Computational study and rational design of pluriZymes

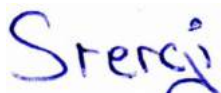Sergi Rodà Llordés

Departament de Química Física

Facultat de Química

Universitat de Barcelona

Programa de doctorat en Química Teòrica i Modelització Computacional

Computational study and rational design of pluriZymes

Memòria presentada per:

Sergi Rodà Llordés

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

Dirigida per:

Prof. Víctor Guallar Tasies   Dr. Gerard Santiago Morcillo

Tutora:

Prof. Mercè Deumal Solé

UNIVERSITAT DE BARCELONA

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Barcelona, Novembre de 2022

"Every living thing is a masterpiece, written by nature and edited by evolution."

Neil Degrasse Tyson

Contents

Acknowledgements

I have met a lot of people along this journey for which I am extremely grateful.

First and foremost, I would like to thank my first thesis director, prof. Víctor Guallar, for allowing me to pursue a Ph.D. in such an exciting field in his lab. He gave me plenty of freedom to work on whatever research I was interested in. He also taught me how to be a great scientist, highlighting the importance of being able to make one's research understandable.

I also want to thank my other thesis director, Dr. Gerard Santiago, for giving me good tips along the Ph.D., always being there when I needed him, and for giving me a perspective of what an industrial career can bring.

I want to thank my UB tutor, prof. Mercè Deumal, for helping me throughout the Ph.D. and facilitating the whole process.

I must thank all people I have worked with in Guallar's lab. I want to list them all, so here it goes; Joan, Alberto, Pep, Carles, Martí, Martin, Ana, Anna D., Anna L., Rubén, Mireia, Masoud, Alexis, Dani, Ruite, Ignasi, Ignacio, Roc, Miguel, Laura, Biel, Júlia, Marina, Quique, Albert, Isaac, Núria. Each and every one of you taught me something or gave me a laugh during the Ph.D., so I appreciate it.

It has been a true honor and a pleasure to collaborate with Prof. Manuel Ferrer from CSIC, and I will always be grateful for all he did for me during my thesis. In relation to his persona and lab, I have to thank all his students, especially, Laura Fernandez-Lopez, as she worked on all the experimental parts of our articles, and I could not have finished my Ph.D. without her outstanding job. Thanks to all the other experimental collaborators I have worked with, it has been a delight doing science together.

I am truly grateful for the opportunity Prof. Aleksej Zelezniak granted me to go for an internship in his lab and learn more about deep learning in bioinformatics. Once I was in Sweden, I was deeply fortunate to meet all the people from his lab; Xiaozhi, Alexander, Danish, Nikolaos, Sandra, and Samuel. Thanks to all of you for making the internship such an amazing experience.

Aside from work, I want to thank all the people from the BSC, especially those partici-pating in the BSC Basketball games during the Ph.D. It always made a bad Tuesday a

iv

good one.

Now is my turn to thank all my family and friends who gave me support during the whole journey. My parents taught me that success and achieving goals comes with hard work and sacrifice, so I thank them for always giving me those values and helping me persevere. My friends eased the passage through tough times and made life happier, so thanks as well.

Last but not least, I want and will always thank my girlfriend and future wife, Julieta. Her unconditional support and encouragement help me bring up the best version of myself every day. I love you.

Abstract

Being able to tailor biocatalysts at will is crucial for switching from inorganic catalysts to enzymes in the industry. The advances in computational power, followed by the repertoire of modeling tools, are pushing the design of the next generation of biocatalysts. This thesis aims at developing a novel concept, which refers to the pluriZymes. PluriZymes are proteins with plural catalytic active sites where one (at least) of them is artificially designed. The type of introduced active site has been the hydrolase site due to its simplicity (only 3 catalytic residues needed) and does not need a cofactor. Moreover, the studied systems were transaminases and esterases because they are widely used in the market, and extensively studied by our collaborators. Thus, the thesis presents a set of research articles where several pluriZymes have been engineered and a new algorithm to ease the design of artificial active sites. With all these advances in the design of pluriZymes, we hope people in the field get attracted to the idea.

List of Abbreviations

3D Three-dimensional

$d_{O^{Acid}_{\delta/\epsilon}-H^{His}_{\delta N}}$ Acid-His distance

$d_{O^{Asp}_{\delta}-H^{His}_{\delta N}}$ Asp-His distance

$d_{H^{Ser}_{\gamma}-N^{His}_{\epsilon}}$ Ser-His distance

$\boldsymbol{\Delta G}$ Gibbs free energy of the system

$\boldsymbol{\Delta G^{\ddagger}}$ Gibbs free energy of activation

$\boldsymbol{h}$ Planck constant

$\boldsymbol{\kappa}$ Transmission coefficient

$\boldsymbol{k_B}$ Boltzmann constant

$\boldsymbol{k_{cat}}$ Turnover number

$\boldsymbol{K_M}$ Michaelis constant

$\dfrac{\boldsymbol{k_{cat}}}{\boldsymbol{K_M}}$ Kinetic efficiency

AMBER **A**ssisted **M**odel **B**uilding with **E**nergy **R**efinement

ANM **A**nisotropic **N**etwork **M**odel

CalB *Candida antarctica* lipase **B**

CHARMM **C**hemistry **HAR**vard **M**acromolecular **M**echanics

CPU **C**entral **P**rocessing **U**nit

EC **E**nzyme **C**ommission number

GPU **G**raphical **P**rocessing **U**nit

GROMOS **GRO**ningen **MO**lecular **S**imulation

H-bond **H**ydrogen bond

MC **M**onte **C**arlo

MD **M**olecular **D**ynamics

MHET **M**ono**h**ydroxy**e**thyl **T**erephthalate

MM **M**olecular **M**echanics

OPLS **O**ptimized **P**otentials for **L**iquid **S**imulations

PCR **P**olymerase **C**hain **R**eaction

PDB **P**rotein **D**ata **B**ank

PELE **P**rotein **E**nergy **L**andscape **E**xploration

PES **P**otential **E**nergy **S**urface

PET **P**oly**e**thyl**e**ne **T**erephthalate

PFE *P**seudomonas** F**luorescens* esterase

QM **Q**uantum **M**echanics

RMSD **R**oot-**m**ean-**s**quare **D**eviation of atomic positions

SASA **S**olvent-**a**ccessible **S**urface **A**rea

TS **T**ransition **S**tate

WT **W**ild **T**ype

1 | Introduction

1.1 Enzymes

Enzymes are biological macromolecules that catalyze chemical reactions without being consumed during the process (Nelson and Cox, 2008a,e). In nature, enzymes are the elements that make life kinetically possible. In fact, some decarboxylation reactions are estimated to have half-time values close to the age of the Earth in the absence of enzymes (Neet, 1998; Wolfenden*, 2006; Wolfenden* and Snider, 2001). In contrast, enzymes can reduce these mentioned half-time values up to milliseconds. Thus, life, as we know it, could not be conceived without enzymes.

Although the details of the mechanism and the factors that enable enzymes to catalyze reactions are not fully comprehended yet (Knowles, 1991; Kraut, 1988), they either reduce the reaction TS energy barrier (decrease the value of $\Delta G^{\ddagger}$) or enhance the productive cross of the TS by the reactants (increase the value of κ) according to the widely known TS theory (*Figure 1.1*) (Eyring, 1935; Truhlar, 2015; Zinovjev and Tuñón, 2017).

The first theories to address how enzymes decreased the activation energy barrier were Fischer's "lock-and-key" model and Koshland's "induced-fit" model. The first model stated that the shape of the enzyme's active site (where the reaction occurs) is complementary to the substrate one (Fischer, 1894). The second, more recent one proposed that the shape of the active site of the enzyme changes along the reaction and is prepared to stabilize the TS, rather than the substrate (Koshland, 1958). Albeit the high usage of these theories in the field of enzymology, they lack a lot of key concepts like allosteric effects, macromolecule dynamics, and more (Kamerlin and Warshel, 2010; Kohen, 2014; Sutcliffe and Scrutton, 2000; Agarwal, 2006).
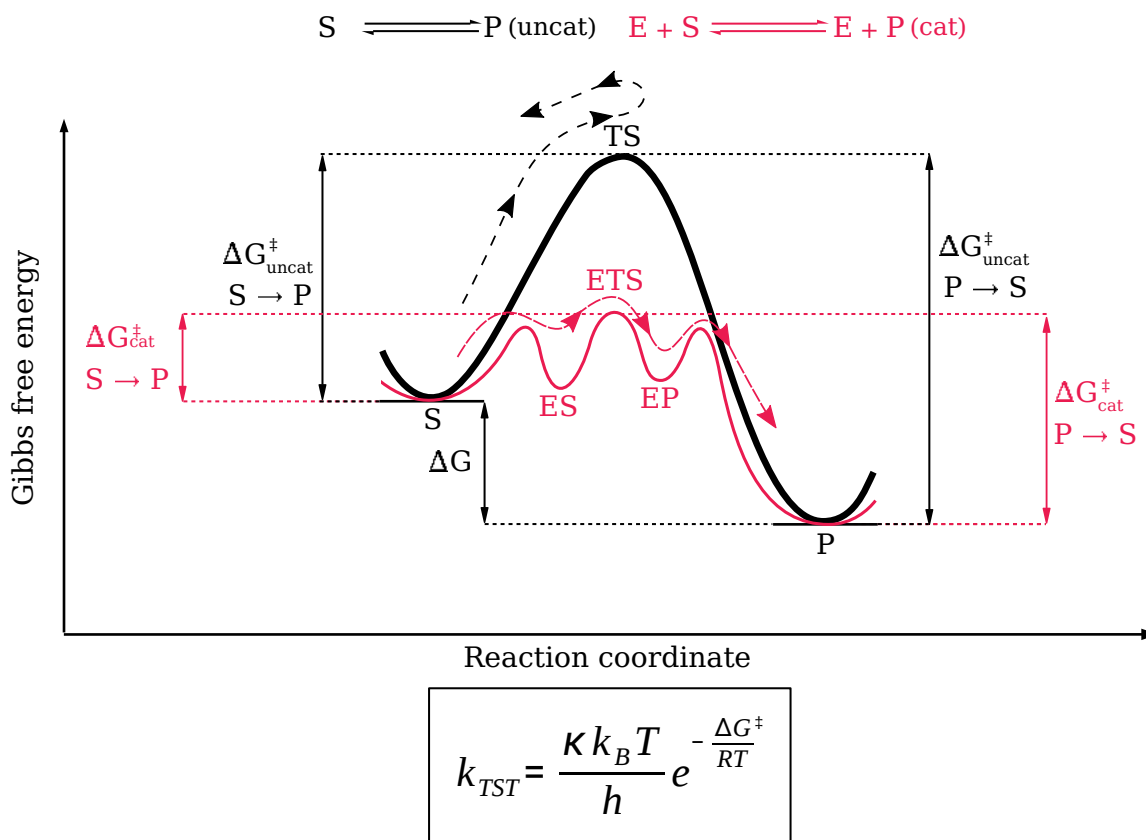
FIGURE 1.1: Free energy profile of a reaction without enzyme (in black, uncatalyzed) and with enzyme (in red, catalyzed). As it can be seen, enzymes can either decrease the value of $\Delta G^{\ddagger}$ or increase the number of productive crosses (as it can be seen by the curve with arrows along the reaction coordinate). The formation of the enzyme-substrate complex and the dissociation of the enzyme-product complex account for the two additional peaks in the catalyzed free energy profile. The S, P, TS, and E stand for substrate, product, transition state, and enzyme, respectively. Reaction coordinate corresponds to a metric that represents the progress of the chemical reaction. Figure adapted from (Roda et al., 2020).

1.1.1   Structure of enzymes

Typically, enzymes are proteins, meaning they are made from chains of the 21 canonical amino acids used in living organisms. These canonical amino acids are α amino acids with a carboxyl and an amino group bound to the α carbon (known as the main chain or backbone). What makes one amino acid different from another is the third group bound to the α carbon, which is named the side chain or R group (*Figure 1.2*).
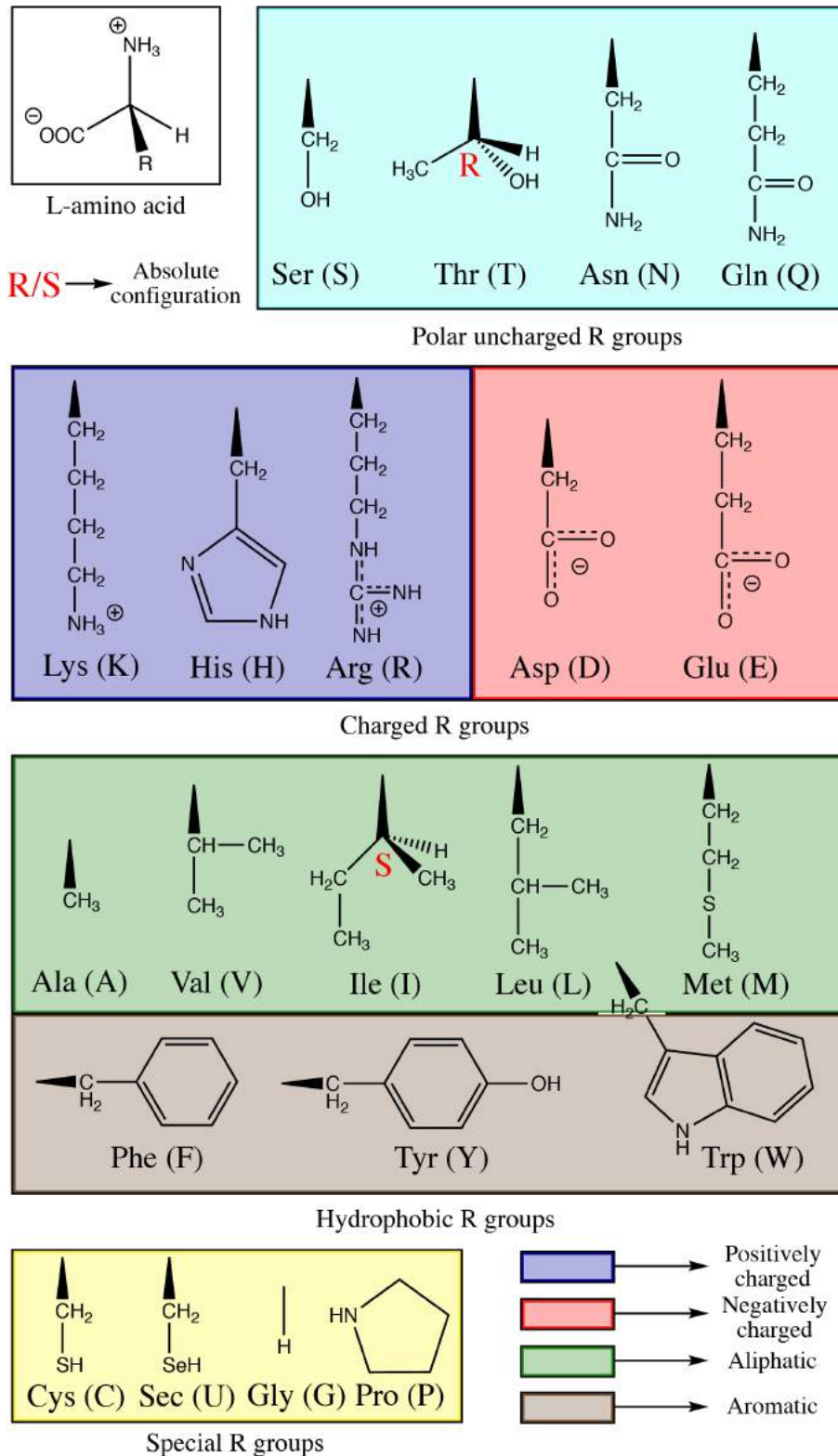
FIGURE 1.2: Classification of the 21 common amino acids according to their R group and their charge at physiological pH (7.4). Still, more classifications according to the nature of the R group exist due to the complexity of amino acids. Figure drawn with ChemDraw.

In proteins, these amino acids are covalently bound through the peptide bond and are called residues inside a protein. Each protein has a unique sequence of residues (sequence length, order, and the type of amino acids) that defines its structure and function. This sequence of residues is known as the primary protein structure, which includes the disulfide bonds of cysteine residues. After this first structural level, three others exist derived from the primary one. The arrangement of the backbone of the residues can acquire structural patterns due to local H-bonds in the structure. These patterns include the α-helix, the β-sheet, the turn, and the random coil. The structural disposition of these local arrangements in the space is a higher structural level, the tertiary structure. This tertiary structure explains the core function of the protein and is folded thanks mainly to non-local non-covalent interactions and entropic effects (like the hydrophobic core, salt bridges, post-translational modifications, and more). The last structural level, known as quaternary structure, happens when protein subunits bind to generate a larger complex (Nelson and Cox, 2008c,d) (*Figure 1.3*).
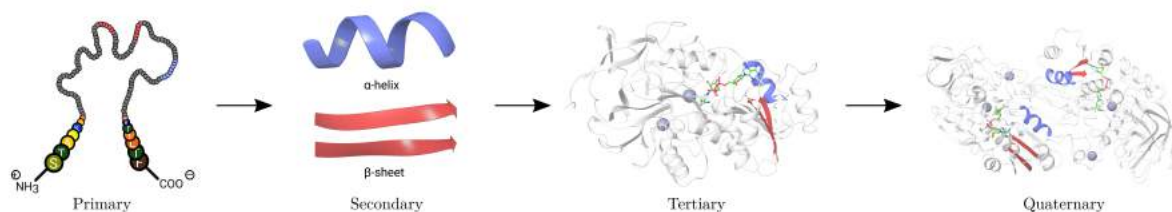

FIGURE 1.3: Representation of the four levels of structure in proteins in the case of an alcohol dehydrogenase (PDB ID: 2OHX).

Enzymes can also contain cofactors (including inorganic ions and complex organic molecules) and chemicals different from amino acids required for catalytic activity (and usually for folding). Hence, the wide variety that amino acids and cofactors offer explains the large number of enzymes found in nature and factors that can modulate their catalytic activity (such as pH, ionic concentration, temperature, presence or absence of the cofactor or activator/inhibitor molecules, and more (Robinson, 2015)).

Although the vast majority of enzymes are of proteic nature, RNA-based ("ribozymes"), DNA-based ("deoxyribozymes"), and even lipid-based enzymes (meaning that they show catalytic properties) have been found in nature (Breaker and Joyce, 2014; Silverman, 2004; Kruger et al., 1982; Breaker, 1997; Guerrier-Takada et al., 1983; Kahana and Lancet, 2021; Yeung et al., 2013; Taylor et al., 2017; Ihara et al., 1983). It has been recently hypothesized that catalytic lipid assemblies could have been the protocell precursors without replicating catalytic biopolymers (RNA or DNA) due to their catalytic properties shown in micelles and vesicular bilayers along with their autocatalytic proliferation and their prebiotic compatibility (Kahana and Lancet, 2021). Still, these lipids lost their catalytic properties through the evolution timeline and gained a more structural role (containment), leaving biopolymers (proteins, RNA, and DNA) as the preferred choice for catalysis. The

abundance of protein-based enzymes in nature compared to D/RNA-based enzymes can be explained due to the higher number of combinations that amino acids offer compared to nitrogenous bases ($21^n >> 4^n$, having more different sequences with the same length n of the macromolecule) (Doudna and Lorsch, 2005). However, ribozymes are still key in living organisms (they are the catalytic part of the ribosome, the machine that synthesizes proteins), and it is also thought that they were the first catalysts in the origin of life (Gilbert, 1986).

1.1.2 Enzymes' theory

As mentioned earlier, enzymes are catalysts, meaning the study of the kinetics of the catalyzed reaction must follow. We usually assume that enzymes follow the Michaelis-Menten kinetic model, which is based on a single-substrate reaction like the following:

$$E + S \xrightleftharpoons[k_{-1}]{k_1} ES \xrightarrow{k_2} E + P \tag{1.1}$$

Based on this drawn chemical reaction, the substrate must bind to the enzyme's active site (with a certain rate of association, k_1). Then, the reaction can occur, leading to product formation (with a specific catalytic rate, k_2), or the substrate can dissociate from the enzyme's active site and return to the solvent (with a particular rate of dissociation, k_{-1}). This leads to the reaction rate, depending on both the concentration of the substrate ($[S]$) and the concentration of the enzyme ($[E]$), as well as the kinetic constants of the specific enzyme (Johnson and Goody, 2011).

$$v = \frac{d[P]}{dt} = V_{max}\frac{[S]}{K_M + [S]} = k_{cat}\frac{[E][S]}{K_M + [S]} \tag{1.2}$$

This equation states that at $[S] \downarrow$, the reaction rate increases linearly with the $[S]$, as not all enzyme molecules are occupied. Still, at $[S] \uparrow$, the reaction rate approaches the V_{max} value, and thus, the speed of the reaction only depends on the $[E]$ and the turnover number of the enzyme (k_{cat}). k_{cat} is the number of times a substrate is converted into a product per unit of enzyme and time, and the constant can coincide with k_2 when there are not several intermediates involved in the catalytic mechanism. The Michaelis-Menten constant (K_M) is the substrate concentration at which the reaction rate is $\frac{V_{max}}{2}$. This constant is defined as:

$$K_M = \frac{k_2 + k_{-1}}{k_1} \tag{1.3}$$

Thus, K_M could be seen as a measure of the substrate's affinity for the enzyme (meaning the lower the K_M, the lower $[S]$ will be needed to reach the V_{max}), as it is described by

all constants affecting the enzyme-substrate complex.

Although the Michaelis-Menten model makes some assumptions (unimolecular reaction with no intermediates, $[S] >> [E]$, free diffusion ...), it is the most commonly employed model of enzyme kinetics, and it is used (with its assumptions) to experimentally characterize the k_{cat} and K_M values of an enzyme against a particular substrate in a specific set of conditions (for instance, temperature and pH).

Computationally predicting these two constants is not straightforward since they can depend on multiple rate constants. In the case of the k_{cat}, it can be estimated, by modeling the transition state and using Eyring's equation (*Figure 1.1*), on simple reactions where the rate-limiting step is known. Still, the exact catalytic mechanism of many enzymes is unknown (due to the existence of several hypotheses), and that is where modeling enzymes with molecular modeling hybrid (QM/MM) techniques become essential.

1.2   Enzyme engineering

Even though we mentioned the high "catalytic proficiency" of enzymes, only a few of them are known to have reached the maximal performance an enzyme can accomplish, which is the physical limit of diffusion rate ($\frac{k_{cat}}{K_M} > 10^9\,M^{-1} \cdot s^{-1}$) (Goldsmith and Tawfik, 2017; Bar-Even et al., 2011). In fact, the average $\frac{k_{cat}}{K_M}$ reported in the literature is 4 orders of magnitude below the perfect enzyme (Bar-Even et al., 2011). The reason behind this is that enzymes are optimized according to the physiological needs of the cell. Thus, their $\frac{k_{cat}}{K_M}$ value is limited by many factors occurring simultaneously in a living organism (such as the flux of metabolites) (Goldsmith and Tawfik, 2017).

