., Todorovic, S., … Martins, L. O. (2008). Proximal mutations at the type 1 copper site of CotA laccase: spectroscopic, redox, kinetic and structural characterization of I494A and L386A mutants. *Biochemical Journal*, *412*(2), 339–346.

Elouarzaki, K., Cheng, D., Fisher, A. C., & Lee, J.-M. (2018). Coupling orientation and mediation strategies for efficient electron transfer in hybrid biofuel cells. *Nature Energy*. https://doi.org/10.1038/s41560-018-0166-4

Eriksen, D. T., Lian, J., & Zhao, H. (2014). Protein design for pathway engineering. *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2013.03.011

Fodor, C., Gajewska, B., Rifaie-Graham, O., Apebende, E. A., Pollard, J., & Bruns, N. (2016). Laccase-catalyzed controlled radical polymerization of N-vinylimidazole. *Polymer Chemistry*, *7*(43), 6617–6625.

Fraczkiewicz, R., & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of Computational Chemistry*. https://doi.org/3.0.co;2-w">10.1002/(sici)1096-987x(199802)19:3<319::aid-jcc6>3.0.co;2-w

Freese, E. (1959). The specific mutagenic effect of base analogues on Phage T4. *Journal of Molecular Biology*. https://doi.org/10.1016/s0022-2836(59)80038-3

Fukuda, K., Kiyokawa, Y., Yanagiuchi, T., Wakai, Y., Kitamoto, K., Inoue, Y., & Kimura, A. (2000). Purification and characterization of isoamyl acetate-hydrolyzing esterase encoded by the IAH1 gene of Saccharomyces cerevisiae from a recombinant Escherichia coli. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s002530051662

Gasser, C. A., Ammann, E. M., Shahgaldian, P., & -X. Corvini, P. F. (2014). Laccases to take on the challenge of emerging organic contaminants in wastewater. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-014-6177-6

Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., & Hallewell, R. A. (1992). Faster superoxide dismutase mutants designed by enhancing electrostatic

guidance. *Nature*. https://doi.org/10.1038/358347a0

Ghanem, A. (2007). Trends in Lipase-Catalyzed Asymmetric Access to Enantiomerically Pure/Enriched Compounds. *ChemInform*. https://doi.org/10.1002/chin.200721237

Gonzalez-Perez, D., Molina-Espeja, P., Garcia-Ruiz, E., & Alcalde, M. (2014). Mutagenic Organized Recombination Process by Homologous IN vivo Grouping (MORPHING) for directed enzyme evolution. *PloS One*, *9*(3), e90919.

Gupta, A., Nederlof, I., Sottini, S., Tepper, A. W. J. W., Groenen, E. J. J., Thomassen, E. A. J., & Canters, G. W. (2012). Involvement of Tyr108 in the enzyme mechanism of the small laccase from Streptomyces coelicolor. *Journal of the American Chemical Society*, *134*(44), 18213–18216.

Hakulinen, N., & Rouvinen, J. (2015). Three-dimensional structures of laccases. *Cellular and Molecular Life Sciences: CMLS*, *72*(5), 857–868.

Hellinga, H. W., & Richards, F. M. (1991). Construction of new ligand binding sites in proteins of known structure. *Journal of Molecular Biology*. https://doi.org/10.1016/0022-2836(91)90510-d

Horne, I., Harcourt, R. L., Sutherland, T. D., Russell, R. J., & Oakeshott, J. G. (2002). Isolation of aPseudomonas monteillistrain with a novel phosphotriesterase. *FEMS Microbiology Letters*. https://doi.org/10.1111/j.1574-6968.2002.tb10985.x

Howard, G. T., Crother, B., & Vicknair, J. (2001). Cloning, nucleotide sequencing and characterization of a polyurethanase gene (pueB) from Pseudomonas chlororaphis. *International Biodeterioration & Biodegradation*. https://doi.org/10.1016/s0964-8305(01)00042-7

Huang, N., Kalyanaraman, C., Bernacki, K., & Jacobson, M. P. (2006). Molecular mechanics methods for predicting protein–ligand binding. *Phys. Chem. Chem. Phys.* https://doi.org/10.1039/b608269f

Hult, K., & Berglund, P. (2007). Enzyme promiscuity: mechanism and applications. *Trends in Biotechnology*, *25*(5), 231–238.

Hurley, J. H., Chen, R., & Dean, A. M. (1996). Determinants of Cofactor Specificity in Isocitrate Dehydrogenase: Structure of an Engineered NADP → NAD Specificity-Reversal Mutant†,‡. *Biochemistry*. https://doi.org/10.1021/bi953001q

Husain, Q., & Qayyum, S. (2013). Biological and enzymatic treatment of bisphenol A and other endocrine disrupting compounds: a review. *Critical Reviews in Biotechnology*. https://doi.org/10.3109/07388551.2012.694409

Jahangir, R., McCloskey, C. B., Mc Clung, W. G., Labow, R. S., Brash, J. L., & Santerre, J. P. (2003). The influence of protein adsorption and surface modifying macromolecules on the hydrolytic degradation of a poly(ether–urethane) by cholesterol esterase. *Biomaterials*. https://doi.org/10.1016/s0142-9612(02)00269-7

Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., … Baker, D. (2008). De novo computational design of retro-aldol enzymes. *Science*, *319*(5868), 1387–1391.

Jochens, H., & Bornscheuer, U. T. (2010). Natural Diversity to Guide Focused Directed Evolution. *ChemBioChem*. https://doi.org/10.1002/cbic.201000284

Jochens, H., Hesseler, M., Stiba, K., Padhi, S. K., Kazlauskas, R. J., & Bornscheuer, U. T. (2011). Protein Engineering of α/β-Hydrolase Fold Enzymes. *ChemBioChem*. https://doi.org/10.1002/cbic.201000771

Kapoor, M., & Gupta, M. N. (2012). Lipase promiscuity and its biochemical applications. *Process Biochemistry*. https://doi.org/10.1016/j.procbio.2012.01.011

Karplus, M. (2002). Molecular dynamics simulations of biomolecules. *Accounts of Chemical Research*, *35*(6), 321–323.

Karplus, M., & Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature*. https://doi.org/10.1038/347631a0

Karra-Chaabouni, M., Pulvin, S., Touraud, D., & Thomas, D. (1996). Enzymatic synthesis of geraniol esters in a solvent-free system by lipases. *Biotechnology Letters*. https://doi.org/10.1007/bf00129736

Kermasha, S., Bisakowski, B., Ismail, S., & Morin, A. (2000). The effect of physical and chemical treatments on the esterase activity from Pseudomonas fragi CRDA 037. *Food Research International*. https://doi.org/10.1016/s0963-9969(00)00080-6

Kim, G.-J., Choi, G.-S., Kim, J.-Y., Lee, J.-B., Jo, D.-H., & Ryu, Y.-W. (2002). Screening, production and properties of a stereospecific esterase from Pseudomonas sp. S34 with high selectivity to (S)-ketoprofen ethyl ester. *Journal of Molecular Catalysis B: Enzymatic*. https://doi.org/10.1016/s1381-1177(01)00077-7

Kim, Y.-H., Cha, C.-J., & Cerniglia, C. E. (2002). Purification and characterization of an erythromycin esterase from an erythromycin-resistantPseudomonassp. *FEMS Microbiology Letters*. https://doi.org/10.1111/j.1574-6968.2002.tb11187.x

Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D., & Houk, K. N. (2013). Computational Enzyme Design. *Angewandte Chemie International Edition*. https://doi.org/10.1002/anie.201204077

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949.

Kontkanen, H., Tenkanen, M., Fagerström, R., & Reinikainen, T. (2004). Characterisation of steryl esterase activities in commercial lipase preparations. *Journal of Biotechnology*. https://doi.org/10.1016/j.jbiotec.2003.11.003

Kries, H., Blomberg, R., & Hilvert, D. (2013). De novo enzymes by computational design. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2013.02.012

Kuchner, O., & Arnold, F. H. (1997). Directed evolution of enzyme catalysts. *Trends in Biotechnology*. https://doi.org/10.1016/s0167-7799(97)01138-4

Kuipers, R. K., Joosten, H.-J., van Berkel, W. J. H., Leferink, N. G. H., Rooijen, E., Ittmann, E., … Schaap, P. J. (2010). 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins*, *78*(9), 2101–2113.