However, enzymes can have their $\frac{k_{cat}}{K_M}$ improved by engineering them (Fasan et al., 2008; Goldsmith and Tawfik, 2017). This property and many others (an increase of thermal stability, switch of optimum pH, allowance to work in different types of solvents, acceptance of more substrates...) can be improved by enzyme engineering (Lutz and Bornscheuer, 2011; Jemli et al., 2016; Bell et al., 2021). Biocatalysts have gained relevance in the industry in recent years thanks to the improvements enzyme engineering can bring (Jemli et al., 2016; Bell et al., 2021; Hughes, 2018). Clear examples of this approach include the optimization of a ω-transaminase to enable the biosynthesis of sitagliptin (an antidiabetic drug) (Savile et al., 2010), the *de novo* design of enzymes to catalyze non-natural complex reactions (Siegel et al., 2010; Röthlisberger et al., 2008; Jiang et al., 2008), the design of an enzyme with multiple active sites to enable interesting one-pot cascade reactions (Santiago et al., 2018; Filice et al., 2015; Bos et al., 2012, 2015; Zhou and Roelfes, 2020), and many more (Lin et al., 2012; Santiago et al., 2016; Tournier et al., 2020). This engineering can be performed either with experimental or computational procedures (or by combining both of them (Steiner and Schwab, 2012; Chica et al., 2005)), and they will be further

presented below.

1.2.1 Experimental strategies

Enzyme engineering was born with the advances in molecular biology techniques, which included the usage of recombinant DNA for the production of the desired protein by the user (Cohen et al., 1973; Goeddel et al., 1979), the amplification of this DNA by the polymerase chain reaction (PCR) (Mullis, 1990), the site-specific manipulation of DNA with restriction nucleases and ligases (Winter et al., 1982; Wilkinson et al., 1983; Hutchison 3rd et al., 1978), and DNA sequencing (Sanger et al., 1977) (in fact, all these methods were awarded 3 Nobel Prizes, 2 in Chemistry and 1 in Physiology). These methods allowed for the specific mutation of some particular residues in a protein sequence, which was quickly used to make an enzyme (subtilisin) resistant to chemical oxidation with a single mutation (Met to Ser, Ala, or Leu) (Estell et al., 1985). Although the enzyme lost some catalytic activity, they could tailor the biocatalyst to the desired conditions. These first studies of enzyme engineering mainly used the knowledge of the crystallographic structure of the protein. This use of the available knowledge of the protein to smartly generate a small subset of variants that are expected to improve the desired property is what we call rational design (Chen, 1999, 2001). However, when there is little or no knowledge of the system, one can take advantage of the potential of natural selection to highly improve a property of an enzyme.

Directed evolution

Directed evolution uses random mutagenesis (Cadwell and Joyce, 1992) with the creation of a library of mutated versions of the studied gene, followed by the selection of the best ones according to the better performance of the protein variant reflected in the cell fitness (Chen, 2001; Kuchner and Arnold, 1997; Romero and Arnold, 2009). Then, the best-selected variants can be used for the next rounds of evolution (until the property you want to improve reaches the desired value). Thus, directed evolution requires: a) a method to randomly introduce changes in the studied enzyme; b) an efficient selection/screening method to be able to distinguish between the enhanced and deleterious variants; c) a practical evolution strategy that allows the improvement of the desired property in a short time and is cost-effective.

A lot of methods to develop random mutations on a DNA sequence exist (Labrou, 2010), such as error-prone PCR (where the used *Taq* DNA polymerase has low fidelity in checking the nitrogenous bases added in the synthesized DNA chain) (Cadwell and Joyce, 1992; Wilson and Keefe, 2001; McCullum et al., 2010), chemical damaging of the DNA (Chu et al., 1979; Shortle and Botstein, 1983), DNA shuffling (based on genetic recombination) (Zhao and Arnold, 1997), the use of bacterial mutator strains that lack DNA repair mech-

anisms (Cox, 1976; Bornscheuer et al., 1998), and even the newer CRISPR-Cas9 system can help (Jakočiūnas et al., 2018; Li et al., 2020). Each procedure has its limitations (Labrou, 2010; Lutz and Iamurri, 2018), but by combining them, large random libraries can be easily obtained. High-throughput screening/selection methods to test the library of randomly generated variants include the use of microtiter plates with colorimetric or fluorometric assays (where the aid of robots comes in handy) (Watt et al., 2000), direct digital imaging of the colonies considering the color observed in each pixel (Delagrave et al., 2001), flow cytometry with fluorometric assays (Yang and Withers, 2009), droplet-based microfluidics (a newer method that was created to overcome the limitations of the other ones) (Agresti et al., 2010), and the display techniques where the gene and the protein are directly or indirectly linked, and then, the protein is tested against the desired property (the most famous being the phage display (Smith, 1985; Smith and Petrenko, 1997), which was awarded half share of the Nobel Prize in Chemistry in 2018) (Xiao et al., 2015).

The strategy to follow for improving the desired property of your particular enzyme system with the mentioned methods depends on them. In other words, the higher number of mutations you try, the more options you will have to boost your catalyst. However, this trial of enormous libraries of variants can be extremely costly and time-consuming. To get an idea, if only a third of positions in a protein of 250 residues were allowed to mutate, the number of possible combinations is 20^{83}. This is where the user's system's available knowledge can help to create smaller libraries with a higher probability of containing sounder mutations (Steiner and Schwab, 2012; Chica et al., 2005), as well as fine-tuned reliable methods that allow the generation of a diversified set of mutants (Gonzalez-Perez et al., 2014). Despite its limitations, directed evolution has proven to be an essential tool for the improvement of enzymes with several reference studies (Savile et al., 2010; Heinzelman et al., 2009; Chen and Arnold, 1993; Moore and Arnold, 1996) (and even allows the creation of new chemistry (Hammer et al., 2017; Kan et al., 2016; Zhang et al., 2018)), and it has also been awarded the other half share of the Nobel Prize in Chemistry in 2018.

1.2.2 Computational strategies

The development of computational methods and more refined algorithms combined with the improvement in the power of computation (predicted by Moore's law) have taken enzyme engineering to a new level (Steiner and Schwab, 2012; Barrozo et al., 2012; Roda et al., 2020, 2021). Since the first trials of computational tools for protein (and enzyme) engineering (DeGrado et al., 1985; Craik et al., 1985; Hellinga et al., 1991; Wilson et al., 1991; Hurley et al., 1992), this field has thrived with several methods that can use as input any of the explained four structural levels of proteins (Roda et al., 2020). Although sequence-based approaches can be helpful to improve some properties of enzymes and
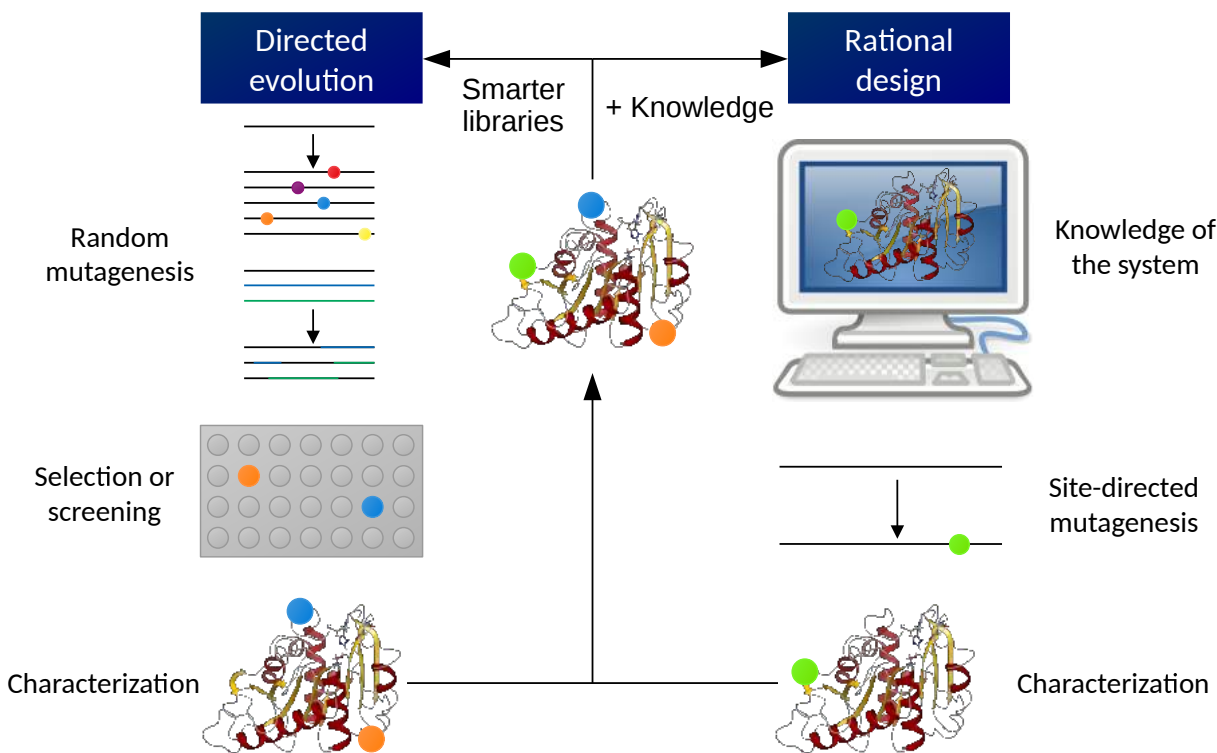
FIGURE 1.4: Comparison of directed evolution vs rational design for enzyme engineering.

primarily design smaller libraries for directed evolution, structural-based modeling is the primary technique used for computational enzyme engineering (Roda et al., 2020, 2021).

For instance, the computational modeling of an enzyme can help characterize its catalytic mechanism (Friesner and Guallar, 2005; Ranaghan and Mulholland, 2010; Lonsdale et al., 2012; Kamp and Mulholland, 2013), study its interactions with substrates (Gilabert et al., 2018; Roda et al., 2020), and determine its key conformational dynamics (Kamerlin and Warshel, 2010; Petrović et al., 2018). Then, this acquired information of the system can be applied to change the pH dependency of an enzyme based on the study of the dynamics of a loop (Shen et al., 2021), to enhance the activity of the enzyme based on analyzing the electrostatic environment in its active site (Santiago et al., 2016), or to switch the type of reaction the enzyme catalyzes based on modeling the active site residues and the TS of the reaction at the QM level of theory (Branneby et al., 2003).

Until recently, these types of computational studies and following rational redesigns of enzymes could only be done when the 3D structure was available, or a homologous sequence of high identity had a 3D structure that could be used as a template (since the structure is more conserved than sequences (Rost, 1999; Illergård et al., 2009)). This problem was a significant limitation as only a small portion of the protein sequences have an experimentally solved 3D structure deposited in the PDB. In fact, scientists have tried to unveil the physical rules that dictate a particular sequence of amino acids folds into its native

structure for half a century (Dill and MacCallum, 2012). Now, thanks to the power of deep learning, DeepMind, a subsidiary company from Google, and Baker's lab have created AlphaFold (Jumper et al., 2021) and RoseTTAFold (Baek et al., 2021), respectively, complex software that use multiple sequence alignments as input to extract evolutionary constraints of the structure, and then, the residues of the desired sequence are modeled based on these extracted evolutionary constraints plus physics and geometric-based ones, and the final structure is obtained. This breakthrough will allow the structural-based modeling of any enzyme, skipping the painful and challenging part of experimentally obtaining its 3D structure. Since the amount of data has grown during the last 25 years, machine learning is also starting to be applied in enzyme engineering (Mazurenko et al., 2019). Still, it has just begun to be employed, and it faces a big problem in comparability and consistency between data sets of different groups.

What makes computational approaches stand out over experimental ones is the possibility to design *de novo* enzymes to catalyze complex reactions, which are not known to exist in nature (some enzymes were discovered to be able to catalyze some of these reactions (Kim et al., 2011; Miao et al., 2017), but after the publication of the *de novo* designs). A pioneer in this subject is Baker's lab, where they created retro-aldolases (Jiang et al., 2008), Kemp eliminases (Röthlisberger et al., 2008), and Diels-Alderases (Siegel et al., 2010) with the help of their famous software, Rosetta (Rohl et al., 2004). Nonetheless, $\dfrac{k_{cat}}{K_M}$ values are modest compared to the most proficient naturally occurring enzymes, and they require directed evolution to reach higher values (Althoff et al., 2012; Khersonsky et al., 2012; Zanghellini, 2014).

1.3   Computational methodology

Although the articles presented in the thesis required experimental characterization of the computational analysis performed, this section will only focus on explaining the computational methodology, as it is the one that represents all the work I have done.

To model enzymes in the computer, we must keep in mind that they are macromolecules (50 kDa or 450 residues), having a massive number of variables and making it complex to simulate their behavior *in silico*. This vast number of variables comes from its conformational space (the number of conformations the enzyme can explore), leading to a well-known case of a combinatorial explosion in chemistry (Howard and Kollman, 1988). To picture this, Levinthal's paradox states that if a protein of 100 residues is considered and only three conformations of the backbone are applied, the amount of possible conformations is 3^{100}. This means that protein folding cannot be random and computational biochemists must sample proteins (and thus, enzymes) smartly.

Several methods have been created working at different levels of theory. Next, a summary of molecular modeling is given.

1.3.1 Molecular modeling

Molecular modeling is the name we give to all techniques that simulate (bio)chemical and physical systems *in silico*. These methods can be classified either if they use Newton's laws (classical physics), which we term molecular mechanics (MM), or if they use the Schrödinger equation (quantum physics), which is called quantum mechanics (QM). Both have their advantages and limitations, MM is more simple and, thus, less expensive computationally. Nevertheless, QM calculi are more accurate and allow a better estimation of the user's parameters of interest (Jensen, 2007c).

Molecular mechanics

MM simplifies the system by treating atoms as the smallest and indivisible units without considering subatomic particles. Atoms are connected by bonds, which are considered harmonic oscillators. This description of molecules with atoms and their bonds is known as the "ball-and-stick" model. Thus, the system's potential energy is parametrized according to the coordinates of the atoms, their types, and their local rearrangement (meaning bonds, angles, and dihedrals), which is commonly known as force fields (FF) (*Figure 1.5*).

$$E_{System} = E_{Bonding} + E_{Non-bonding} \tag{1.4}$$

Where $E_{Bonding}$ and $E_{Non-bonding}$ are comprised by:

$$E_{Bonding} = E_{Stretching} + E_{Bending} + E_{Torsions} \tag{1.5}$$

$$E_{Non-bonding} = E_{Van\ der\ Waals} + E_{Electrostatic} \tag{1.6}$$

The numerical values assigned to the constants of the energy terms, the functional form used, and the number of parameters and terms taken into account define and distinguish one FF from another (*Figure 1.5*). Also, these numerical values are extracted either from experimental or computational (from QM calculi) data. Thus, the accuracy of the energy calculation of the user's system will depend, in principle, on the applied FF; at a practical level, most FFs give similar results on standard systems. A handful of widely used FFs exist, including AMBER (Bayly et al., 1995), CHARMM (Brooks et al., 1983), OPLS (Jorgensen et al., 1983), and GROMOS (Berendsen et al., 1995). One of the other main advantages of using FFs is that they are transferable, meaning that the parameters of an amino acid are the same for the different positions of that residue in the protein system and different systems. Still, this limits the incorporation of polarization effects in a specific residue position, and bonds cannot be formed nor broken.
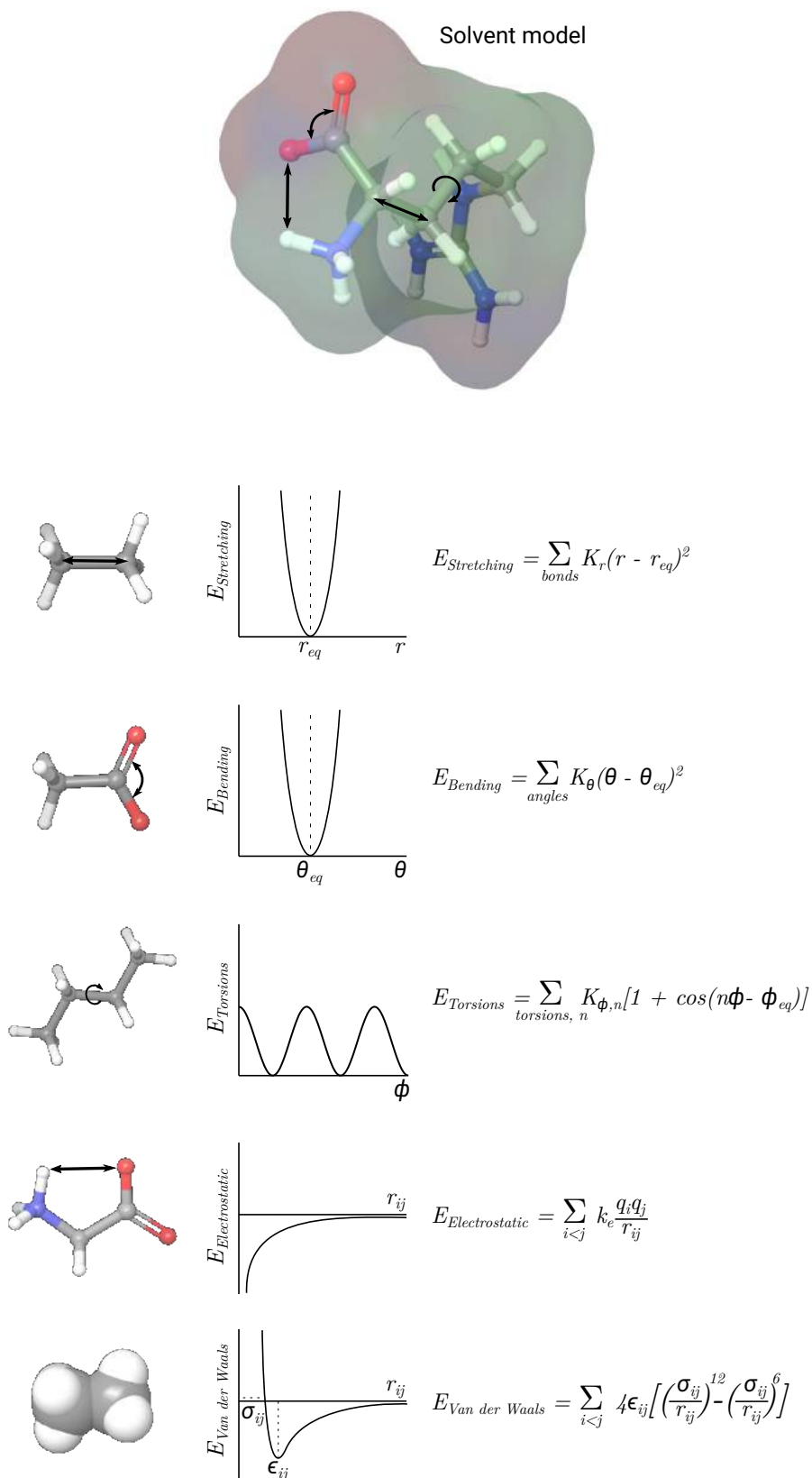
Solvent model

$$E_{Stretching} = \sum_{bonds} K_r(r - r_{eq})^2$$

$$E_{Bending} = \sum_{angles} K_\theta(\theta - \theta_{eq})^2$$

$$E_{Torsions} = \sum_{torsions,\ n} K_{\phi,n}[1 + cos(n\phi - \phi_{eq})]$$

$$E_{Electrostatic} = \sum_{i<j} k_e \frac{q_i q_j}{r_{ij}}$$

$$E_{Van\ der\ Waals} = \sum_{i<j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

FIGURE 1.5: Illustrative scheme of the different interactions that define the energy of the system in a classical FF.

Since enzymes typically work in aqueous media, water must be considered in the simulations. There are two main ways to model the solvent with explicit waters (meaning modeling every water molecule with the used FF) or implicit ones (meaning the solvent is represented as a continuum medium). The number of explicit water molecules required to model the solvent surrounding an enzyme is too large, making it unnecessarily computationally demanding. The solution for modeled explicit solvents is the usage of periodic boundary conditions (PBC), in which the system is confined in a box (called a unit cell), and it is repeated throughout the 3D space by rigid translations.

If one's system is big and the simulation time is large, implicit solvents can be used, saving computational resources and allowing more extended simulations. However, they cannot account for the H-bonds formed in the system (between the solvent and the protein). Implicit solvents consider electrostatic and van der Waals effects, as well as the free energy penalty, derived from creating a cavity in the solvent bulk (cavitation). Inside implicit solvation, we have three main methods (Cramer and Truhlar, 1999).

- SASA-based methods; no electrostatic effect is considered.

- Poisson-Boltzmann model; the exact electrostatic effect of the solute is considered, and it is the most precise one.

- Generalized Born model; the electrostatic effect is also considered but less accurate. They are an approximation to the exact Poisson-Boltzmann model, where the solute is regarded as a set of spheres with a particular dielectric constant.

Quantum mechanics

QM describes the system using subatomic particles, electrons, and nuclei. The Schrödinger equation must be solved to study these entities explicitly (compared to MM methods, where subatomic particles are considered implicitly). To obtain the energy and the wavefunction of the system, computational biochemists use the time-independent Schrödinger equation, because the protein is treated as a stationary system.

$$\hat{H}\Psi = (\hat{T}_N + \hat{T}_e + \hat{U}_{NN} + \hat{U}_{eN} + \hat{U}_{ee})\Psi = E\Psi \tag{1.7}$$

Where $\hat{H}\Psi$ is the Hamiltonian operator, $\hat{T}$ represents the kinetic part, $\hat{U}$ represents the potential part, e states for electrons, and N for nuclei. Thus, all interactions are taken into account, including between electrons (ee), nuclei (NN), and electron-nuclei (eN). Due to the high complexity of this equation, the Born-Oppenheimer approximation is used (Jensen, 2007d). This notion considers that the position of the nuclei is fixed compared to the electrons' positions because the much higher mass of the nuclei (3 orders of magnitude higher) makes their movement negligible compared to the electronic motion. This

approximation leads us to a simplified version of the equation, the electronic Schrödinger equation:

$$\hat{H}_e\Psi_i(R,r) = (\hat{T}_e + \hat{U}_{NN} + \hat{U}_{eN} + \hat{U}_{ee})\Psi_i(R,r) = E_i(R)\Psi_i(R,r) \tag{1.8}$$

R denoting nuclear positions and r electron coordinates. The analytical solution of this equation can only be obtained for a small list of systems, and they do not contain multiple electrons. Many-electron systems require a numerical answer and depending on the accuracy desired, several methods exist (Jensen, 2007d). The three main groups are:

- **Hartree Fock (HF)**: The method is based on the assumption that the exact N-body wavefunction of the system can be approximated by a single Slater determinant of mono-electronic wavefunctions or spin-orbitals (ϕ_i) (Jensen, 2007d). ϕ_i are built from the linear combination of atomic spin-orbitals χ, which are mono-electronic functions based on the atom nuclei (so, $\phi_i = \sum_{\alpha}^{M_{basis}} c_{\alpha i}\chi_\alpha$ where α denotes a specific spin quantum number). χ is a basis set, and the type of basis function can be any (exponential, Gaussian, polynomial, cube functions, wavelets ...). Depending on the basis set used (and the particular problem we are facing), the accuracy of the results can vary a lot (Jensen, 2007d). HF is an accurate method, but the electron-electron interactions are only considered as an average effect. The explicit treatment of the electron-electron interactions is neglected, and thus, the correlation energy is not properly estimated. To have a more accurate and exact solution, post-HF methods that consider correlation energy have been developed (including configuration interaction, coupled cluster, and the MØller-Plesset methods). Still, these post-HF methods are limited to only calculations with small systems (1000 electrons at best), since the computational effort increases by N^6.

- **Density Functional Theory**: This methodology is based on the Hohenberg-Kohn theorem, which states that the ground-state electronic energy is determined only by the electron density (ρ) (Jensen, 2007e). It is said that the integral of the electron density defines the number of electrons, the cusps in the electron density relate to the position of the nuclei, and the heights of the cusps define the corresponding nuclear charges (Steiner, 2004). The main advantage of this methodology is that the electron density depends only on three spatial coordinates and, thus, is independent of the number of electrons (compared to wavefunction methods that have $4N$ variables, three spatial and one spin coordinate for each electron). Still, the exact functional that connects the electron density with the ground state energy is unknown. To address this, the Khon-Sham theory expresses the electron density of the system as a linear combination of basis functions (similar to HF) and assumes electrons are not interacting to calculate the kinetic energy (and the remaining kinetic energy is transferred to an exchange-correlation term). The main advantage of DFT is that the computational cost is lower than in HF methods while not losing accuracy.

Also, calculations can be correlated with experiments, since electron density can be measured (for instance, X-ray diffraction techniques). Its main problem comes with the inability to systematically improve its results, compared to HF, where additional determinants can be added to bear in mind the correlation energy.

- **Semi-empirical methods**: These methods come from adding more approximations to HF methods, mainly neglecting all integrals involving three or more nuclei when constructing the Fock matrix (Jensen, 2007d). These integrals are transformed into parameters and fitted to experimental data. Likewise, the only electrons treated explicitly are the valence ones, and a minimal number of basis sets is used. Several semi-empirical methods are defined depending on how many integrals we neglect, the parametrization used, and the number of basis sets used for the valence electrons. The main advantage of this approach is reducing the computational cost compared to the other two methods, allowing for the simulation of more extensive systems (like biomolecules). Also, they can be handy when a lot of experimental data is available for the studied system. However, they perform poorly when systems with limited experimental data are simulated (like FF-based methods, although semi-empirical QM methods can describe bond-breaking and forming reactions).

Quantum mechanics / molecular mechanics

To study the catalytic mechanism of an enzyme, we need to simulate the system at the QM level. Still, the system's size is too big to study it all at that level of theory. Even though we could simulate only the part of the enzyme where the reaction occurs, the influence of the rest of the protein would not be taken into account, leading to wrong answers. To solve this problem, QM/MM methods were created (Warshel and Levitt, 1976; Field et al., 1990), where the part in which the reaction takes place is modeled at the QM level, while the remaining portion of the system (protein+solvent) is modeled at the MM level (*Figure 1.6*).

One of the main problems in QM/MM methods is representing the interaction between both regions (QM and MM). Thus, a simple way to think of the energy of the whole system would be:

$$E_{system} = E_{QM} + E_{MM} + E_{QM/MM} \tag{1.9}$$

The interface bonds must be dealt with when the MM region and the QM one are covalently bound. Three main methods exist; frozen orbitals (orbitals that cap the QM region and represent the cut bond), link atoms (usually a hydrogen atom, and it is used to saturate the valence of the frontier QM atom), or boundary atoms (the frontier MM atom appears in both QM and MM calculi). The covalent terms in the energy involving the border QM/MM atoms are treated classically, while in the electrostatic terms, frontier
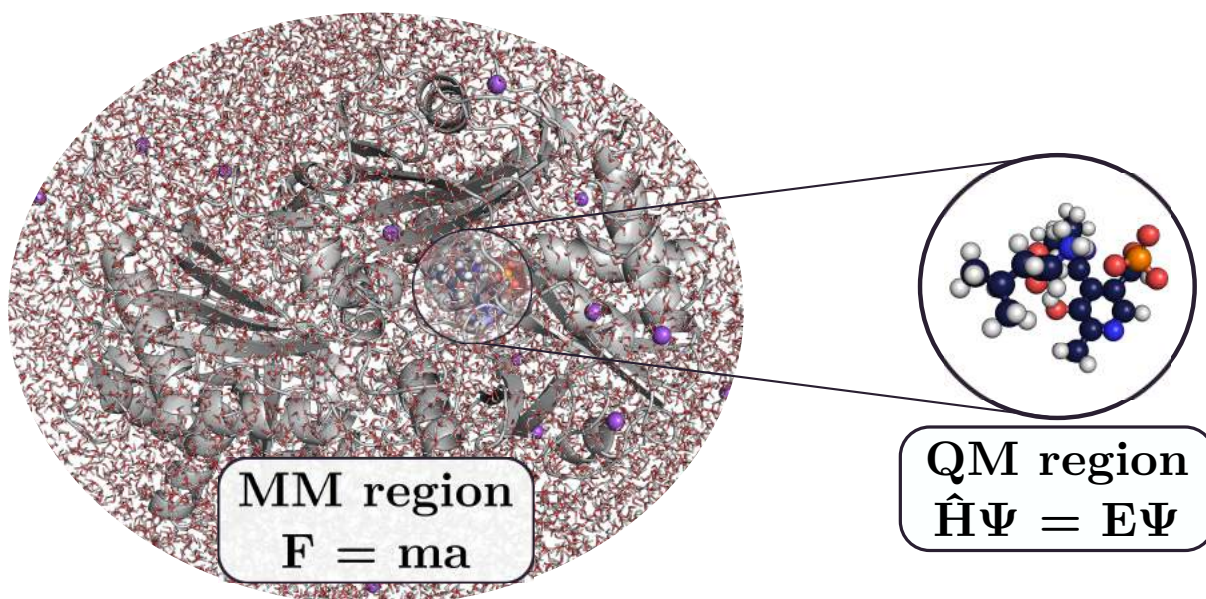
FIGURE 1.6: Illustrative scheme of a QM/MM calculation. The system is
an ω-transaminase (PDB ID: 5FR9).

MM atoms have their charges replaced by spherical gaussian charges (and redistributed) to avoid over-polarization of the charges in the frontier QM atoms.

The development of QM/MM methods was awarded the Nobel Prize in Chemistry in 2013 since they enabled the study of complex chemical systems *in silico*. This Nobel Prize could not have been conceived without another Nobel Prize in Chemistry, the one in 1998 given to the development of DFT and computational methods for quantum chemistry.

1.3.2   Optimization and simulation techniques

The main methodology to study enzymes and obtain the energy of the whole system at different levels of theory has been described. Likewise, the parameters or notions from enzymes that computational biochemists want to study have also been explained.

Still, I have not explained how to use molecular modeling to sample the mentioned conformational space of a protein smartly. The energy of all the conformations of a system represented as a function of its degrees of freedom is what is known as its potential energy surface (PES). In statistical mechanics, it is well-known that the conformations with the lowest energies (meaning local minima of the PES) are the more frequent and representative of the whole system. Thus, computational chemists tend to explore the PES of a system to find stationary points (points in the PES where the first derivative is zero) (Jensen, 2007a). The energy of a system can be optimized with different numerical methods and by iteratively slightly modifying its coordinates until a convergence criterion is reached. These optimization techniques are mainly steepest descent, conjugate gradient,

and Newton-Raphson methods. These methods will lead to a final geometry where the energy is lower than the initial one and depending on if we obtain imaginary frequencies in the Hessian matrix diagonalization or not, a minimum or a saddle point (transition state) will have been reached (Jensen, 2007a).

However, not all initial conformations of the system are equally suitable for optimization. Moreover, several local minima of a system might be studied or even less represented conformations of the PES as well. Two main methods exist to achieve a representative sampling of the PES space: Monte Carlo (MC) and molecular dynamics (MD) (Jensen, 2007b).

Molecular dynamics

Molecular dynamics simulate the behavior of a system and its coordinates over time by applying Newton's equations of motion (Jensen, 2007a,b). Since the force can be expressed as the gradient of the potential energy ($\dfrac{dV}{dr_i}$), the forces of the system can be obtained for each particle in the system (with its specific mass and coordinates in the space; m_i and r_i) by evaluating the PES along the time axis:

$$a_i(t) = \frac{F_i}{m_i} = \frac{-1}{m_i}\frac{dV}{dr_i} \tag{1.10}$$

The forces are conservative, meaning they do not depend on the taken path, but only the V and r_i. Therefore, an MD simulation starts with an initial conformation of the system (with the positions of all its particles), and a set of initial velocities for each atom in the system are sampled using the Boltzmann distribution at the user's defined temperature. Then, we update the positions and velocities of all atoms in the system based on the calculated accelerations after each time step. The time step used to integrate Newton's laws of motion must be much smaller (an order of magnitude) than the time required for the system to execute the fastest movement. The frequency of vibration of hydrogens (the lightest particles in an MM-based simulation) has a typical value of around $\sim 10^{-14}s^{-1}$, meaning that the time step should be around 1 (where H atom vibrations are considered) or 2 (where H bonds are fixed) fs to consider the system classically (Jensen, 2007b).

New positions and velocities of all atoms in the system are obtained after every time step until the desired simulation time is reached. The concatenation of all the coordinates in every time step is known as the system's trajectory. A basic scheme of an MD simulation is depicted in *Figure 1.8*.

Despite the explained fundamental functioning of an MD simulation, several parameters and concepts must be included to simulate a protein system properly. The system must be prepared and adequately equilibrated before running an MD simulation. This preparation means solvating the protein structure (as they do not exist in the vacuum or gas
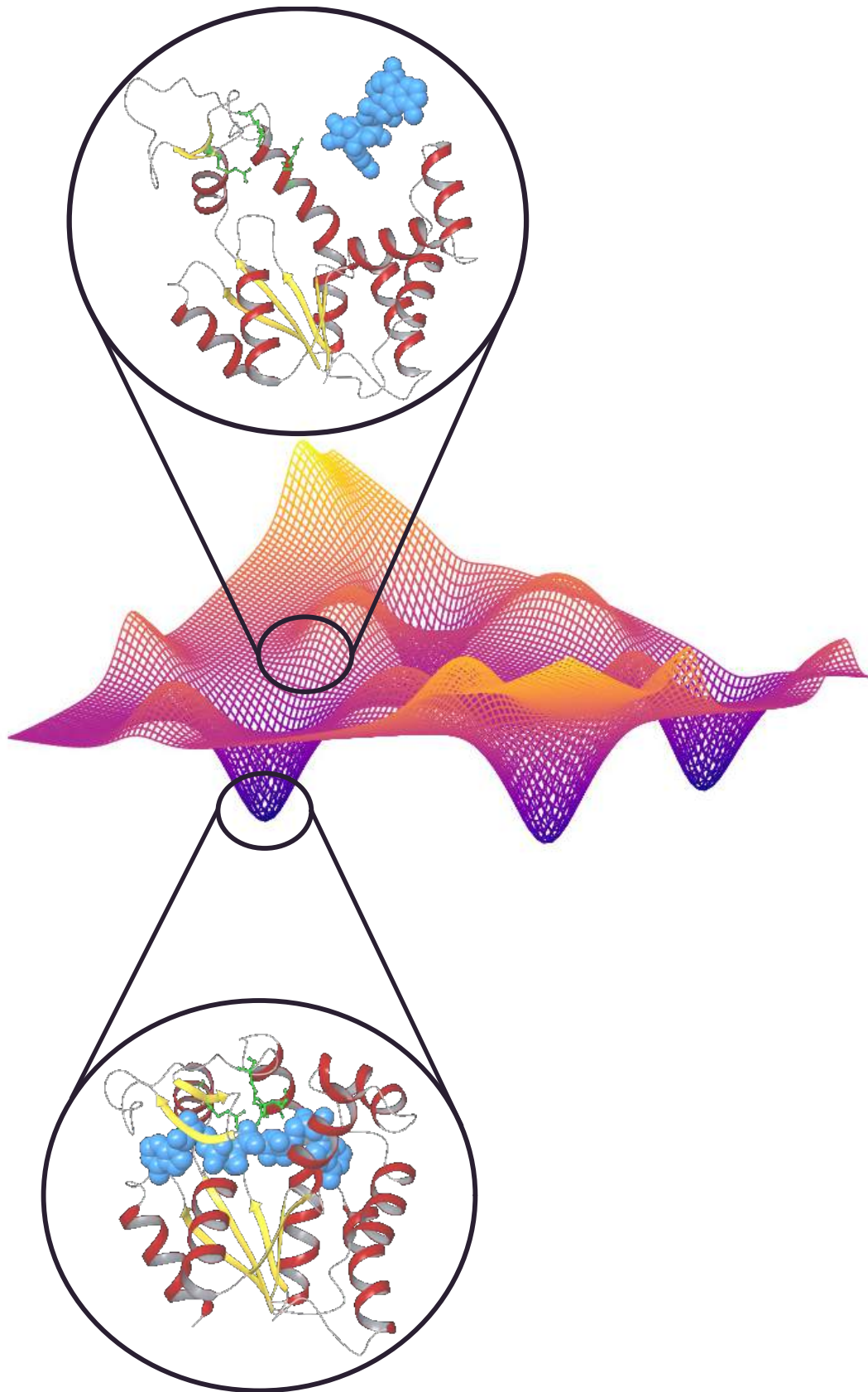
FIGURE 1.7: Illustrative scheme representing the PES of an enzyme-substrate complex where the bound state represents the local minimum. Both conformations of the adenylate kinase are extracted from the PBD codes: 2RH5 and 2RGX (with the bound substrate analog).
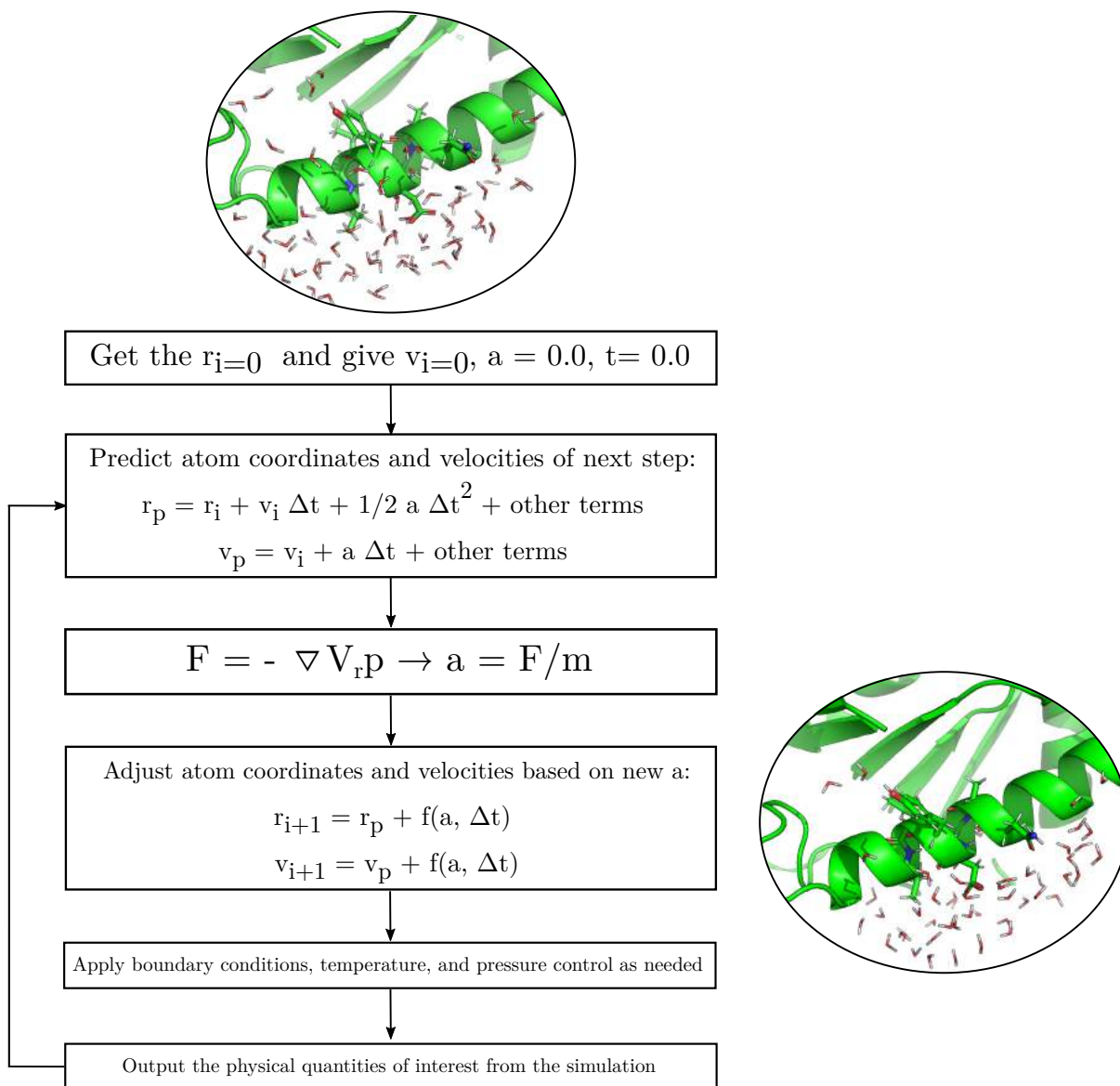
Get the $r_{i=0}$ and give $v_{i=0}$, a = 0.0, t= 0.0

Predict atom coordinates and velocities of next step:

$r_p = r_i + v_i \, \Delta t + 1/2 \, a \, \Delta t^2$ + other terms

$v_p = v_i + a \, \Delta t$ + other terms

$$F = - \nabla V_r p \rightarrow a = F/m$$

Adjust atom coordinates and velocities based on new a:

$r_{i+1} = r_p + f(a, \Delta t)$

$v_{i+1} = v_p + f(a, \Delta t)$

Apply boundary conditions, temperature, and pressure control as needed

Output the physical quantities of interest from the simulation

FIGURE 1.8: General scheme of an MD simulation.

phase), defining a specific thermodynamic ensemble (NVT, with constant volume and a thermostat, or NPT, with both a barostat and a thermostat), laying thermodynamic and spatial boundary conditions, and deciding the integration algorithm (typically the Verlet algorithm).

Although usually the energy of the whole system is obtained with a certain FF (AMBER, CHARMM, OPLS, GROMOS...), MD simulations can calculate a section of the system by means of QM and the Schrödinger equation (what is known as *ab initio* MD), allowing to study chemical reactions of enzymes over time.

Nowadays, the usage of graphical processing units (GPUs) to accelerate the simulation of all-atom MD simulations allows running 0.5 $\frac{\mu s}{day}$ in a protein of ~ 300 residues. Thus, we can currently study side chain rotations, loop motions, ligand/substrate binding, and some domain folding. If we want to see phenomenons that occur at the micro and milliseconds timescale (like folding/unfolding, allosterism, or catalysis), more ingenious ways to sample the system (Lane et al., 2013) or finer hardware (Shaw et al., 2008, 2014) are needed.

Monte Carlo

MC-based methods are another main alternative besides MD simulations to smartly explore the system's PES. (Jensen, 2007b). MC algorithms generate the new conformations based on stochastic perturbations of the system instead of using Newton's laws of motion. Hence, this heuristic approach for exploring the conformational space of a system is time-independent. The performed perturbation is accepted or rejected based on the final energy of the system compared to the energy before the perturbation. The generated conformation is always accepted if $\Delta E < 0$. Otherwise, the Boltzmann factor is used ($e^{\frac{-\Delta E}{k_B T}}$), and its value is compared to a random number ($0 \leq R \leq 1$). If $e^{\frac{-\Delta E}{k_B T}} > R$, the new geometry is accepted. Else, the new geometry is rejected, and a new perturbation is tried from the initial configuration. This manner of deciding whether to accept or reject the executed perturbation in the system, called the Metropolis criterion, is summarized in the following equation:

$$P_{i \to j} = \begin{cases} 1 & \Delta E_{i \to j} < 0 \\ e^{\frac{-\Delta E_{i \to j}}{k_B T}} > R & \Delta E_{i \to j} > 0 \end{cases} \tag{1.11}$$

The differences between MC algorithms reside in how the perturbation is done. Compared to MD simulations, these methods require less computational resources, but they lack the perspective of time. Moreover, the random perturbation of MC methods hinders the study of correlated motions, meaning that exploring whole proteins in an explicit solvent is inefficient since multiple perturbations must synchronize to generate good conformations. Thus, MC methods are handy for studying relatively small molecules' conformational space around protein systems (Jensen, 2007b; Gilabert et al., 2018; Jorgensen and Tirado-Rives, 2005).