Lai, Y.-P., Huang, J., Wang, L.-F., Li, J., & Wu, Z.-R. (2004). A new approach to random mutagenesis in vitro. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.20066

Lang, D. A., & Dijkstra, B. W. (1998). Structural investigations of the regio- and enantioselectivity of lipases. *Chemistry and Physics of Lipids*. https://doi.org/10.1016/s0009-3084(98)00035-8

Lang, D. A., Mannesse, M. L. M., De Haas, G. H., Verheij, H. M., & Dijkstra, B. W. (1998). Structural basis of the chiral selectivity of Pseudomonas cepacia lipase. *European Journal of Biochemistry*. https://doi.org/10.1046/j.1432-1327.1998.2540333.x

Larsson, D. G. J., & Joakim Larsson, D. G. (2014). Antibiotics in the environment. *Upsala Journal of Medical Sciences*. https://doi.org/10.3109/03009734.2014.896438

Laurent, C. V. F. P., Breslmayr, E., Tunega, D., Ludwig, R., & Oostenbrink, C. (2019). Interaction between Cellobiose Dehydrogenase and Lytic Polysaccharide Monooxygenase. *Biochemistry*, *58*(9), 1226–1235.

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*, *10*, 168.

Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., & Chatonnet, A. (2013). ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Research*, *41*(Database issue), D423–D429.

Librando, V., & Pappalardo, M. (2013). In silico bioremediation of polycyclic aromatic hydrocarbon: A frontier in environmental chemistry. *Journal of Molecular Graphics and Modelling*. https://doi.org/10.1016/j.jmgm.2013.04.011

Lind, M. E. S., & Himo, F. (2013). Quantum chemistry as a tool in asymmetric biocatalysis: limonene epoxide hydrolase test case. *Angewandte Chemie* , *52*(17), 4563–4567.

Lomize, A. L., Hage, J. M., & Pogozheva, I. D. (2018). Membranome 2.0: database for proteome-wide profiling of bitopic proteins and their dimers. *Bioinformatics* , *34*(6), 1061–1062.

Lomize, A. L., Lomize, M. A., Krolicki, S. R., & Pogozheva, I. D. (2017). Membranome: a database for proteome-wide analysis of single-pass membrane proteins. *Nucleic Acids Research*, *45*(D1), D250–D255.

Majeau, J.-A., Brar, S. K., & Tyagi, R. D. (2010). Laccases for removal of recalcitrant and

emerging pollutants. *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2009.10.087

Malisi, C., Kohlbacher, O., & Höcker, B. (2009). Automated scaffold selection for enzyme design. *Proteins: Structure, Function, and Bioinformatics*. https://doi.org/10.1002/prot.22418

Manco, G., Porzio, E., & Suzumoto, Y. (2018). Enzymatic detoxification: a sustainable means of degrading toxic organophosphate pesticides and chemical warfare nerve agents. *Journal of Chemical Technology & Biotechnology*. https://doi.org/10.1002/jctb.5603

Martí, M. A., Bidon-Chanal, A., Crespo, A., Yeh, S.-R., Guallar, V., Luque, F. J., & Estrin, D. A. (2008). Mechanism of product release in NO detoxification from Mycobacterium tuberculosis truncated hemoglobin N. *Journal of the American Chemical Society*, *130*(5), 1688–1693.

Marti, M. A., Capece, L., Bidon-Chanal, A., Crespo, A., Guallar, V., Luque, F. J., & Estrin, D. A. (2008). Nitric oxide reactivity with globins as investigated through computer simulation. *Methods in Enzymology*, *437*, 477–498.

Martínez, A. T., Ruiz-Dueñas, F. J., Camarero, S., Serrano, A., Linde, D., Lund, H., … Alcalde, M. (2017). Oxidoreductases on their way to industrial biotransformations. *Biotechnology Advances*, *35*(6), 815–831.

Martí, S., Andrés, J., Moliner, V., Silla, E., Tuñón, I., & Bertrán, J. (2008). Predicting an Improvement of Secondary Catalytic Activity of Promiscuos Isochorismate Pyruvate Lyase by Computational Design. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja078334c

Maté, D., García-Burgos, C., García-Ruiz, E., Ballesteros, A. O., Camarero, S., & Alcalde, M. (2010). Laboratory evolution of high-redox potential laccases. *Chemistry & Biology*, *17*(9), 1030–1041.