Molecular docking

Docking techniques aim to find the preferred binding orientation between the ligand (either small molecule or protein) and the receptor (protein). In this case, the methodology aims at quickly finding the most stable conformation of a ligand-receptor pair rather than exploring the PES of the system. Typically, docking methods provide multiple ligand

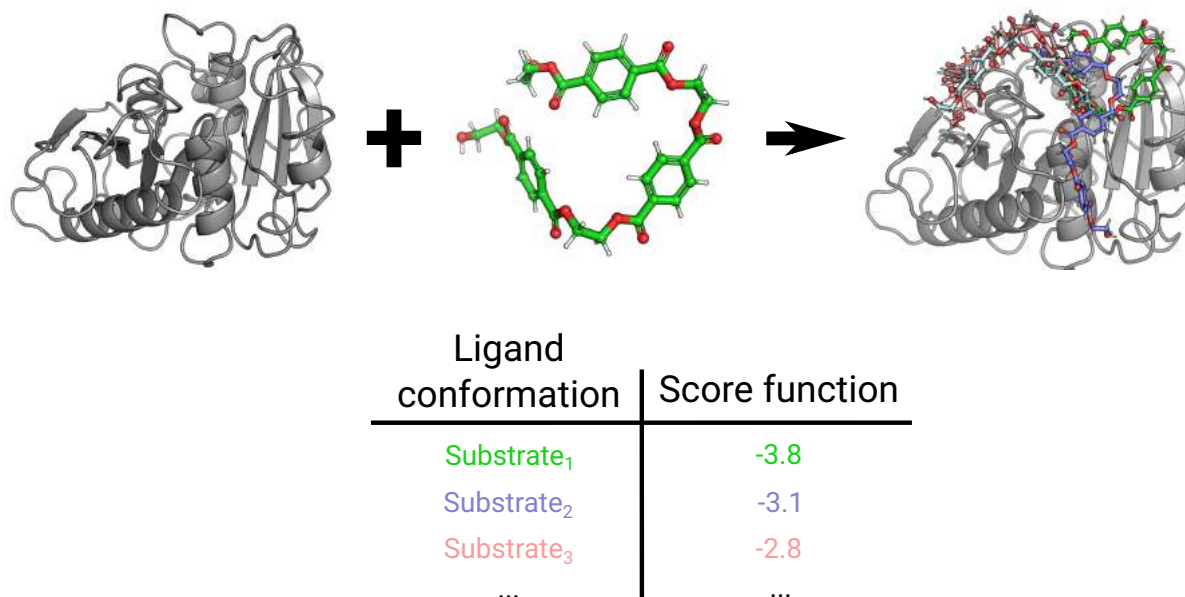| Ligand conformation	Score function
Substrate$_1$	-3.8
Substrate$_2$	-3.1
Substrate$_3$	-2.8
...	...

FIGURE 1.9: Illustrative representation of molecular docking. Some conformations of the ligand inside the protein cavity are shown and tagged with a specific score in the table.

conformations that are later ranked by a scoring function (*Figure 1.9*). This scoring function can either be knowledge-based, empirical-based, or physics-based (where FFs are used). These methods are extremely fast compared to MC and MD simulations. Thus, one can evaluate thousands of ligands in the same receptor, allowing structural-based virtual screening of compounds for the pharmaceutical industry (Sousa et al., 2013; Kitchen et al., 2004). The problem with docking is the requirement of defining a grid box to explore around a protein cavity (meaning the active site must be known). Likewise, docking methods tend to use a unique rigid protein conformation, which can be inconvenient when the receptor is not in the bound conformation. Ensemble docking tries to address this issue by docking the ligand against different conformations of the receptor (probably obtained from an MD simulation).

Protein Energy Landscape Exploration

Protein Energy Landscape Exploration (PELE) is an MC-based algorithm developed in our laboratory that aims to sample the protein-ligand conformational space, exceeding the limitation of classical MC methods mentioned earlier (Kenneth W. Borrelli et al., 2005; Municoy et al., 2020). This heuristic algorithm combines a random perturbation of the ligand (translations and rotations) with protein structure prediction techniques, considering the movement of both entities. The software starts by sampling the different microstates of the ligand through small rotations and translations. Then, the protein's flexibility is also contemplated by applying low-frequency normal modes through the

anisotropic network model (ANM) approach. The protein is depicted as a network of elastic forces connected by the alpha carbons of its residues. The obtained modes/vectors allow the protein's backbone to move (Atilgan et al., 2001). After perturbing the whole system, side chains of the residues close to the ligand are sampled to evade steric clashes. Finally, the entire system is minimized with a truncated Newton minimization (by using the OPLS2005 force field (Banks et al., 2005) and a generalized Born implicit solvent model), and the step is accepted or rejected based on the previously mentioned Metropolis criterion. The explained protocol is illustrated in *Figure 1.10*.

The sampling of PELE can be enhanced with a clustering/spawning technique called Adaptive-PELE, based on an adaptive reinforcement learning procedure (Lecina et al., 2017). Adaptive-PELE consists of concatenated short PELE simulations, named epochs, where the less explored regions (smaller clusters) have more sampling in the next epoch. These regions are defined based on a clustering algorithm and a specific metric.

Albeit PELE was intended for mapping the ligand diffusion into the active site (Kenneth W. Borrelli et al., 2005), it has been employed for many other purposes, such as global exploration for binding sites of a specific substrate in a protein (Diáz et al., 2020), fragment growing onto a scaffold while exploring the protein-ligand energy landscape (for hit-to-lead drug design) (Perez et al., 2020), estimation of absolute binding free energies through the usage of Markov state models (Takahashi et al., 2014; Gilabert et al., 2019), or discovery of the location of buried waters (hydration sites) in proteins (Municoy et al., 2020).
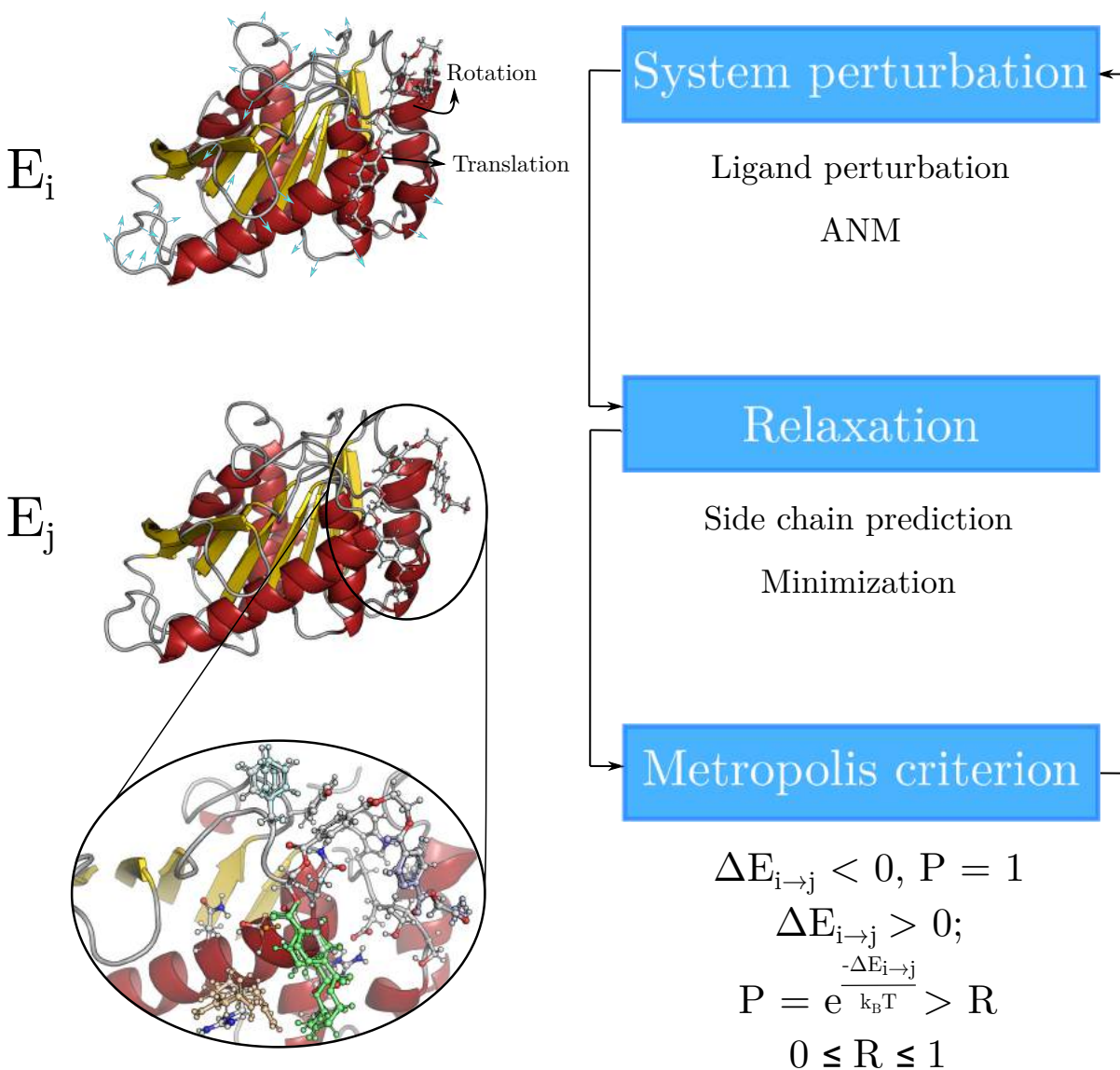
FIGURE 1.10: Illustrative scheme of a step from the PELE algorithm. A step is comprised of the system perturbation, its later relaxation, and acceptance or rejection according to the Metropolis criterion.

1.4   Systems of study

This section presents a summary of the studied enzymes in this thesis, mainly structural information, catalytic mechanism, the importance of these types of enzymes and their applications. The studied biocatalysts are serine hydrolases and ω-transaminases, which are a big market in the food, detergent, and pharmaceutical industry (Jemli et al., 2014; Kelly et al., 2018; Wu et al., 2021).

1.4.1   Serine hydrolases

Serine hydrolases are part of the big family of hydrolases (Enzyme Commission [EC] number 3). This family of enzymes catalyze the cleavage of chemical bonds by using water (hydrolysis) (Nelson and Cox, 2008b), and the difference between hydrolases resides in the nature of the broken chemical bond. Serine hydrolases include esterases and proteases, being one of the biggest (they represent 1 % of the human proteome and are abundant in other eukaryotic and prokaryotic proteomes (Simon and Cravatt, 2010)) and most important subfamily of hydrolases.

Their general catalytic mechanism can be separated into two steps. First, the electrophilic carbon in the reactive bond/group is nucleophilically attacked by the catalytic serine residue. The nucleophilic O_γ of the catalytic serine residue is activated by a charge-relay system comprised of a histidine residue (that acts as a base) and an acid residue (aspartate or glutamate). These three residues are named catalytic triad, the key elements of a serine hydrolase. The histidine residue extracts a proton of the serine residue in this first step of the reaction. The acid residue stabilizes the position of the histidine residue and polarizes it by stabilizing the positive charge. After the nucleophilic attack, a covalent bond between the enzyme and the substrate is formed, known as the first tetrahedral intermediate. The carbonyl oxygen of the substrate gets a negative charge (oxyanion) in this stage, which is stabilized by either two amino groups of the backbone amide bonds of two residues or by positively charged residues. This preorganized environment to stabilize the mentioned negative charge is named oxyanion hole. The tetrahedral intermediate is subsequently disrupted by recovering the carbonyl double bond in the substrate and ejecting part of it as the first product, activated by the donation of a proton by the protonated histidine residue. The first part of the reaction ends with the remaining part of the substrate covalently bound to the serine residue (known as the acyl-enzyme intermediate) (Nelson and Cox, 2008e; Dodson and Wlodawer, 1998; Rauwerdink and Kazlauskas, 2015).

The second step of the reactions procceds with the nucleophilic attack of a water molecule to the acyl-enzyme intermediate. The O of the water molecule is activated by the same histidine residue. The attack results in the formation of second tetrahedral intermediate, followed by the release of the second product of the reaction. A general scheme of the
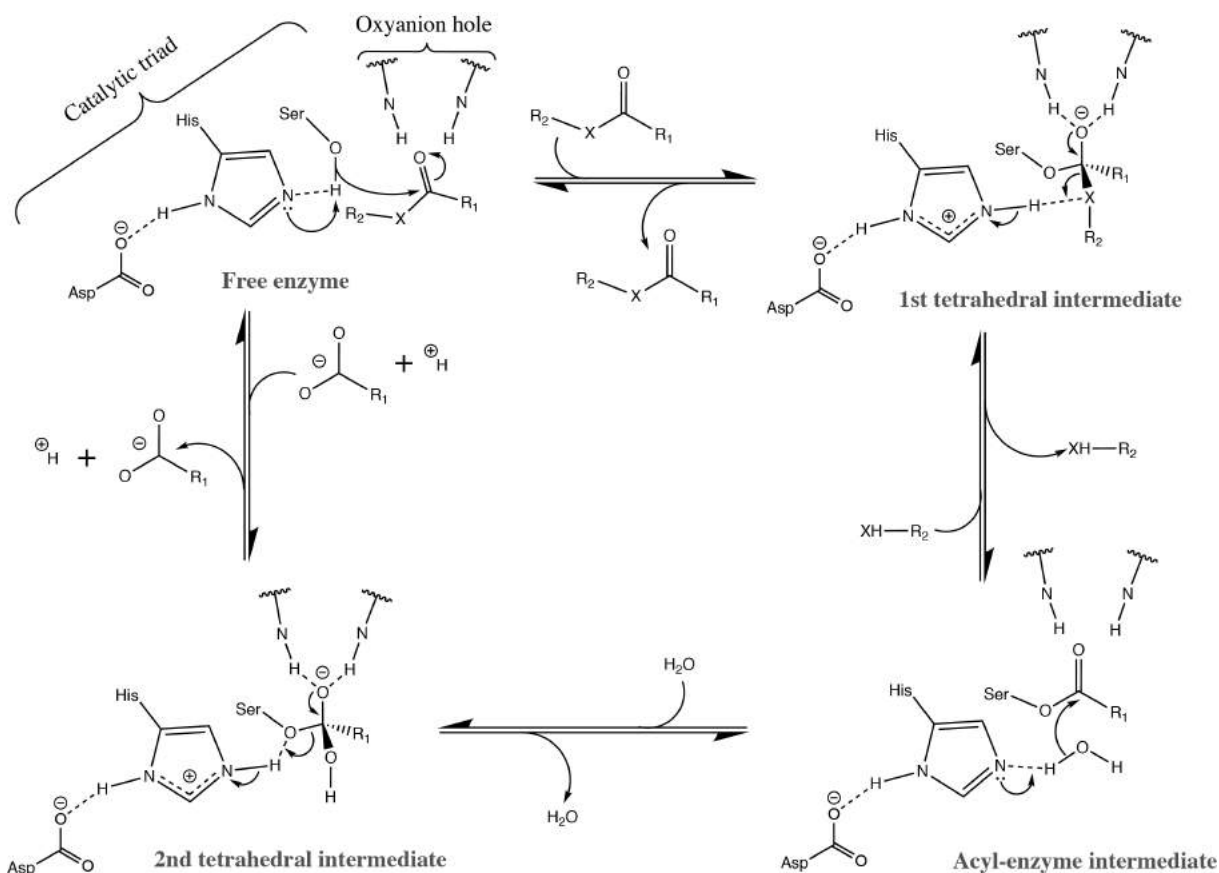
FIGURE 1.11: General scheme of the catalytic mechanism of serine hydro-
lases. The catalytic triad and oxanion hole are tagged with curly brackets.
The X can be -NH in amides and O in esters. The scheme was drawn with
ChemDraw.

catalytic mechanism can be found (*Figure 1.11*).

Regarding their structural features, the vast majority of serine hydrolases share a classical α/β fold. The fold can be described as parallel β-sheets (usually eight) that are inter-connected with α-helices (usually 6). Both the catalytic triad and the oxanion hole are frequently located on the loops between β-sheets and α-helices (Ollis et al., 1992; Lenfant et al., 2013) (*Figure 1.12*). Likewise, a common and important feature found in serine hydrolases is the "nucleophilic elbow", a sequence motif around the nucleophilic serine residue: Gly-X-Ser-X-Gly (X referring to any of the 21 common amino acids) (Nardini and Dijkstra, 1999).

Applications

Since serine hydrolases include lipases, esterases, and proteases, they are one of the most used enzyme families (estimated to cover around 75 % of the commercially used enzymes) in the industry (Prakash et al., 2013; Jemli et al., 2014). Hydrolases (mainly proteases)
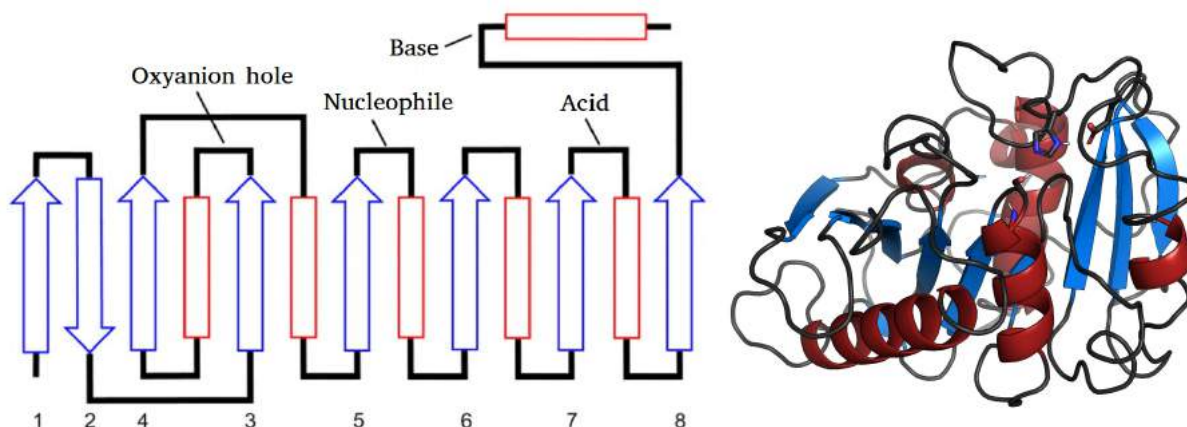
FIGURE 1.12: Illustration of the α/β hydrolase fold represented with the 6 α-helices (red rectangles) and 8 β-sheets (blue arrows) (right). The location of the three catalytic residues and the oxyanion hole are displayed in both the scheme and the 3D representation.

have been extensively used in the detergent industry since the availability of enzyme production from microbial origins in the 1960s (Maurer, 2004), allowing to remove difficult stains from different sources at milder temperatures (affecting less the properties of the tissues' clothes). Subtilisins (alkaline proteases from *Bacillus* species) are the main proteases used in the detergent industry (Maurer, 2004; Jemli et al., 2014) due to their high stability, high substrate promiscuity, and the fact that the enzyme is secreted outside of the cell (simplifying the separation of the enzymes from the bacteria/cells). However, there is an interest towards finding psychrophilic enzymes that could work at low temperatures (10-20 °C) with the same advantages of subtilisins to have environmental benefits by saving energy (Maurer, 2004; Jemli et al., 2014).

The food processing industry has a long story with the usage of enzymes, dating back to their earliest applications in 6000 BC with beer production, bread baking, cheese making, and other fermentation processes (Jemli et al., 2014). Nowadays, the food processing industry is one of the biggest in the market and enzymes are widely used in this area. Esterases and lipases are essential in producing flavours due to their specificity toward fats and oils (Casas-Godoy et al., 2012; Jemli et al., 2014), either with hydrolysis of the ester bonds or by catalyzing the transesterification of desired esters. They can also be applied in the juice, baking, and brewing industries (Casas-Godoy et al., 2012; Jemli et al., 2014).

The pulp and paper, textile, and leather industries also employ hydrolases for pitch control in pulping processes (Kontkanen et al., 2004), refinement of cotton/polyester fibers for final product quality (Degani et al., 2002; Vertommen et al., 2005), dehairing of animal skins to obtain leather (Dayanandan et al., 2003). Serine hydrolases can be used for the chiral synthesis of drugs in the pharmaceutical industry by enantioselective hydrolysis of racemic mixtures of esters (Wu et al., 2021). A relevant example includes the use of a mutant CalB (I189K) to efficiently synthesize the chiral intermediate of moxifloxacin (Shen

et al., 2018), an antibiotic employed to treat penumonias, conjunctivitis, endocarditis, tuberculosis, and sinusitis (Hoogkamp-Korstanje and Roelofs-Willemse, 2000).

Recently, the value of hydrolases in the market and society has dramatically increased after the discovery of a microorganism, named *Ideonella sakaiensis* that feeds from polyethylene terephthalate (PET) (Yoshida et al., 2016). *Ideonella sakaiensis* contains two serine hydrolases (called PETase and MHETase), one extracellular involved in degrading the polymer into the monomer (mono-(2-hydroxyethyl)terephthalate or MHET) and the other is anchored to the membrane, breaking the monomer into smaller building blocks that the bacteria can use as carbon source. After this finding, several studies have tried to create engineered versions of the PETase (Austin et al., 2018; Liu et al., 2018; Son et al., 2019; Cui et al., 2021; Lu et al., 2022) or other cutinases (Tournier et al., 2020) that can act against PET as well. Moreover, other PET hydrolyzing enzymes have been discovered from other sources and with different optimal temperatures, turnover rates, and thermostabilities (Buchholz et al., 2022). Thus, these enzymes could be substantially valuable for PET recycling and a circular economy of PET (Rosenboom et al., 2022).

1.4.2 ω-transaminases

Transaminases (EC 2.6.1) are enzymes that catalyze the transfer of an amine group from an amine donor to an acceptor (frequently being a ketone or aldehyde) (Ghislieri and Turner, 2014; Kelly et al., 2018). Transaminases (TAs) are a big family of enzymes, and they are classified according to the amine group's relative position to the substrate's carboxyl moiety. α-TAs have the amine group bound to the same C atom as the carboxyl group, whereas ω-TAs can have the amine group in any position relative to the carboxyl group. Thus, ω-TAs are more suited for industrial processes since they can accept a broader range of substrates.

TAs require a pyridoxal 5'-phosphate (PLP) cofactor to catalyze the reaction. PLP is covalently bound to the protein forming a Schiff base (internal aldimine) with the ϵ-amino group of the catalytic lysine residue (Ghislieri and Turner, 2014; Kelly et al., 2018). The reaction starts with a nucleophilic addition of the amine group in the substrate to the PLP cofactor, breaking the internal aldimine and forming one with the substrate. The deprotonated ϵ-amino group of the catalytic lysine residue extracts a proton from the C next to the bound amine group, activated by the electron sink nature of PLP. The extracted electrons by PLP are returned, allowing the C=C next to the pyridine ring to deprotonate the catalytic lysine residue again. Finally, an activated water molecule, through deprotonation by the catalytic lysine residue, nucleophilically attacks the C=N bond, releasing the initial substrate in its corresponding keto acid and producing pyridoxamine-5'-phosphate (PMP) (Eliot and Kirsch, 2004; Malik et al., 2012). This first part of the reaction is known as oxidative deamination of the amine donor.

The second part of the reaction, called reductive amination of the amine acceptor, continues with the amine group of PMP nucleophilically attacking the carbonyl carbon of the amine acceptor (second substrate), forming a Schiff base and releasing a water molecule (deprotonating the catalytic lysine residue in the process). The catalytic lysine residue deprotonates the C atom next to the N of the newly formed Schiff base, activated again by the mentioned electron sink nature of PLP. The N+=C bond uses this pair of extracted electrons to deprotonate the catalytic lysine residue again. Lastly, the deprotonated catalytic lysine residue binds again to PLP, resulting in the release of the amine acceptor and the regeneration of the PLP cofactor. A scheme depicting the explained mechanism is shown in *Figure 1.13*. The electron sink nature of PLP is enhanced in the active site by a close interaction of the nitrogen in the pyridine ring with an acid residue (Eliot and Kirsch, 2004).

The structures of TAs are comprised of two major domains, a small (S) one and a large (L) one. The L domain has a Rossmann-like fold (with a seven-stranded parallel β sheet). The S domain can be divided into the N-terminal and C-terminal regions. As mentioned earlier, the structure contains a catalytic lysine residue that binds to the PLP cofactor, and the protein is a functional dimer (*Figure 1.13*).

Applications

Even though TAs are not as widely spread as serine hydrolases, they are a crucial family of enzymes in the pharmaceutical industry. It is estimated that 40 % of pharmaceutical drugs contain a chiral amino group within their structure (Ghislieri and Turner, 2014; Kelly et al., 2018). Chirality is critical in a drug, going from treatment to severe adverse effects on someone's health, depending on the purity of the right enantiomer. Thus, the synthesis of an enantiopure drug is a must. Enzymes come in handy as they are catalysts with high enantioselectivities, compared to inorganic chemical catalysts. Thus, TAs are key for synthesizing a lot drugs.
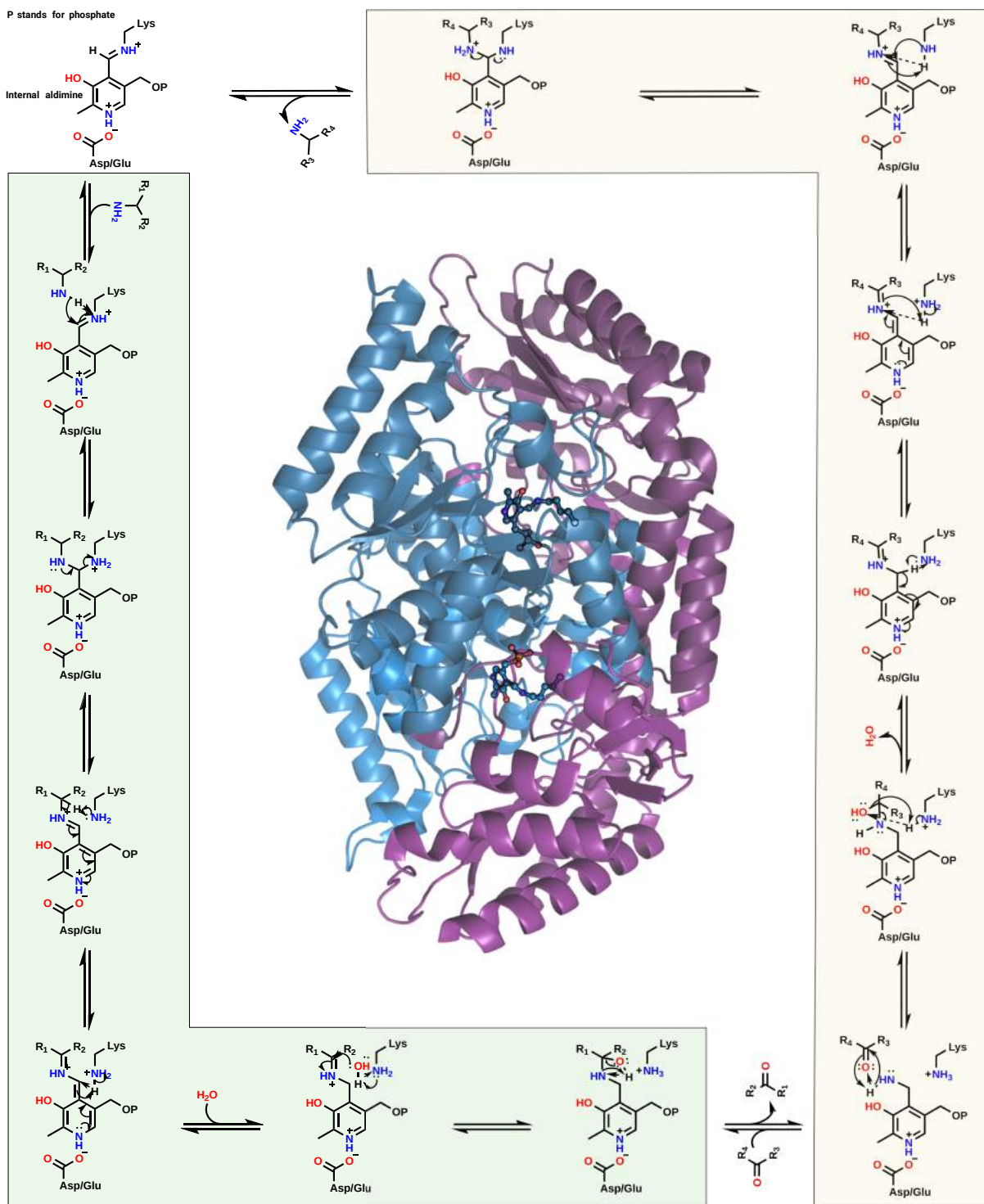
FIGURE 1.13: General scheme of the catalytic mechanism of TAs and their archetypical fold. The oxidative deamination is highlighted with a green frame, while the reductive amination with a yellow one. The Rs are the other possible chemical groups in the substrates. The catalytic Lys residue bound to the PLP cofactor is displayed. The L domain is stained in blue, while the S domain is shown in purple (PDB ID: 4E3R). The scheme of the catalytic mechanism was drawn with ChemDraw.
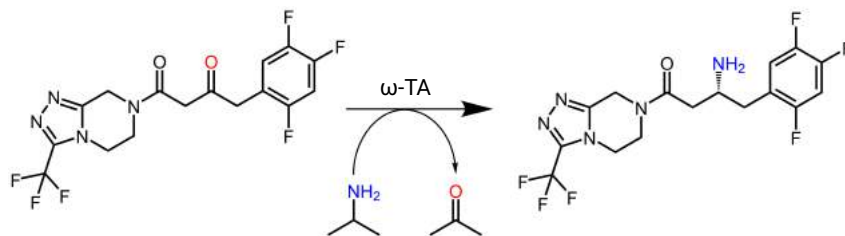
FIGURE 1.14: Asymmetric synthesis of sitagliptin by an evolved (R)-selective ω-TA with isopropylamine as amine donor.

The synthesis of chiral amines by ω-TAs can be accomplished via kinetic resolution of a racemic mixture or asymmetric synthesis from the corresponding prochiral ketone (Bezborodov and Zagustina, 2016). The kinetic resolution works by using the undesired enantiomer as the amine donor in the transaminase. One successful example of using an ω-TA for synthesizing a vital drug refers to an evolved (R)-selective ω-TA to bind prositagliptin and synthesize sitagliptin (Savile et al., 2010) (*Figure 1.14*). Sitagliptin is one of the 15-20 best-selling drugs (with revenue of 5.288 billion USD in 2021). Replacing the former rhodium-catalyzed asymmetric enamine hydrogenation with the transaminase-catalyzed amination for its synthesis allowed a 10-13 % improvement in overall yield, 53 % increase in productivity (as the enzyme can work under milder conditions), and 19 % reduction in total waste (Savile et al., 2010; Kelly et al., 2018). This improvement shows the urgent need for transaminases in the pharmaceutical industry for a more sustainable future due to the enormous amounts of drugs manufactured per day.

1.5 PluriZymes

The one-pot cascade reactions that allow the synthesis of compounds of industrial interest are really compelling (Devine et al., 2018; Sperl and Sieber, 2018; Huffman et al., 2019; Wu et al., 2021; Bell et al., 2021). Enzymes are suited for biocatalytic cascade reactions, since they are much more regioselective and stereoselective compared to inorganic catalysts and can work under milder conditions. Plus, enzymes can be further engineered, as we explained, to further improve e.g., enzyme activity for industrial production and compete with current chemical processes to synthesize the same compound (McIntosh and Owens, 2021; Nazor et al., 2021). For instance, Merck & Co. engineered (evolved) 5 enzymes of a designed nine-enzyme cascade process to efficiently synthesize islatravir, a potential drug for HIV treatment (Huffman et al., 2019).

Our lab computationally designed a pluriZyme, an enzyme with added active site(s), enabling it to catalyze different types of chemical reactions in the same protein scaffold (Santiago et al., 2018; Alonso et al., 2019; Roda et al., 2021). The first pluriZyme consisted of a serine ester hydrolase, where a second catalytic triad was added. Although they performed the same chemistry, one of the active sites (the WT one) was converted into a metal-complex chemocatalytic site by binding a suicide inhibitor that could coordi-

nate a transition-metal ion (Alonso et al., 2019). This incorporation allowed the designed pluriZyme to perform hydrolysis, as well as oxidation and Friedel–Crafts alkylation reactions (*Figure 1.15*A). Thus, pluriZymes could be employed to do some one-pot cascade reactions, saving the amount of expressed protein of a particular enzyme cascade reaction and more likely to be more efficient.

In spite of the fact that other studies related to pluriZyme-like systems have been published, the approaches and the designed catalyst were different, giving several strategies to tackle the desired cascade process (Roda et al., 2021). One of them consisted in the self-assembly of histidine-tyrosine peptides, which had esterase activity, were electrochemically active, and could polymerize polypyrrole (Díaz-Caballero et al., 2020) (*Figure 1.15*B). Another group used a non-catalytic protein scaffold with a big hydrophobic cleft in its dimer interface (Lactococcal multidrug resistance Regulator, LmrR) to add two abiological active sites (Zhou and Roelfes, 2020). First, they bound a Cu(II) complex in the dimer interface, and then, they introduced a non-canonical amino acid (p-aminophenylalanine). These two changes allowed the generated catalyst to perform Michael additions where the added non-canonical amino acid activates an enal through iminium ion formation, and the Cu(II) complex activates the nucleophile by enolization and binds it to the preferred prochiral face of the activated enal (*Figure 1.15*C).

Another approach aims at designing pluriZymes by coupling small metal nanoparticles with an enzyme, combining their catalytic properties (Filice et al., 2013, 2015; Benavente et al., 2018; Palomo, 2019; Benavente et al., 2020). An example includes the combination of the lipase B from *Candida antarctica* (CalB) with Pd nanoparticles that enabled both the one-pot cascade conversion of p-nitrophenyl propionate to p-aminophenol in water at room temperature, as well as the production of N-[(1R)-phenylethyl]acetamide coming from a racemic mixture of 1-phenylethylamine in toluene at 70°C (*Figure 1.15*D). Thus, this outlook of enzyme-metal hybrids allows one-pot cascade reactions in truly different conditions. Lastly, a recent publication showed an original way to replace the native NADH cofactor of an enzyme for a new molecule that emulates both the native cofactor and the flavin analog, allowing the direct proton and electron transfer between cofactors without being consumed nor affected by diffusion (Zhao et al., 2020). They took a formate dehydrogenase and replaced the NADH cofactor with Zn-ZPA, the complex containing a dihydropyridine amido motif, which mimics the NADH. This introduced complex has the space to confine both the substrate and riboflavin tetraacetate (RFT), the catalyst of the monooxygenation reaction. The designed catalyst–enzyme coupled system enhanced the Baeyer–Villiger oxidation of substituted 3-phenylcyclobutanones and substituted thioanisoles (*Figure 1.15*E).

The recent development of pluriZymes and their potential motivated me to pursue a PhD in Guallar's lab and push the concept forward with new ideas and some hypotheses to test in mind.
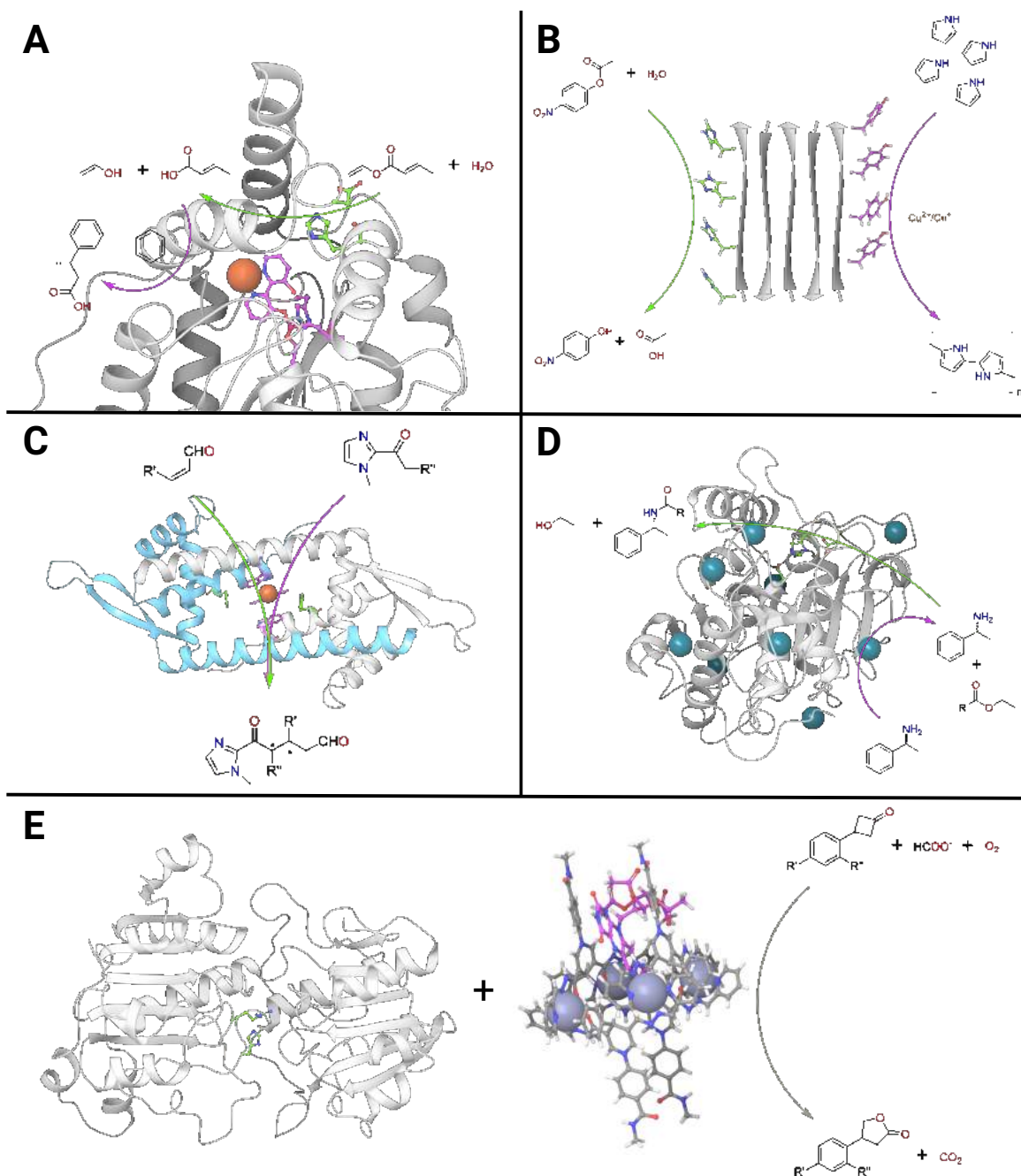
FIGURE 1.15: Current designed pluriZymes accompanied by an example of the catalyzed cascade reaction. A) Designed chemo-biocatalyst hydrolyzing vinyl crotonate followed by the Friedel-Crafts alkylation of the acid product of the previous reaction. B) Amyloid-like fibrils of histidine-tyrosine peptides hydrolyzing p-nitrophenyl acetate and polymerizing pyrrole. C) Modified LmrR protein can perform Michael additions. D) Enzyme-metal hybrid based on CalB and Pd-nanoparticles performing the dynamic kinetic resolution of 1-phenylethylamine using the transesterification with ethyl esters. E) Designed catalyst–enzyme coupled system improved the monooxygenation of substituted 3-phenylcyclobutanones. Protein structures are shown with the ribbon representation. The residues or molecules in the active sites have the C atoms stained in green and purple, respectively (except from the metal ions, which use the color of the element). Likewise, the reaction arrows are colored according to the active site catalyzing the reaction. All 3D representations of residues or molecules are displayed with the ball-and-stick model (except from the metal ions, for which the CPK model is used). The PDB codes are 6RKY, 6I8N, 4K6G, and 2NAC. The RFT molecule was extracted from PubChem with compound identification number: 9915369. The Zn-ZPA complex was obtained from the CCDC database with the number: 1920570.

2 | Objectives

The purpose behind this thesis has been to further advance the concept of pluriZymes. Thus, the focus of the objectives has been specially given to the design of active sites in functional enzymes to create catalysts with multiple functionalities and the enhancement of these artificial functional sites. The generated variants were later experimentally validated by our collaborators to test the performance of our methods/ideas to create compelling pluriZymes. The work is depicted as a compendium of research articles that address the objectives of the thesis:

- **To obtain an enzyme with more than two active sites (active sites $\geq$ 3)**: What is the limit of functionalities in a single protein scaffold? To answer this question, a recent design of a pluriZyme with two active sites was used as starting point. A new type of active site was added to add different chemistries to the newly designed biocatalyst.

- **To design pluriZymes in interesting enzymes according to the academic community or the industrial sector, aiming to attract investors and other research groups to the idea**: Until now, our pluriZymes have only been designed for serine esterases, and the added functionality has been a hydrolase site. Demonstrating that pluriZymes can be generalized in other enzyme families would show their potential for cascade reactions in a single protein scaffold without requiring suicide inhibitors. As mentioned in *Chapter 1*, ω-TAs are incredibly pivotal for synthesizing enantiopure drugs in the pharmaceutical industry, making them a great candidate to highlight the capabilities of pluriZymes.

- **To enable the tailoring of artificial hydrolase sites from pluriZymes at will**: Can we improve the activity of an added hydrolase site towards a specific substrate? Can we make the active site accept a broader range of substrates? The latter has been explored in this thesis with the study of substrate promiscuity on serine esterases, followed by transforming a low-promiscuous enzyme into a prominent one while not affecting its activity.

- **To ease and improve the computational design of pluriZymes**: The goal is to democratize the development of functional sites to the community, by automatizing the process and lowering the computational costs. The knowledge obtained from previous objectives was essential to achieving this goal, giving us a benchmark of systems. A heuristic method was developed to test different variants computationally and rank them according to the system's total energy, imposing the proper catalytic architecture to design a new active site. The code was tested on an esterase with experimental validation of the variants.

3 | Results

The results chapter is encompassed as a compendium of publications (either published or under review). The articles included in the thesis are presented according to the proposed objectives. My contribution to each paper is briefly but clearly explained next.

Articles of the thesis:

- **Article I**: Fernandez-Lopez L, Roda S, Gonzalez-Alfonso JL, Plou FJ, Guallar V, Ferrer M. Design and Characterization of In-One Protease-Esterase PluriZyme. International Journal of Molecular Sciences. 2022; 23(21):13337
  My contribution: I performed the whole computational part of the study and came up with the design of the enzyme variant. I played an important role in writing the main manuscript.

- **Article II**: Roda S, Fernandez-Lopez L, Benedens M, Bollinger A, Thies S, Schumacher J, Coscolín C, Kazemi M, Santiago G, Gertzen CGW, Gonzalez-Alfonso JL, Plou FJ, Jaeger KE, Smits SHJ, Ferrer M, Guallar V. A Plurizyme with Transaminase and Hydrolase Activity Catalyzes Cascade Reactions. Angewandte Chemie International Edition. 2022; 61(37), e202207344
  My contribution: I performed the whole computational part of the article and developed the different mutants. I played a major role in writing the main manuscript.

- **Article III**: Roda S, Fernandez-Lopez L, Cañadas R, Santiago G, Ferrer M, Guallar V. Computationally driven rational design of substrate promiscuity on serine ester hydrolases. ACS Catalysis. 2021; 11(6), 3590–3601
  My contribution: I performed half of the work related to computer simulations and played a leading role in writing the main manuscript.

- **Article IV**: Roda S, Terholsen H, Meyer JRH, Guallar V, Bornscheuer U, Kazemi M. AsiteDesign: A semi-rational algorithm for automated enzyme design. The Journal of Physical Chemistry B. 2022 (In review)
  My contribution: I helped in the development of the new algorithm. I mainly

performed the computational benchmark of the developed tool, selected the variants, and wrote a big part of the manuscript.

Articles not included, but related to the thesis:

- **Article V**: Municoy M, <u>Roda S</u>, Soler D, Soutullo A, Guallar V. AquaPELE: A Monte Carlo-Based Algorithm to Sample the Effects of Buried Water Molecules in Proteins. Journal of Chemical Theory and Computation. 2020; 16(12), 7655–7670
  My contribution: I collaborated in the validation of the conceived method and helped in the writing of the manuscript.

- **Article VI**: <u>Roda S</u>, Robles-Martín A, Xiang R, Kazemi M, Guallar V. Structural-Based Modeling in Protein Engineering. A Must Do. The Journal of Physical Chemistry B. 2021; 125(24), 6491–6500
  My contribution: I mainly wrote the review and performed the shown *in silico* simulations as well.

- **Article VII**: Vidal P, Martínez-Martínez M, Fernandez-Lopez L, <u>Roda S</u>, Méndez-García C, Golyshina OV, Guallar V, Peláez AI, Ferrer M. Metagenomic Mining for Esterases in the Microbial Community of Los Rueldos Acid Mine Drainage Formation. Frontiers in Microbiology. 2022; 13
  My contribution: I performed the computational analysis shown in the article, wrote that part of the manuscript, and helped revise the document.

- **Book chapter I**: <u>Roda S</u>, Santiago G, Guallar V. Mapping enzyme-substrate interactions: its potential to study the mechanism of enzymes. Advances in Protein Chemistry and Structural Biology. 2020; 122, 1–31
  My contribution: I wrote the whole book chapter and prepared all figures except figure 4 of the manuscript.

Article

Design and Characterization of In-One Protease-Esterase PluriZyme

Laura Fernandez-Lopez [1,†], Sergi Roda [2,†] , Jose L. Gonzalez-Alfonso [1] , Francisco J. Plou [1] , Víctor Guallar [2,3,*] and Manuel Ferrer [1,*]

1   Department of Applied Biocatalysis, ICP, CSIC, 28049 Madrid, Spain
2   Department of Life Sciences, Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
3   Institution for Research and Advanced Studies (ICREA), 08010 Barcelona, Spain
*   Correspondence: victor.guallar@bsc.es (V.G.); mferrer@icp.csic.es (M.F.)
†   These authors contributed equally to this work.