Mate, D. M., & Alcalde, M. (2015). Laccase engineering: From rational design to directed evolution. *Biotechnology Advances*. https://doi.org/10.1016/j.biotechadv.2014.12.007

Mateljak, I., Monza, E., Lucas, M. F., Guallar, V., Aleksejeva, O., Ludwig, R., … Alcalde, M. (2019). Increasing Redox Potential, Redox Mediator Activity, and Stability in a Fungal Laccase by Computer-Guided Mutagenesis and Directed Evolution. *ACS Catalysis*.

https://doi.org/10.1021/acscatal.9b00531

McCullum, E. O., Williams, B. A. R., Zhang, J., & Chaput, J. C. (2010). Random Mutagenesis by Error-Prone PCR. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-60761-652-8_7

Mentus, S., Cirić-Marjanović, G., Trchová, M., & Stejskal, J. (2009). Conducting carbonized polyaniline nanotubes. *Nanotechnology*, *20*(24), 245601.

Monza, E., Fatima Lucas, M., Camarero, S., Alejaldre, L. C., Martínez, A. T., & Guallar, V. (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.jpclett.5b00225

Morozova, O. V., Shumakovich, G. P., Shleev, S. V., & Iaropolov, A. I. (2007). [Laccase-mediator systems and their applications: a review]. *Prikladnaia biokhimiia i mikrobiologiia*, *43*(5), 583–597.

Murphy, R. B., Philipp, D. M., & Friesner, R. A. (2000). A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *Journal of Computational Chemistry*. https://doi.org/3.0.co;2-o">10.1002/1096-987x(200012)21:16<1442::aid-jcc3>3.0.co;2-o

Nardini, M., & Dijkstra, B. W. (1999). α/β Hydrolase fold enzymes: the family keeps growing. *Current Opinion in Structural Biology*. https://doi.org/10.1016/s0959-440x(99)00037-8

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., … Goldman, A. (1992). The α/β hydrolase fold. *"Protein Engineering, Design and Selection."* https://doi.org/10.1093/protein/5.3.197

Oulton, R. L., Kohn, T., & Cwiertny, D. M. (2010). Pharmaceuticals and personal care products in effluent matrices: A survey of transformation and removal during wastewater treatment and implications for wastewater management. *Journal of Environmental Monitoring*. https://doi.org/10.1039/c0em00068j

Panda, T., & Gowrishankar, B. S. (2005). Production and applications of esterases. *Applied*

Microbiology and Biotechnology. https://doi.org/10.1007/s00253-004-1840-y

Pardo, I., Santiago, G., Gentili, P., Lucas, F., Monza, E., Medrano, F. J., … Camarero, S. (2016). Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. *Catalysis Science & Technology*. https://doi.org/10.1039/c5cy01725d

Philipp, D. M., & Friesner, R. A. (1999). Mixedab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *Journal of Computational Chemistry*. https://doi.org/3.0.co;2-0">10.1002/(sici)1096-987x(19991115)20:14<1468::aid-jcc2>3.0.co;2-0

Potapov, V., Cohen, M., & Schreiber, G. (2009). Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Engineering, Design & Selection: PEDS*, *22*(9), 553–560.

Rao, M. A., Scelza, R., Acevedo, F., Diez, M. C., & Gianfreda, L. (2014). Enzymes as useful tools for environmental purposes. *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2013.12.059

Ribeiro, D. S., Henrique, S. M. B., Oliveira, L. S., Macedo, G. A., & Fleuri, L. F. (2010). Enzymes in juice processing: a review. *International Journal of Food Science & Technology*. https://doi.org/10.1111/j.1365-2621.2010.02177.x

Richter, F., Blomberg, R., Khare, S. D., Kiss, G., Kuzin, A. P., Smith, A. J. T., … Baker, D. (2012). Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja3037367

Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, *59*, 1–41.

Rodríguez-Couto, S. (2019). Fungal Laccase: A Versatile Enzyme for Biotechnological Applications. *Springer Proceedings in Mathematics & Statistics*. https://doi.org/10.1007/978-3-030-10480-1_13

Rodríguez Couto, S., Couto, S. R., & Herrera, J. L. T. (2006). Industrial and biotechnological applications of laccases: A review. *Biotechnology Advances*, *24*(5), 500–513.

Rodríguez-Delgado, M., & Ornelas-Soto, N. (2017). Laccases: A Blue Enzyme for Greener Alternative Technologies in the Detection and Treatment of Emerging Pollutants. *Green Technologies and Environmental Sustainability*. https://doi.org/10.1007/978-3-319-50654-8_2

Sancho, F., Santiago, G., Amengual-Rigo, P., & Guallar, V. (n.d.). CHAPTER 10. Modeling O2-dependent Heme Enzymes: A Quick Guide for Non-experts. *Dioxygen-Dependent Heme Enzymes*. https://doi.org/10.1039/9781788012911-00222

Santhanam, N., Vivanco, J. M., Decker, S. R., & Reardon, K. F. (2011). Expression of industrially relevant laccases: prokaryotic style. *Trends in Biotechnology*, *29*(10), 480–489.

Schmid, R. D., & Verger, R. (1998). Lipases: Interfacial Enzymes with Attractive Applications. *Angewandte Chemie* , *37*(12), 1608–1633.

Schmidt, M., Hasenpusch, D., Kähler, M., Kirchner, U., Wiggenhorn, K., Langel, W., & Bornscheuer, U. T. (2006). Directed Evolution of an Esterase from Pseudomonas fluorescens Yields a Mutant with Excellent Enantioselectivity and Activity for the Kinetic Resolution of a Chiral Building Block. *ChemBioChem*. https://doi.org/10.1002/cbic.200500546

Schrag, J. D., & Cygler, M. (1997). [4] Lipases and hydrolase fold. *Methods in Enzymology*. https://doi.org/10.1016/s0076-6879(97)84006-2

Senn, H. M., & Thiel, W. (2007). QM/MM studies of enzymes. *Current Opinion in Chemical Biology*, *11*(2), 182–187.

Sen, S. K., Raut, S., Bandyopadhyay, P., & Raut, S. (2016). Fungal decolouration and degradation of azo dyes: A review. *Fungal Biology Reviews*. https://doi.org/10.1016/j.fbr.2016.06.003

Sewell, S. L. (2014). The Spatial Diffusion of Beer from its Sumerian Origins to Today. In *The Geography of Beer* (pp. 23–29).

Shaik, S., Cohen, S., Wang, Y., Chen, H., Kumar, D., & Thiel, W. (2010). P450 enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chemical Reviews*, *110*(2), 949–1017.

Shen, D., Xu, J.-H., Wu, H.-Y., & Liu, Y.-Y. (2002). Significantly improved esterase activity of

Trichosporon brassicae cells for ketoprofen resolution by 2-propanol treatment. *Journal of Molecular Catalysis B: Enzymatic*. https://doi.org/10.1016/s1381-1177(02)00099-1

Shi, L., Ma, F., Han, Y., Zhang, X., & Yu, H. (2014). Removal of sulfonamide antibiotics by oriented immobilized laccase on Fe3O4 nanoparticles with natural mediators. *Journal of Hazardous Materials*. https://doi.org/10.1016/j.jhazmat.2014.06.070

Shu, X.-G., Su, J.-H., Du, K.-J., You, Y., Gao, S.-Q., Wen, G.-B., … Lin, Y.-W. (2016). Rational Design of Dual Active Sites in a Single Protein Scaffold: A Case Study of Heme Protein in Myoglobin. *ChemistryOpen*. https://doi.org/10.1002/open.201500224

Siegbahn, P. E. M., & Himo, F. (2009). Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry*, *14*(5), 643–651.

Siegbahn, P. E. M., & Himo, F. (2011). The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. https://doi.org/10.1002/wcms.13

Singh, R., Kumar, M., Mittal, A., & Mehta, P. K. (2016). Microbial enzymes: industrial progress in 21st century. *3 Biotech*, *6*(2), 174.

Sirim, D., Wagner, F., Wang, L., Schmid, R. D., & Pleiss, J. (2011). The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. *Database: The Journal of Biological Databases and Curation*, *2011*, bar006.