Abstract: Proteases are abundant in prokaryotic genomes (~10 per genome), but their recovery encounters expression problems, as only 1% can be produced at high levels; this value differs from that of similarly abundant esterases (1–15 per genome), 50% of which can be expressed at good levels. Here, we design a catalytically efficient artificial protease that can be easily produced. The *PluriZyme* EH_{1AB1} with two active sites supporting the esterase activity was employed. A Leu24Cys mutation in EH_{1AB1}, remodelled one of the esterase sites into a proteolytic one through the incorporation of a catalytic dyad (Cys24 and His214). The resulting artificial enzyme, EH_{1AB1C}, efficiently hydrolysed (azo)casein at pH 6.5–8.0 and 60–70 °C. The presence of both esterase and protease activities in the same scaffold allowed the one-pot cascade synthesis (55.0 ± 0.6% conversion, 24 h) of L-histidine methyl ester from the dipeptide L-carnosine in the presence of methanol. This study demonstrates that active sites supporting proteolytic activity can be artificially introduced into an esterase scaffold to design easy-to-produce in-one protease-esterase *PluriZymes* for cascade reactions, namely, the synthesis of amino acid esters from dipeptides. It is also possible to design artificial proteases with good production yields, in contrast to natural proteases that are difficult to express.

Keywords: esterase; *PluriZyme*; protease; protein engineering; computational chemistry; cascade reaction

1. Introduction

One-pot cascade reactions are chemical processes highly appealing to the industrial sector, as they allow the synthesis of complex products, starting from relatively simple reaction conditions [1–4]. Nevertheless, the implementation of these reactions is a demanding task, and usually, it requires the engineering of each catalyst from each internal chemical reaction. Moreover, catalysts must be specific for their reactant, avoiding unpleasant side products as much as possible. For that reason, enzymes are excellent candidates to set up multistep tandem reactions, since they work under milder conditions than inorganic catalysts and are also regio- and stereo-selective/specific [2–4]. Additionally, recent advances in the rational design and directed evolution of enzymes have introduced a significant success rate in improving diverse enzymatic properties, allowing them to compete with conventional catalysts [2–5]. Remarkably, current developments include designing enzymes with new-to-nature catalytic activities, such as carbene transfer in an engineered cytochrome P450 enzyme [6], expanding the variety of the cascade reactions we can tackle. Another interesting example of the usage of enzymes for cascade reactions is the one developed by Merck & Co., to synthesize islatravir (a potential drug for HIV treatment) from relatively simple building blocks [7].

Performing the cascade reaction within a single enzyme introduces complexities. For this reason, the study and design of biocatalysts performing different chemical reactions

in either the same protein scaffold [8–13] or by linking multiple domains [14–17] have become a hot topic in protein engineering through a number of strategies [18,19]. One of these strategies is to benefit from the capabilities of protein engineering supported by the computational resources for the design of artificial enzymes with a superior or novel performance, compared to natural enzymes. Beyond the engineering of enzymes by these techniques, the possibility of incorporating active centers into protein scaffolds opens the opportunity to design artificial biocatalysts. A number of computational methods have been successfully applied to introduce biological active sites into protein scaffolds. They include, the Rosetta-like methods and the Protein Energy Landscape Exploration (PELE) software, through which biological sites supporting ester-hydrolysis have been incorporated in different protein scaffolds [8,9,19]. However, PELE is the only described method to generate several artificial biological actives sites in the same enzyme scaffold, which open a range of possibilities in cascade reactions, as demonstrated using different enzyme scaffolds into which two same or different biological activities were introduced [8,20]. These two-active sites enzymes, named *PluriZymes*, could introduce an ideal scenario for one-pot cascade reactions, by reducing the costs of producing two enzymes separately. In some cases, their use may also allow increased yields by facilitating the transfer of reaction intermediates between active sites within the same protein, compared to the transfer between sites of different proteins [20]; however, this will depend on the architecture and positioning of these sites.

One example is our recent work building *PluriZymes* in two different enzyme families. First, we developed an esterase, EH_{1AB1}, where a second catalytic triad (Ser-His-Asp/Glu) was added, creating an enzyme with artificial and native active sites supporting ester hydrolysis (T_{opt} of 8–45 °C) [8]. Importantly, the native active site could be transformed into a metal-complex chemocatalytic site by adding a suicide inhibitor, allowing the oxidation and Friedel–Crafts alkylation reactions. Thus, one-pot cascade reactions could be constructed from this biocatalyst [8]. Second, we took an ω-transaminase and added an artificial site supporting ester hydrolysis through the introduction of a catalytic triad (Ser-His-Asp/Glu). A polypeptide having two biotic sites catalysing different types of chemical reactions was thus designed, which could transform oxo-esters into amino acids in a one-pot reaction [20].

In the present study, we aimed at a different approach, pushing the limits of a hydrolase site by adding extra biochemistry through the addition of a cysteine-histidine catalytic dyad. Using our *PluriZyme* EH_{1AB1}, we introduced the protease activity by designing a single mutant, Leu24Cys, which was capable of recycling a histidine residue, His214, from an already existing catalytic triad. A priori, computational analyses indicate that the mutation should not disrupt the esterase activity of the recycled esterase site. Therefore, the newly designed *PluriZyme*, herein referred to as EH_{1AB1C}, included three potential sites. The first supports ester hydrolysis through a native catalytic triad (Ser161, Asp256 and His286) and an oxyanion hole (Gly88, Gly89 and Gly90), with Ser161 being the nucleophile. The second, also supporting ester hydrolysis, would employ an artificial catalytic triad (Ser211, Asp25 and His214) with Ser211 as the nucleophile and an oxyanion hole (Gly207, Tyr208 and Phe209). The third would support the protease activity through a catalytic dyad (Cys24 and His214).

Adding a site supporting proteolytic activity was targeted, as proteases are pivotal enzymes for the hydrolysis of peptide bonds in materials where proteins are abundant components and are also widely used in organic synthesis [21]. This is why they constitute 60–65% of the global industrial market, growing at an annual growth rate of 5.6% [22]. Through evolution, proteases have adapted to the wide range of conditions found in complex organisms (variations in pH, reductive environment, etc.) and use different catalytic mechanisms for substrate hydrolysis [23]; their mechanism of action classifies them as either serine, cysteine or threonine proteases (amino-terminal nucleophile hydrolases) or aspartic, metallo and glutamic proteases (with glutamic proteases being the only subtype not found in mammals thus far) [24]. Proteases specifically cleave protein substrates either from the N or C termini (aminopeptidases and carboxypeptidases, respectively) and/or in

the middle of the molecule (endopeptidases). Proteases can be easily screened by functional screens or in silico predictions in microorganisms or microbial communities by applying genomic and metagenomic approaches [22]. Of course, it should be stressed that the novelty itself does not guarantee a better enzymatic performance and better opportunities for commercialization, whose analysis requires laborious wet lab work. This is not a trivial exercise, given that not all genes in a genome or a metagenome can be successfully cloned and expressed. This is especially important in the cases of proteases that suffer major problems of expression, compared to other types of enzymes that are as easy to be screened as proteases but better to be produced at high levels [22]. Proteases are used in a broad range of applications, including biorefineries targeting a broad range of biomasses [22]. However, the known and new proteases positively impact additional processes, such as cascade reactions where proteases have a pivotal role [25]. Here, we targeted a cascade reaction involving peptide bond cleavage and ester bond formation and demonstrated that the newly designed EH$_{1AB1C}$ *PluriZyme* was capable of converting the dipeptide L-carnosine (β-alanine-L-histidine) into L-histidine methyl ester, an intermediate for the design of Schiff base ligands [26].

2. Results

2.1. Molecular Simulations

This work is based on the design of *PluriZymes* (latin root pluri: multiplicity), an enzyme design in which a single polypeptide harbours two different active centers, one native and one artificial [8,20]. The idea is based on locating a native enzyme, through the Protein Energy Landscape Exploration (PELE) software, existing binding pockets where a target substrate could be accommodated and turning them into catalytic active sites by introducing all the residues needed for the catalysis. Following from this, we have recently successfully found and designed, by introducing a few mutations, a second artificial active site (Ser-His-Asp) in an esterase containing a native site to generate a *PluriZyme*, EH$_{1AB1}$, with two efficient biological active sites for the ester hydrolysis that coexists in a close region. Compared to the natural triad (referred to as EH$_{1A1}$), the newly introduced artificial site (referred to as EH$_{1B1}$) is slightly more solvent exposed and located at an ~10 A distance (Figure 1A). We have considered increasing the number of chemical reactions within this enzymatic design. The computational protocol employed herein is similar to the one used in our *PluriZyme* designs, largely described in [8,9,19,27], but we have focused more on a local region exploration.
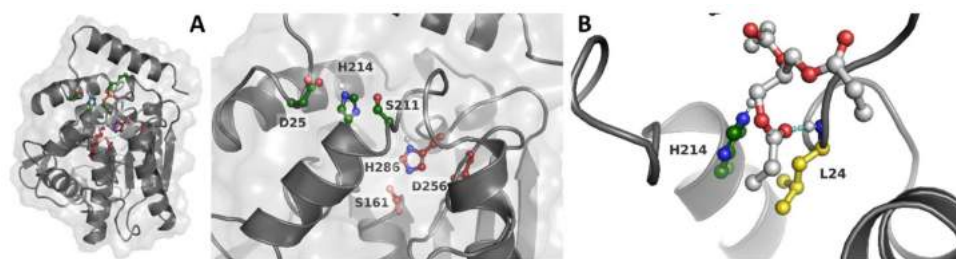

Figure 1. (**A**) Slice of the 3D structure of EH$_{1AB1}$ representing the main and artificial active sites. The C atoms are stained in maroon and dark green in the main active site and artificial active site, respectively. (**B**) Representative binding pose of glyceryl tripropionate (with C atoms stained in gray) where we can see the position of Leu24. Leu24 (in yellow) forms a hydrogen bond with the carbonyl O atom in the ester bond by the NH group in the backbone (indicated by a cyan dashed line) and is close to the His214 residue.

We first attempted to perform an additional global surface exploration for locating potential sites supporting the ester hydrolysis using glyceryl tripropionate as a probe. As shown in Figures S1 and S2, we did not obtain additional alternatives. We then considered the incorporation of a new active center to support another reaction, namely, proteolysis,

to explore the possibility of designing an in-one protease-esterase *PluriZyme*. For this, we focused on the local analysis of the catalytic triad regions (Figure 1A), where we clearly observed how the more exposed site, EH_{1B1}, could accommodate bulkier substrates (as in the case of peptides). Following the inspection of the best enzyme-substrate poses (Figure 1B), we decided to introduce the Leu24Cys variant, named EH_{1AB1C}, which, together with the catalytic histidine of EH_{1B1}, His214, could introduce a catalytic dyad similar to those seen in proteases.

We proceeded by preparing multiple dipeptides, namely, AH, AQ, DI, EA, FF, KA, LA, LL, NV, PF, QQ, RG, SW, TM, YN and YY, and performed a local PELE exploration for each dipeptide. By doing so, we can model the propensity to form catalytically active positions between each peptide bond and the newly designed catalytic dyad: Cys24 and His214. Here, we were not searching for the best peptide but to see if a diverse set of them could reach catalytically active conformations. Interestingly, as shown in Figure 2, Table S1, and Figures S3–S5, all dipeptides, except for YY were able to find the EH_{1B1} site and reach catalytic poses around the engineered cysteine-histidine dyad. Likewise, it can be seen that the catalytic hydrogen bond distances (Cys to His) were quite good during the simulation (Table S2, Figures S6 and S7).
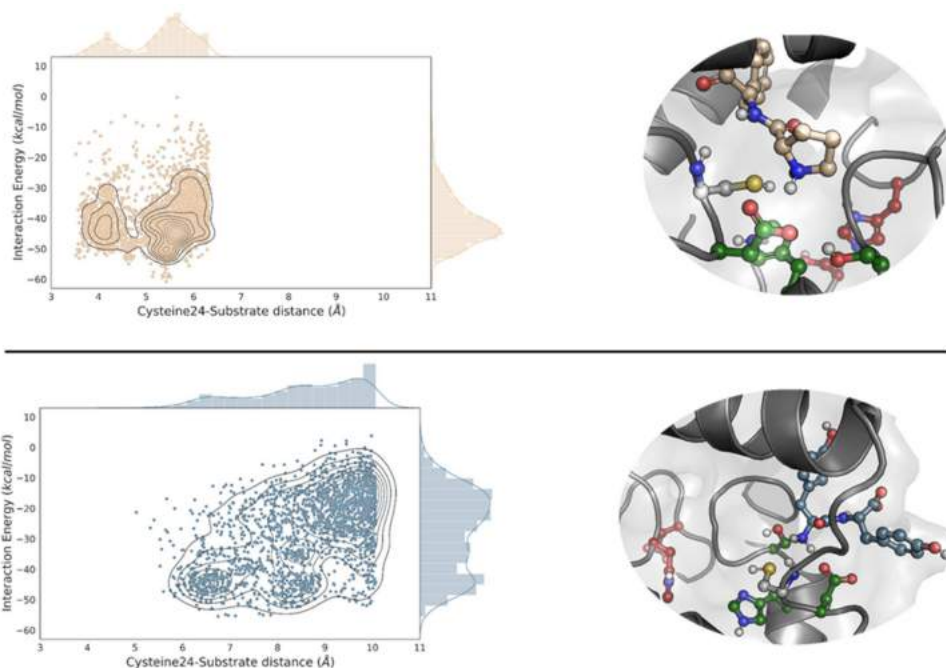
Figure 2. PELE's energy analysis and representative geometries. EH_{1AB1C} density plots of the distribution of the catalytic cysteine-substrate distances against the interaction energy for the best-bound dipeptide substrate (PF, top) and the worst one (YY, bottom). Only the 10% lowest percentile, regarding the distance is shown. In the right panel, we display representative catalytic poses of the dipeptide substrate in the putative protease site. The C atoms are stained in maroon, dark green and yellow in the EH_{1A1} active site residues, EH_{1B1} and Cys24, respectively, and each substrate is shown using the density plot colour (in wheat for PF and blue for YY). The energy profiles were created with the Matplotlib library [28].

2.2. Experimental Validation: EH_{1AB1C} Is an Efficient Protease

The recombinant mutant, hereafter referred to as EH_{1AB1C}, was successfully expressed in soluble form in *Escherichia coli* and purified by nickel affinity chromatography. The purified protein (approx. 10 mg per litre of culture) was desalted by ultrafiltration, and its proteolytic activity was tested through a general fluorescence assay, namely, the BODIPY-FL-casein assay using the EnzChek® Protease Assay Kit, which is insensitive to pH changes.

First, the pH profile of the enzyme was obtained (Figure 3A). Its optimal pH for activity was 7.0, retaining more than 70% of the maximal activity at pH values from 6.5 to 8.0. We then analysed its temperature profile using the chromogenic substrate azocasein. At pH 7.0, EH_{1AB1C} showed maximal activity at 70–75 °C, retaining more than 70% of the maximum activity at 50–85 °C (Figure 3B). The specific activity of EH_{1AB1C} was compared to that of the commercial protease Neutrase 0.8 L (Novozymes A/S, Bagsvaerd, Denmark). At pH 7.0 and 30 °C, the specific activity of EH_{1AB1C} was 2.63 ± 0.06 U/mg protein, while that of the commercial Neutrase 0.8 L was 1.86 ± 0.11 U/mg protein. Note that the original design, EH_{1AB1}, did not show any proteolytic activity with BODIPY-FL-casein or azocasein, demonstrating that the incorporation of the dyad Cys24-His214 introduced proteolytic activity.
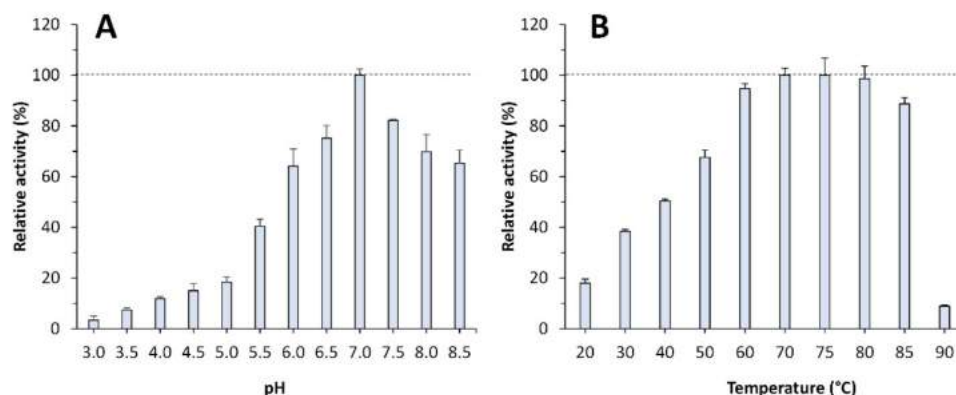
Figure 3. pH and temperature profiles of the purified EH_{1AB1C}. (**A**) pH profile at 30 °C. (**B**) Temperature profile at pH 7.0. The maximal activity was defined as 100%, and the relative activity is shown as the percentage of the maximal activity (mean $\pm$ SD of triplicates), determined under standard reaction conditions with BODIPY-FL-casein (in panel (**A**)) and azocasein (in panel (**B**)) as the substrates. The graphics were created with Excel version 14.0.

The esterase activity of the EH_{1AB1C} mutant was quantified at 30 °C and pH 8.0 with the model ester glyceryl tripropionate and compared to that of the initial construct EH_{1AB1}. This substrate can be converted by both esterase sites (the native and the artificial sites). We found that the EH_{1AB1C} mutant is an efficient esterase capable of hydrolysing this ester at 3160 ± 76 U/g. However, we observed that this value was only 14.5% of that of EH_{1AB1}. To evaluate whether this reduction could be because the artificial esterase center (B site), during its remodelling to add a protease center, has been altered and with it, its activity, would require evaluating the activity with B-site specific esters. However, all tested esters hydrolysed by this site are also hydrolysed by the native esterase center (A-site), due to the broad substrate specificity of the latter [8]. Therefore, we cannot be sure whether the observed reduction in activity is due to a possible effect of the mutation introduced (Leu24Cys) on the architecture of the B-center, to a local effect on the structure, or to the possibility of partial autolysis (self-digestion) by the addition of the protease site.

2.3. Application of EH$_{1AB1C}$ in a One-Pot Cascade Reaction

Following the computational design of a protease site in EH_{1AB1} and the characterization of the successful variant EH_{1AB1C}, we wanted to test the ability of this validated *PluriZyme* to catalyse a cascade reaction of interest. As a model reaction, we chose to synthesize L-histidine methyl ester, an intermediate for the design of Schiff base ligands [26], from the dipeptide L-carnosine (β-alanine-L-histidine) (Figure 4).

The simulation of the reaction by a local PELE exploration showed efficient L-alanine-L-histidine (AH) catalytic (hydrolytic) binding poses at the proteolytic site (Table S1). We also obtained good catalytic poses for L-carnosine (β-alanine-L-histidine) (Figure 5; Table S1). Thus, we expected that L-carnosine would be hydrolysed at the protease site, due to its similarity with the L-alanine-L-histidine. We set up two reactions at 40 °C and pH 7.0. The first one contained L-carnosine (5 mM) in the buffer, 40 mM HEPES, at pH 7.0. The second one represented a one-pot cascade reaction with all the reagents necessary for the hydrolysis of L-carnosine and the esterification of the corresponding reaction products with methanol, i.e., L-carnosine and methanol. Following the addition of the *E. coli* cells expressing the EH$_{1AB1C}$ *PluriZyme*, the levels of the substrate L-carnosine, the intermediates β-alanine and L-histidine, and the possible products β-alanine methyl ester and L-histidine methyl ester, were quantified using high-performance liquid chromatography (HPLC) over 20 h (Figure 6 and Figure S8). Note that for these tests, cells expressing EH$_{1AB1C}$ were used instead of purified protein to increase the stability of the biocatalysts, which in soluble form may be inactivated.
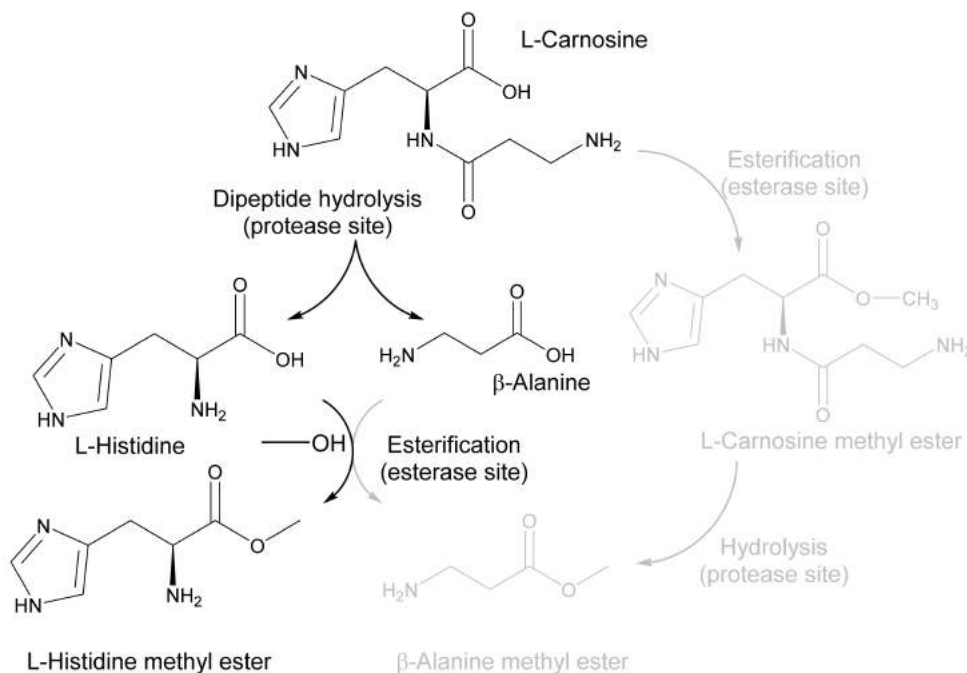
Figure 4. Schematic representation of the main product obtained in a one-pot reaction with the dipeptide L-carnosine (β-alanine-L-histidine) and methanol. A schematic representation of the two reaction intermediates and the two possible products (β-alanine and L-histidine methyl esters) is shown. As shown in the figure, the preferential route found is the production of L-histidine methyl ester via the hydrolysis of L-carnosine at the protease site and the selective esterification with methanol of L-histidine (but not β-alanine) at the esterase site. No appreciable formation of L-carnosine methyl ester and β-alanine methyl ester (gray colour) was detected. The figure was created using ChemDraw 18.2.
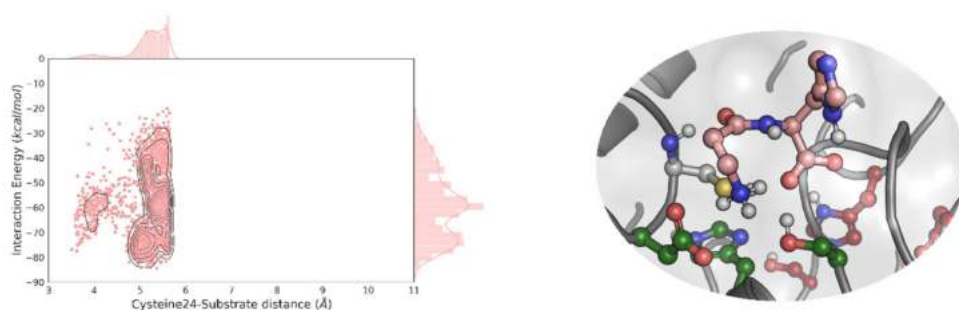
Figure 5. EH_{1AB1C} density plot of the distribution of the catalytic cysteine-substrate distance against the interaction energy for L-carnosine. Only the 10% lowest percentile, regarding the distance is shown. In the right panel, we display the representative catalytic poses of L-carnosine in the putative protease site. The C atoms are stained in maroon, dark green, and yellow in the EH_{1A1} active site residues, EH_{1B1}, and Cys24, respectively, and L-carnosine is shown using the pink colour used in the density plot. The energy profiles were created with the Matplotlib library [28].