Spyrakis, F., Lucas, F., Bidon-Chanal, A., Viappiani, C., Guallar, V., & Luque, F. J. (2013). Comparative analysis of inner cavities and ligand migration in non-symbiotic AHb1 and AHb2. *Biochimica et Biophysica Acta*, *1834*(9), 1957–1967.

Stemmer, W. P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.91.22.10747

Strong, P. J., & Claus, H. (2011). Laccase: A Review of Its Past and Its Future in Bioremediation. *Critical Reviews in Environmental Science and Technology*.

https://doi.org/10.1080/10643380902945706

Suda, T., Hata, T., Kawai, S., Okamura, H., & Nishida, T. (2012). Treatment of tetracycline antibiotics by laccase in the presence of 1-hydroxybenzotriazole. *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2011.10.041

Tanimu, A., Jaenicke, S., & Alhooshani, K. (2017). Heterogeneous catalysis in continuous flow microreactors: A review of methods and applications. *Chemical Engineering Journal*. https://doi.org/10.1016/j.cej.2017.06.161

Tantillo, D. J., Jiangang, C., & Houk, K. N. (1998). Theozymes and compuzymes: theoretical models for biological catalysis. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/s1367-5931(98)80112-9

Thomson, A. R., Wood, C. W., Burton, A. J., Bartlett, G. J., Sessions, R. B., Leo Brady, R., & Woolfson, D. N. (2014). Computational design of water-soluble α-helical barrels. *Science*. https://doi.org/10.1126/science.1257452

Uppenberg, J., Hansen, M. T., Patkar, S., & Alwyn Jones, T. (1994). The sequence, crystal structure determination and refinement of two crystal forms of lipase B from Candida antarctica. *Structure*. https://doi.org/10.1016/s0969-2126(00)00031-9

Vaissier Welborn, V., & Head-Gordon, T. (2018). Computational Design of Synthetic Enzymes. *Chemical Reviews*. https://doi.org/10.1021/acs.chemrev.8b00399

Valero, F. (2012). Heterologous expression systems for lipases: a review. *Methods in Molecular Biology* , *861*, 161–178.

Viswanath, B., Rajesh, B., Janardhan, A., Kumar, A. P., & Narasimha, G. (2014). Fungal Laccases and Their Applications in Bioremediation. *Enzyme Research*. https://doi.org/10.1155/2014/163242

Wang, Y., Arandiyan, H., Scott, J., Bagheri, A., Dai, H., & Amal, R. (2017). Recent advances in ordered meso/macroporous metal oxides for heterogeneous catalysis: a review. *Journal of Materials Chemistry A*. https://doi.org/10.1039/c6ta10896b

Warshel, A., & Florián, J. (2004). The Empirical Valence Bond (EVB) Method. *Encyclopedia of*

Computational Chemistry. https://doi.org/10.1002/0470845015.cu0002

Winkler, F. K., D'Arcy, A., & Hunziker, W. (1990). Structure of human pancreatic lipase. *Nature*. https://doi.org/10.1038/343771a0

Wittstock, A., Biener, J., & Bäumer, M. (2010). Nanoporous gold: a new material for catalytic and sensor applications. *Physical Chemistry Chemical Physics*. https://doi.org/10.1039/c0cp00757a

Xu, F., Palmer, A. E., Yaver, D. S., Berka, R. M., Gambetta, G. A., Brown, S. H., & Solomon, E. I. (1999). Targeted Mutations in aTrametes villosaLaccase. *The Journal of Biological Chemistry*, *274*(18), 12372–12375.

Y.-J., C., -J., C. Y., & B., L. (2001). Culture conditions for the production of esterase from Lactobacillus casei CL96. *Bioprocess and Biosystems Engineering*. https://doi.org/10.1007/s004490100233

Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., … Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*. https://doi.org/10.1126/science.aad6359

Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Meiler, J., Althoff, E. A., … Baker, D. (2006). New algorithms and an in silico benchmark for computational enzyme design. *Protein Science*. https://doi.org/10.1110/ps.062353106

Zeng, J., Zhu, Q., Wu, Y., & Lin, X. (2016). Oxidation of polycyclic aromatic hydrocarbons using Bacillus subtilis CotA with high laccase activity and copper independence. *Chemosphere*, *148*, 1–7.

Zhao, H. (1997). Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Research*. https://doi.org/10.1093/nar/25.6.1307