We found that the L-carnosine dipeptide (concentration of 5 mM) was fully converted (>93.0 ± 0.1%) after 20 h in the buffer, obtaining β-alanine (4.5 ± 0.1 mM) and L-histidine (4.75 ± 0.05 mM) (Figure 6). When the reaction was performed in methanol, L-carnosine was converted after 20 h (90.5 ± 12.3%), with the main products being β-alanine (4.5 ± 0.1 mM), L-histidine (1.8 ± 0.1 mM) and L-histidine methyl ester (2.8 ± 0.5 mM). β-Alanine methyl ester and L-carnosine methyl ester were not found as products. This result demonstrates that L-carnosine is hydrolysed at the artificial proteolytic site; furthermore, the esterase site only recognized L-histidine but not β-alanine as a substrate, which is esterified to yield L-histidine methyl ester as the only amino acid ester.
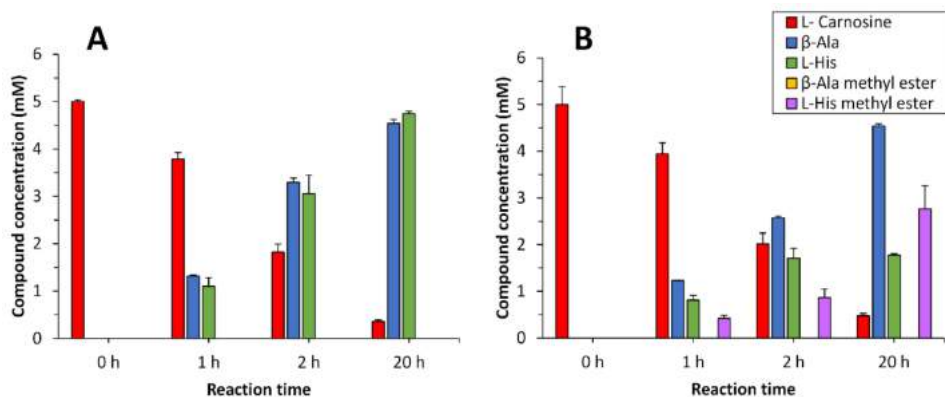

Figure 6. Concentrations of the substrates, intermediates and final products obtained for the conversion of L-carnosine by the EH_{1AB1C} *PluriZyme* in the absence (**A**) or presence (**B**) of methanol. Reaction conditions in A: [cells expressing EH_{1AB1C}]: 5 μL of resuspended cells at $OD_{600\,nm}$ of 15.0; [L-carnosine]: 5 mM; buffer: 92.5 μL 40 mM HEPES, pH 7.0; reaction final volume: 100 μL. Reaction conditions in A: [cells expressing EH_{1AB1C}]: 5 μL of resuspended cells at $OD_{600\,nm}$ of 15.0; [L-carnosine]: 5 mM; solvent: 92.5 μL methanol; reaction final volume: 100 μL (92.5% methanol). Note, that the amount of the β-Ala methyl ester was below the detection limit under our experimental conditions, and this is why yellow bars are not visible in the figure. The figure was created using SigmaPlot 14.0 software.

To highlight, we found that when using as biocatalyst cells expressing the original EH_{1AB1} *PluriZyme*, L-carnosine was not converted to any of the intermediates or final products (Figure S9A); this indicates that the proteases in *E. coli* do not affect the results and that in the absence of the artificial proteolytic center, the cascade reaction is not feasible. Further, both the original EH_{1AB1} *PluriZyme* and the mutant EH_{1AB1C} *PluriZyme* biocatalysts

were shown to convert L-histidine to L-histidine methyl ester in the presence of methanol, but not β-alanine to the corresponding methyl ester (Figure S9B,C). These control tests, together with the results presented in Figure 6, demonstrated that the incorporation of the dyad Cys24-His214 in the EH$_{1AB1C}$ *PluriZyme* was responsible for the proteolysis of L-carnosine, and that the esterase(s) active site(s) originally present in EH$_{1AB1C}$ supported the acylation of L-histidine and the production of the amino acid ester (L-histidine methyl ester).

3. Discussion

Proteases that can break ester bonds and esterases and lipases that can break amide bonds exist in nature; they are an example of catalytically promiscuous enzymes [29–34]. The presence of a Ser or a Cys catalytic residue, along with differences in the substrate binding site, seems to be the main reason for substrate discrimination and the promiscuous behavior. When attempting to engineer such activity from scratch (ab initio) or when introducing enough mutations to accommodate a dyad/triad, we typically find low activity profiles [35–37]. Nevertheless, these results have paved the way for more recent implementations, such as our recent *PluriZymes*, where the extensive in silico optimization (catalytic distances) has provided wide substrate promiscuity, high substrate conversion rates, and even the development of cascade reactions [8,9]. In this study, we introduced the protease activity by reusing a (previously engineered) esterase site. We recycled the catalytic histidine and inserted only a cysteine residue, achieving high catalytic rates in the standard peptide assays. As in our previous designs, the potential of running accurate enzyme-substrate induced fit simulations provided enough insights. Thus, these results again illustrate how simulation techniques using molecular modelling are mature enough and capable of providing a realistic description of localized changes. Such modelling potential, along with the propensity of catalytic dyads/triads to form (stabilizing) hydrogen bonds, makes the design of hydrolase active sites an affordable task; we have succeeded in all recent attempts [8,9,20].

The main reason behind our efforts derives from our goals of developing single scaffold enzymes with multiple biochemical activities that are capable of performing one-pot cascade reactions. Previous attempts in the native esterase EH$_1$ focused on introducing a second artificial esterase active site and the incorporation of an irreversible-linked inhibitor containing a meta-chelating moiety to one of the sites, allowing the addition of artificial abiotic oxidative chemistry as a complement to the original biotic esterase activity [8]. More recently, we were able to add artificial biotic esterase activity into an ω-transaminase [20]. With our new catalytically efficient design, EH$_{1AB1C}$, we are now capable of breaking amide bonds and forming ester bonds to yield natural and nonnatural amino acid esters, herein exemplified by L-histidine methyl ester, which are important intermediates in organic synthesis [38]. Amino acid methyl esters, such as L-histidine methyl ester, can be chemically synthesized at room temperature through the esterification of L-histidine with trimethylchlorosilane (TMSCl) and methanol at high yields (88–96%, using 10 mM L-histidine) [38]; the biocatalysts reported herein offer milder and environmentally friendly conditions and good yields for the synthesis of amino acid esters useful in the pharmaceutical industry. In addition, both protease and esterase sites in EH$_{1AB1C}$ may offer specificity, as shown here, for example, by the capacity of the latter to esterify L-histidine but not β-alanine. The capacity of some proteases to catalyse ester hydrolysis has been used to produce amino acid esters with a high optical purity through the selective hydrolysis of D,L-amino acid esters (e.g., methyl and benzyl esters) to provide, for example, L-amino acids and optically pure D-amino acid esters [39,40]. However, in this case, an amino acid ester is needed as the initial substrate, in contrast to our *PluriZyme*, which can directly use dipeptides and possibly longer oligopeptides.

We would like to highlight that the reaction conditions for the substrate reported in this work and others yet to be tested have not been optimized. Thus, this design offers interesting properties to be utilized in industrial settings, such as an integrated biorefinery

for biomass recovery based on proteases for the production of protein hydrolysates, bioactive peptides, and amino acids applicable to a wider range of applications. It also offers a platform to synthesize a wide range of biobased products through in-one cascade reactions involving ester and amine bond hydrolysis and formation.

4. Materials and Methods

4.1. Materials

Azocasein (ref. A2765-1G), β-alanine (ref. 146064-25G), L-histidine (ref. H7750-25G), L-histidine methyl ester (ref. H15403-25G), β-alanine methyl ester (ref. 05210-10G), L-carnosine (ref. C9695-10MG), glyceryl tripropionate (ref. W328618-1KG-K), and methanol (ref. 34966-1L) were ordered from Merck Life Science S.L.U. (Madrid, Spain). The EnzChek® Protease Assay Kit (ref. E6638) was provided by Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA. FMOC chloride (ref. GE3236-1G) was purchased from Glentham Life Sciences, Corsham, UK.

4.2. Protein and Chemical Preparation for the In Silico Analysis

The apo EH_{1A} crystal structure (5JD4) and the holo EH_{1AB1} crystal structure (6RB0) were prepared and protonated at pH 8.0, the pH at which the experimental assays were performed, using Protein Preparation Wizard [41]. This includes fixing side chains and loops missing in the crystal structure using Prime software [42]. The ester compound used as a probe to find noncatalytic hydrolase sites was glyceryl tripropionate; the peptide binding assays of the generated variant used 16 dipeptides (AH, AQ, DI, EA, FF, KA, LA, LL, NV, PF, QQ, RG, SW, TM, YN, and YY) as substrates. All substrates were modelled using the OPLS2005 force field [43]. The atomic charges of glyceryl tripropionate and the catalytic serine residues bound with the methyl hydrogen (*R*)-hexylphosphonate inhibitor [8,20] were calculated with Jaguar [44] using density functional theory with a B3LYP-D3 exchange-correlation functional and the polarized triple-zeta (pVTZ) basis set.

4.3. Protein Energy Landscape Exploration (PELE) Simulations

PELE was used to find the noncatalytic peptide binding sites in EH_{1A}/EH_{1AB1} and check if the catalytic poses can be reached in the functionalized variant [8,9,19,27]. PELE is a Monte Carlo (MC)-based algorithm coupled with protein structure prediction methods [45]. The heuristic MC approach begins with the sampling of different microstates by initially applying small perturbations (translations and rotations) on the ligand. Then, the flexibility of the protein is taken into account by applying normal modes through the anisotropic network model (ANM) approach. Once the system has been perturbed, side chains of the residues near the ligand are sampled with a library of rotamers to avoid steric clashes. Finally, a truncated Newton minimization with the OPLS2005 force field [43] is performed, and the new microstate is accepted or rejected, according to the Metropolis criterion. The variable dielectric generalized Born non-polar (VDGBNP) implicit solvent [46] was applied to mimic the influence of water around the protein.

4.4. Prediction of ΔΔG in the EH_{1AB1} Variant

The ΔΔG(mut-WT) of stability in the experimentally tested variants was calculated using the module of thermodynamic stability from HotSpot Wizard, which uses FoldX to repair possible problems in the protein structure and Rosetta to perform the energy minimization and ΔΔG calculation (according to protocol 3 from Rosetta) [47].

4.5. Source and Production of EH_{1AB1C}

The sequence of EH_{1AB1C} was synthesized by GenScript Biotech (GenScript Biotech, EG Rijswijk, Netherlands) and was codon-optimized to maximize the expression in *E. coli*. The gene was flanked by BamHI and HindIII (stop codon) restriction sites and inserted into a pET-45b(+) expression vector with an ampicillin selection marker (GenScript Biotech, EG Rijswijk, Netherlands), which was further introduced into *E. coli* BL21(DE3). The

soluble N-terminal histidine (His)-tagged protein was produced and purified (98% purity, as determined by SDS–PAGE analysis using a Mini PROTEAN electrophoresis system, Bio-Rad, Madrid, Spain; Figure S10) at 4 °C after binding to a Ni-NTA His-Bind resin (Merck Life Science S.L.U., Madrid, Spain), as previously described [8,20], and stored at −20 °C until use at a concentration of 1.5 mg/mL in 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 7.0). Approximately 10 mg of purified EH_{1AB1C} was obtained on average from a 1-L culture.

4.6. Activity Tests

The proteolytic activity was tested with the EnzChek® Protease Assay Kit containing casein labelled with BODIPY® FL (E6638) dye (green fluorescence), which is insensitive to pH. The assay was prepared, according to the protocol provided by the manufacturer (Thermo Fisher, Madrid, Spain). If otherwise not indicated, the reactions were carried out in 96-well plates by adding 1 μL of BODIBY FL casein (1 mg/mL) to 99 μL of 40 mM HEPES buffer at pH 7.0 with 1.0–5.0 μg of protein and incubated for 1 h, protected from light at 30 °C. The fluorescence (λ_{exc}: 485 ± 20 nm; λ_{emm}: 528 ± 20 nm) was then measured in a Synergy HT multi-mode microplate reader (Biotek Instruments, Winooski, VT, USA). All values were determined in duplicate by determining the fluorescence at 60 min and were corrected for nonenzymatic transformation. For the pH determination (in triplicate), the previous conditions were used but with 50 mM Britton and Robinson buffer at pH 3.0–8.5. To determine the thermal profile (in triplicate), the substrate used was chromogenic azocasein. Briefly, to 25 μL of 2.5% *w/v* azocasein (prepared in water), 15 μL of 40 mM HEPES at pH 7.0 and 10 μL of EH_{1AB1C} (from a 1.4 mg/mL stock) were added. This reaction was incubated for 40 min with shaking at 700 rpm at different temperatures (20–90 °C). Then, 200 μL of trichloroacetic acid was added to stop the reaction, which was then centrifuged for 5 min at 13,000 rpm. Subsequently, 50 μL of this supernatant was added to 150 μL of 0.5 M NaOH in a 96-well plate, and its absorbance at 440 nm was measured in a Synergy HT Multi-Mode Microplate Reader. In all cases, the control reactions with EH_{AB1} under the same experimental conditions were performed. In all cases, the values were corrected for nonenzymatic transformation.

The hydrolysis of the model ester glyceryl tripropionate by EH_{1AB1} and EH_{1AB1C} was assayed using a pH indicator assay in 384-well plates (ref. 781162, Greiner Bio-One GmbH, Kremsmünster, Austria) at 30 °C and pH 8.0 (5 mM 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid (EPPS) buffer containing 0.45 mM phenol red) in a Synergy HT multi-mode microplate reader in continuous mode at 550 nm over 30 min (extinction coefficient (ε) of phenol red: 8450 M^{-1} cm^{-1}), as reported in [8,20]. The conditions for determining the specific activity (U/mg) were as follows: [protein]: 5 μg/mL; [ester]: 10 mg/mL; reaction volume: 44 μL; T: 30 °C; and pH: 8.0. The activity was calculated by determining the absorbance per minute from the slopes generated [8,20]. In all cases, all values in triplicate were corrected for nonenzymatic transformation, with the absence of activity defined as having at least a twofold background signal.

4.7. Cascade Reaction and the HPLC Analysis

L-Carnosine, used as the substrate, and β-alanine, L-histidine, β-alanine methyl ester and L-histidine methyl ester, used as the standards, were obtained from Merck Life Science S.L.U., Madrid, Spain. The reactions were performed using resting cell assays. Briefly, *E. coli* BL21(DE3) expressing EH_{1AB1C}, His-tagged at the N-terminus, was grown at 37 °C on solid LB agar medium supplemented with 100 μg/mL ampicillin, and one colony was picked and used to inoculate 50 mL of Luria Bertani broth plus antibiotic in a 0.25-L flask. The culture was then incubated at 37 °C and 200 rpm overnight. Then, 50 mL of this culture was used to inoculate 1-l of LB medium plus antibiotic in a 2.5-L flask, which was then incubated to an $OD_{600\,nm}$ of approximately 0.7–1.0 at 37 °C. The 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 at 220 rpm. The cells were

harvested by centrifugation at 5000× *g* for 15 min to yield a pellet of 2–3 g/L (wet weight). The wet cell pellet was resuspended in 60 mL of 40 mM HEPES at pH 7.0 and centrifuged as before, and the cells were retained. Finally, the cells were resuspended in 60 mL of 40 mM HEPES at pH 7.0 until an $OD_{600\,nm}$ of 15.0 and used directly for the activity tests.

To evaluate the proteolytic capacity of EH_{1AB1C} against L-carnosine, the following reaction conditions were used in a 1.5 mL Eppendorf tube: [cells expressing EH_{1AB1C}]: 5 μL of resuspended cells at $OD_{600\,nm}$ of 15.0; [L-carnosine]: 2.5 μL of a 200 mM stock solution in 40 mM HEPES at pH 7.0; T: 30 °C; buffer/solvent: 92.5 μL of 40 mM HEPES buffer at pH 7.0; and reaction final volume: 100 μL (100% buffer). To evaluate the capacity of EH_{1AB1C} to produce amino esters in a cascade reaction, the following conditions were used in a 1.5 mL Eppendorf tube: [cells expressing EH_{1AB1C}]: 5 μL of resuspended cells at $OD_{600\,nm}$ of 15.0; [L-carnosine]: 2.5 μL of a 200 mM stock solution in 40 mM HEPES at pH 7.0; T: 30 °C; buffer/solvent: 92.5 μL of methanol; and reaction final volume: 100 μL (7.5% buffer and 92.5% methanol). Control reactions with L-histidine and β-alanine (using stock solutions at the same concentrations as L-carnosine) instead of L-carnosine were set-up. The same reaction tests were performed with cells expressing EH_{AB1} under the same experimental conditions.

Aliquots of 100 μL of the reaction mixture were taken at different times. The aliquots were filtered using 0.45 μm Durapore filter unit inserts (Millipore Corporation, Billerica, MA, USA) to remove the cells and stop the reaction. The solvent was evaporated using a SpeedVac Concentrator (Model 5301, Eppendorf, Hamburg, Germany) at 30 °C. The remaining L-carnosine and the formed reaction products were derivatized by adding 500 μL of $B_4Na_2O_7$ buffer at pH 9.0 (100 mM) followed by the addition of 500 μL FMOC chloride (4 mM) in acetonitrile. The samples were shaken in a vortexer for 1 h. Then, the samples were analysed by HPLC using a quaternary pump (Model 600, Waters, Milford, MA, USA) coupled to an autosampler (Model 420, Varian ProStar, Wainut Creek, CA, USA). The injection volume was 10 μL. The column was a Zorbax Eclipse Plus C-18 (4.6 × 100 mM, 3.5 μm, Agilent Technologies, Santa Clara, CA, USA) at 40 °C, and the mobile phase was an acetonitrile/H_2O gradient with both solvents acidified with 0.1% (*v*/*v*) formic acid and degassed with helium. The gradient (20 min in total) is specified as follows: 0–12 min, 20→90% acetonitrile; 12–15 min, 90% acetonitrile; 15–15.5 min, 90→20% acetonitrile; 15.5–20 min, 20% (initial conditions). The flow rate was 1.0 mL/min. The detection of peaks was carried out using a photodiode array detector (Model 335, Varian ProStar), and integration was carried out using Varian Star LC Workstation 6.41 at a wavelength of 264 nm. The quantification of the reaction products was performed with the calibration curves of the standards of L-carnosine and the corresponding amino acids and esters at concentrations between 0 and 500 μM.

5. Conclusions

To the best of our knowledge, we have designed the first example of an in-one protease-esterase *PluriZyme*. This construct was built by engineering an artificial active site supporting the protease activity into an esterase by remodelling one of the two esterase active sites of a recent *PluriZyme* design through a Let24Cys mutation. The resulting mutant does contain a native Ser-His-Asp catalytic triad supporting esterase activity and an artificial Cys-His dyad supporting proteolysis. We will continue to explore the biocatalytic potential of this new type of biocatalyst, as suggested by the data presented here, particularly for the synthesis of amino acid esters, which are versatile chiral auxiliary groups employed for the asymmetric synthesis of pharmaceuticals [47]. This exploration will include the use of this enzyme design in other chain reactions, and its comparison with traditional bioenzymatic systems in which an esterase and a protease are combined in one-pot, either in soluble form or co-immobilized preparations. That said, we have previously demonstrated that a *PluriZyme* supporting two different activities works as efficiently as when mixing separate enzymes supporting the two activities needed for the cascade reaction [20]. In both cases, the yields and conversions will depend on multiple factors, yet to be explored. For example,

the separation of the two active sites within the protein scaffold in the case of the *PluriZyme* and the transfer of reaction intermediates between them; this can also occur between active centers of different enzymes in bi-enzymatic systems.

References

1. Nicolaou, K.C.; Edmonds, D.J.; Bulger, P.G. Cascade Reactions in Total Synthesis. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 7134–7186. [CrossRef] [PubMed]
2. Wu, S.; Snajdrova, R.; Moore, J.C.; Baldenius, K.; Bornscheuer, U.T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chem. Int. Ed. Engl.* **2021**, *60*, 88–119. [CrossRef] [PubMed]
3. Devine, P.N.; Howard, R.M.; Kumar, R.; Thompson, M.P.; Truppo, M.D.; Turner, N.J. Extending the Application of Biocatalysis to Meet the Challenges of Drug Development. *Nat. Rev. Chem.* **2018**, *2*, 409–421. [CrossRef]
4. Sperl, J.M.; Sieber, V. Multienzyme Cascade Reactions—Status and Recent Advances. *ACS Catal.* **2018**, *8*, 2385–2396. [CrossRef]
5. Nazor, J.; Liu, J.; Huisman, G. Enzyme Evolution for Industrial Biocatalytic Cascades. *Curr. Opin. Biotechnol.* **2021**, *69*, 182–190. [CrossRef]
6. Coelho, P.S.; Brustad, E.M.; Kannan, A.; Arnold, F.H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310. [CrossRef]
7. Huffman, M.A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K.R.; Canada, K.A.; Devine, P.N.; Duan, D.; Forstater, J.H.; Grosser, S.T.; et al. Design of an in Vitro Biocatalytic Cascade for the Manufacture of Islatravir. *Science* **2019**, *366*, 1255–1259. [CrossRef]
8. Alonso, S.; Santiago, G.; Cea-Rama, I.; Fernandez-Lopez, L.; Coscolín, C.; Modregger, J.; Ressmann, A.K.; Martínez-Martínez, M.; Marrero, H.; Bargiela, R.; et al. Genetically Engineered Proteins with Two Active Sites for Enhanced Biocatalysis and Synergistic Chemo- and Biocatalysis. *Nat. Catal.* **2020**, *3*, 319–328. [CrossRef]
9. Santiago, G.; Martínez-Martínez, M.; Alonso, S.; Bargiela, R.; Coscolín, C.; Golyshin, P.N.; Guallar, V.; Ferrer, M. Rational Engineering of Multiple Active Sites in an Ester Hydrolase. *Biochemistry* **2018**, *57*, 2245–2255. [CrossRef]
10. Liu, Z.; Calvó-Tusell, C.; Zhou, A.Z.; Chen, K.; Garcia-Borràs, M.; Arnold, F.H. Dual-Function Enzyme Catalysis for Enantioselective Carbon-Nitrogen Bond Formation. *Nat. Chem.* **2021**, *13*, 1166–1172. [CrossRef]
11. Díaz-Caballero, M.; Navarro, S.; Nuez-Martínez, M.; Peccati, F.; Rodríguez-Santiago, L.; Sodupe, M.; Teixidor, F.; Ventura, S. PH-Responsive Self-Assembly of Amyloid Fibrils for Dual Hydrolase-Oxidase Reactions. *ACS Catal.* **2021**, *11*, 595–607. [CrossRef]
12. Zhou, Z.; Roelfes, G. Synergistic Catalysis in an Artificial Enzyme by Simultaneous Action of Two Abiological Catalytic Sites. *Nat. Catal.* **2020**, *3*, 289–294. [CrossRef]
13. Filice, M.; Romero, O.; Gutiérrez-Fernández, J.; de Las Rivas, B.; Hermoso, J.A.; Palomo, J.M. Synthesis of a Heterogeneous Artificial Metallolipase with Chimeric Catalytic Activity. *Chem. Commun.* **2015**, *51*, 9324–9327. [CrossRef] [PubMed]

14. Christoffel, F.; Igareta, N.V.; Pellizzoni, M.M.; Tiessler-Sala, L.; Lozhkin, B.; Spiess, D.C.; Lledós, A.; Maréchal, J.-D.; Peterson, R.L.; Ward, T.R. Design and Evolution of Chimeric Streptavidin for Protein-Enabled Dual Gold Catalysis. *Nat. Catal.* **2021**, *4*, 643–653. [CrossRef]

15. Knott, B.C.; Erickson, E.; Allen, M.D.; Gado, J.E.; Graham, R.; Kearns, F.L.; Pardo, I.; Topuzlu, E.; Anderson, J.J.; Austin, H.P.; et al. Characterization and Engineering of a Two-Enzyme System for Plastics Depolymerization. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 25476–25485. [CrossRef]

16. Faylo, J.L.; van Eeuwen, T.; Kim, H.J.; Gorbea Colón, J.J.; Garcia, B.A.; Murakami, K.; Christianson, D.W. Structural Insight on Assembly-Line Catalysis in Terpene Biosynthesis. *Nat. Commun.* **2021**, *12*, 3487. [CrossRef]

17. Zhang, Y.; Hess, H. Toward Rational Design of High-Efficiency Enzyme Cascades. *ACS Catal.* **2017**, *7*, 6018–6027. [CrossRef]

18. Chen, K.; Arnold, F.H. Engineering New Catalytic Activities in Enzymes. *Nat. Catal.* **2020**, *3*, 203–213. [CrossRef]

19. Roda, S.; Robles-Martín, A.; Xiang, R.; Kazemi, M.; Guallar, V. Structural-Based Modeling in Protein Engineering. A Must Do. *J. Phys. Chem. B* **2021**, *125*, 6491–6500. [CrossRef]

20. Roda, S.; Fernandez-Lopez, L.; Benedens, M.; Bollinger, A.; Thies, S.; Schumacher, J.; Coscolín, C.; Kazemi, M.; Santiago, G.; Gertzen, C.G.W.; et al. A Plurizyme with Transaminase and Hydrolase Activity Catalyzes Cascade Reactions. *Angew. Chem. Int. Ed. Engl.* **2022**, *61*, e202207344. [CrossRef]

21. Sheldon, R.A.; Brady, D.; Bode, M.L. The Hitchhiker's Guide to Biocatalysis: Recent Advances in the Use of Enzymes in Organic Synthesis. *Chem. Sci.* **2020**, *11*, 2587–2605. [CrossRef] [PubMed]

22. García-Moyano, A.; Diaz, Y.; Navarro, J.; Almendral, D.; Puntervoll, P.; Ferrer, M.; Bjerga, G.E.K. Two-step Functional Screen on Multiple Proteinaceous Substrates Reveals Temperature-Robust Proteases with a Broad-Substrate Range. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 3195–3209. [CrossRef] [PubMed]

23. Dyer, R.P.; Weiss, G.A. Making the cut with protease engineering. *Cell Chem. Biol.* **2022**, *29*, 177–190. [CrossRef]

24. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS Database of Proteolytic Enzymes; their Substrates and Inhibitors in 2017 and a Comparison with Peptidases in the PANTHER Database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [CrossRef] [PubMed]

25. Wang, Y.; Wang, N. Hydrolase-Catalyzed Promiscuous Reactions and Applications in Organic Synthesis. In *Molecular Biotechnology*; Sedykh, S., Ed.; IntechOpen: London, UK, 2019.

26. Ogawa, R.; Sunatsuki, Y.; Suzuki, T. Schiff Base Ligands Derived from l-Histidine Methyl Ester: Characterization, Racemization, and Dimerization of Their Transition-Metal Complexes. *Eur. J. Inorg. Chem.* **2018**, 1733–1742. [CrossRef]

27. Roda, S.; Santiago, G.; Guallar, V. Mapping Enzyme-Substrate Interactions: Its Potential to Study the Mechanism of Enzymes. *Adv. Protein Chem. Struct. Biol.* **2020**, *122*, 1–31. [PubMed]

28. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]

29. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102*, 4501–4524. [CrossRef]

30. Busto, E.; Gotor-Fernández, V.; Gotor, V. Hydrolases: Catalytically Promiscuous Enzymes for Non-Conventional Reactions in Organic Synthesis. *Chem. Soc. Rev.* **2010**, *39*, 4504–4523. [CrossRef]

31. Liljeblad, A.; Kallio, P.; Vainio, M.; Niemi, J.; Kanerva, L.T. Formation and Hydrolysis of Amide Bonds by Lipase A from *Candida antarctica*; Exceptional Features. *Org. Biomol. Chem.* **2010**, *8*, 886–895. [CrossRef]

32. Simons, C.; van Leeuwen, J.G.E.; Stemmer, R.; Arends, I.W.C.E.; Maschmeyer, T.; Sheldon, R.A.; Hanefeld, U. Enzyme-Catalysed Deprotection of N-Acetyl and N-Formyl Amino Acids. *J. Mol. Catal. B Enzym.* **2008**, *54*, 67–71. [CrossRef]

33. Henke, E.; Bornscheuer, U.T. Fluorophoric Assay for the High-Throughput Determination of Amidase Activity. *Anal. Chem.* **2003**, *75*, 255–260. [CrossRef] [PubMed]

34. Kapoor, M.; Gupta, M.N. Lipase Promiscuity and Its Biochemical Applications. *Process Biochem.* **2012**, *47*, 555–569. [CrossRef]

35. Richter, F.; Blomberg, R.; Khare, S.D.; Kiss, G.; Kuzin, A.P.; Smith, A.J.T.; Gallaher, J.; Pianowski, Z.; Helgeson, R.C.; Grjasnow, A.; et al. Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **2012**, *134*, 16197–16206. [CrossRef] [PubMed]

36. Rajagopalan, S.; Wang, C.; Yu, K.; Kuzin, A.P.; Richter, F.; Lew, S.; Miklos, A.E.; Matthews, M.L.; Seetharaman, J.; Su, M.; et al. Design of Activated Serine-Containing Catalytic Triads with Atomic-Level Accuracy. *Nat. Chem. Biol.* **2014**, *10*, 386–391. [CrossRef]

37. 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.; et al. New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *J. Am. Chem. Soc.* **2015**, *137*, 14905–14911. [CrossRef] [PubMed]

38. Miyazawa, T.; Minowa, H.; Miyamoto, T.; Imagawa, K.; Yanagihara, R.; Yamada, T. Resolution of Non-Protein Amino Acids Via Microbial Protease-Catalyzed Ester Hydrolysis: Marked Enhancement of Enantioselectivity by the use of Esters with Longer Alkyl Chains and at Low Temperature. *Tetrahedron Asymmetry* **1997**, *8*, 367–370. [CrossRef]

39. Li, J.; Sha, Y. A Convenient Synthesis of Amino Acid Methyl Esters. *Molecules* **2008**, *13*, 1111–1119. [CrossRef]

40. Chen, S.-T.; Wang, K.-T.; Wong, C.-H. Chirally Selective Hydrolysis of D,L-Amino Acid Esters by Alkaline Protease Amino Acid-Catalyzed Hydrolysis of Amino Acid Esters with the Free α-Amino Group. *J. Chem. Soc. Chem. Commun.* **1986**, *20*, 1514–1516. [CrossRef]

41. Sastry, G.M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221–234. [CrossRef]

42. Jacobson, M.P.; Pincus, D.L.; Rapp, C.S.; Day, T.J.F.; Honig, B.; Shaw, D.E.; Friesner, R.A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins* **2004**, *55*, 351–367. [CrossRef] [PubMed]

43. Banks, J.L.; Beard, H.S.; Cao, Y.; Cho, A.E.; Damm, W.; Farid, R.; Felts, A.K.; Halgren, T.A.; Mainz, D.T.; Maple, J.R.; et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752–1780. [CrossRef] [PubMed]

44. Bochevarov, A.D.; Harder, E.; Hughes, T.F.; Greenwood, J.R.; Braden, D.A.; Philipp, D.M.; Rinaldo, D.; Halls, M.D.; Zhang, J.; Friesner, R.A. Jaguar: A High-performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142. [CrossRef]

45. 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*, 1304–1311. [CrossRef] [PubMed]

46. Bashford, D.; Case, D.A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152. [CrossRef]

47. Sumbalova, L.; Stourac, J.; Martinek, T.; Bednar, D.; Damborsky, J. HotSpot Wizard 3.0: Web Server for Automated Design of Mutations and Smart Libraries Based on Sequence Input Information. *Nucleic Acids Res.* **2018**, *46*, W356–W362. [CrossRef]

Supporting Information

Design and characterization of in-one protease-esterase PluriZyme

Laura Fernandez-Lopez[1,†], Sergi Roda[2,†], Jose L. Gonzalez-Alfonso[1], Francisco J. Plou[1], Víctor Guallar[2,*] and Manuel Ferrer[1*]

Content

Table S1: Absolute and relative number of accepted and total catalytic events (at 3.75 Å, 4.25 Å, and 5 Å) of EH1$_{AB1C}$ from the local exploration for the different dipeptide substrates (A) and L-carnosine (B).

A)

Dipeptide substrate	Number of accepted catalytic events [3.75 Å]	Number of total catalytic events [3.75 Å]	Number of accepted catalytic events [4.25 Å]	Number of total catalytic events [4.25 Å]	Number of accepted catalytic events [5 Å]	Number of total catalytic events [5 Å]
AH	129 (0.265 %)	2121 (0.422 %)	443 (0.91 %)	7708 (1.533 %)	779 (1.6 %)	12349(2.455 %)
AQ	17 (0.04 %)	97 (0.019 %)	105 (0.248 %)	1709 (0.341 %)	421 (0.995 %)	5433 (1.083 %)
DI	3 (0.012 %)	6 (0.001 %)	68 (0.279 %)	1103 (0.22 %)	144 (0.59 %)	1597 (0.318 %)
EA	14 (0.034 %)	101 (0.02 %)	584 (1.416 %)	2887 (0.571 %)	1200 (2.91 %)	11398 (2.255 %)
FF	0	0	7 (0.025 %)	58 (0.012 %)	41 (0.145 %)	1468 (0.296 %)
KA	5 (0.013 %)	10 (0.002 %)	412 (1.093 %)	2840 (0.581 %)	1197 (3.177 %)	11745 (2.402 %)
LA	14 (0.033 %)	77 (0.015 %)	777 (1.858 %)	3748 (0.74 %)	1163 (2.781 %)	7051 (1.391 %)
LL	0	0	17 (0.044 %)	124 (0.025 %)	67 (0.172 %)	621 (0.124 %)
NV	0	0	7 (0.018 %)	49 (0.01 %)	371 (0.935 %)	3211 (0.638 %)
PF	58 (0.171 %)	1294 (0.259 %)	495 (1.456 %)	6355 (1.27 %)	657 (1.933 %)	7984 (1.596 %)
QQ	20 (0.065 %)	233 (0.047 %)	128 (0.415 %)	2656 (0.536 %)	295 (0.955 %)	4529 (0.914 %)
RG	9 (0.029 %)	36 (0.007 %)	171 (0.553 %)	2301 (0.465 %)	1016 (3.286 %)	9615 (1.941 %)
SW	47 (0.147 %)	1267 (0.256 %)	444 (1.393 %)	4699 (0.949 %)	513 (1.609 %)	5177 (1.045 %)
TM	38 (0.095 %)	570 (0.112 %)	404 (1.014 %)	7220 (1.418 %)	635 (1.594 %)	10478 (2.057 %)
YN	20 (0.081 %)	297 (0.06 %)	56 (0.227 %)	838 (0.17 %9	72 (0.292 %)	950 (0.192 %)
YY	0	0	0	0	0	0

B)

Dipeptide substrate	Number of accepted catalytic events [3.75 Å]	Number of total catalytic events [3.75 Å]	Number of accepted catalytic events [4.25 Å]	Number of total catalytic events [4.25 Å]	Number of accepted catalytic events [5 Å]	Number of total catalytic events [5 Å]
L-carnosine	26 (0.102 %)	433 (0.098 %)	166 (0.654 %)	2344 (0.531 %)	425 (1.675 %)	5082 (1.151 %)

Table S2: predicted ΔΔG$_{(mut-WT)}$ of the EH1$_{AB1C}$ and its alternative calculated using the module of thermodynamic stability from HotSpot Wizard (see reference [45]) in both EH1$_A$ and EH1$_{AB1}$ crystal structures.

Variant	ΔΔG$_{(mut-WT)}$ [5JD4]	ΔΔG$_{(mut-WT)}$ [6RB0]
EH1$_{AB1C}$ (L24C)	6	4.2
EH1$_{AB1C*}$ (L24C/V36H)	6.8	31.4

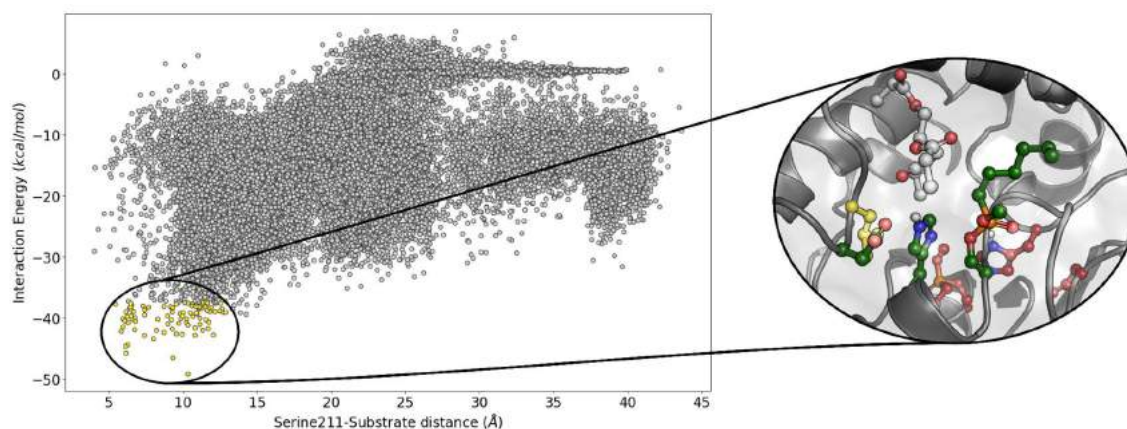**Figure S1.** Global exploration of the potential hydrolase sites in the EH1AB1 structure with the methyl hydrogen (R)-hexylphosphonate inhibitor bound to both catalytic serine residues. Accepted PELE steps around, what we called site C, are highlighted in the energetic profile with a yellow color. On the right, we represent a binding pose of the probe ester in site C. The main active site has the C atoms stained in maroon, the artificial active site has them stained in dark green, and the potential residue to mutate to cysteine to add a protease site has them stained in yellow. The energy profile was created with the Matplotlib library (see reference [28]).
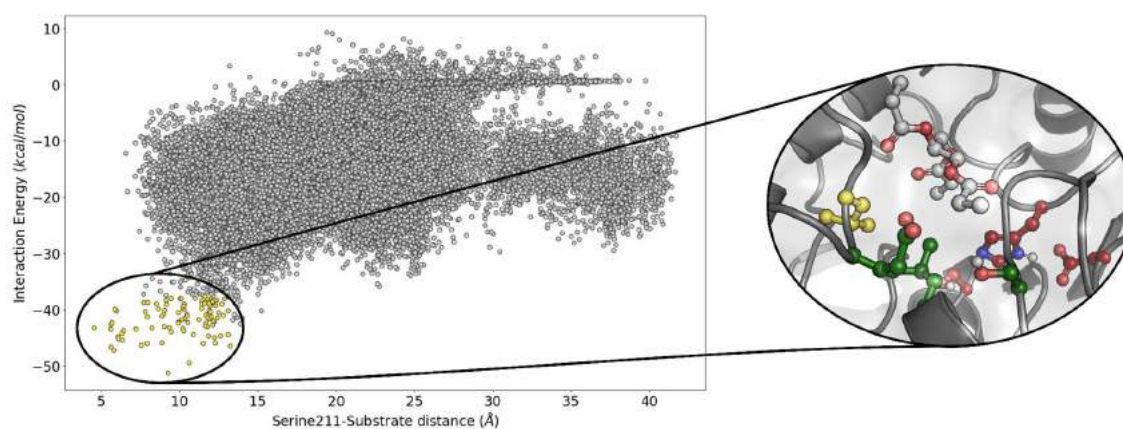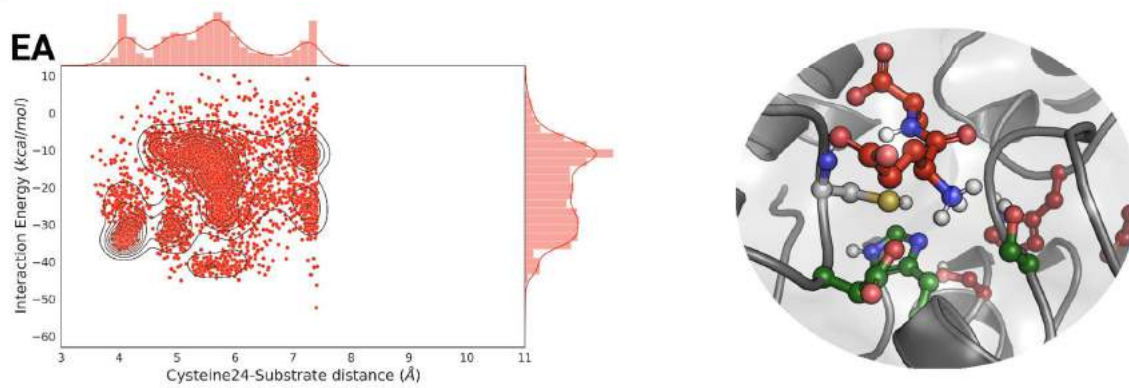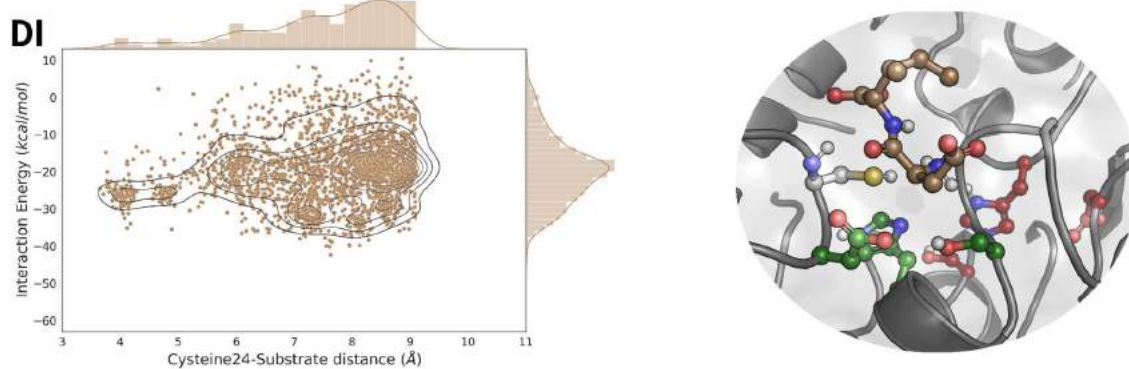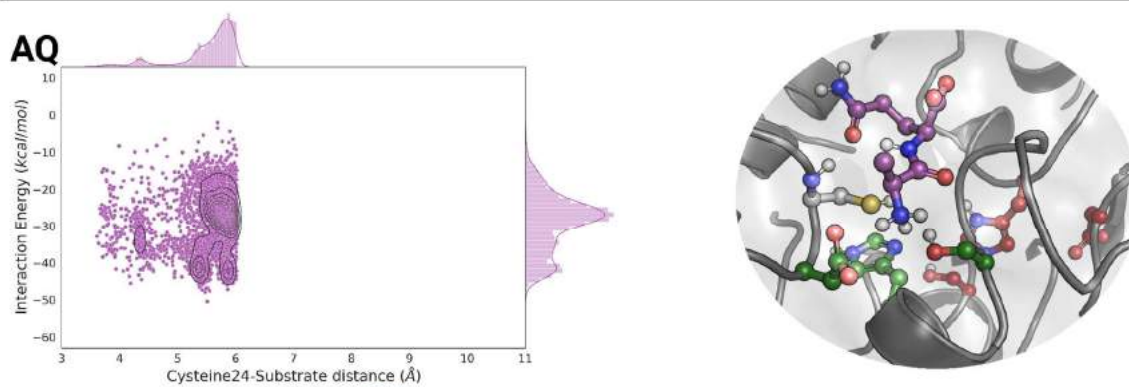

Figure S2. Global exploration of the potential hydrolase sites in the EH1A structure. Accepted PELE steps around, what we called site C, are highlighted in the energetic profile with a yellow color. On the right, we represent a binding pose of the probe ester in site C. The main active site has the C atoms stained in maroon, the artificial active site has them stained in dark green, and the potential residue to mutate to cysteine to add a protease site has them stained in yellow. The energy profile was created with the Matplotlib library (see reference [28]).

Figure S3. EH$_{1AB1C}$ density plots of the distribution of the catalytic cysteine-substrate distance against the interaction energy for the other simulated dipeptide substrates. Only the 10% lowest percentile regarding the distance is shown. On the right, we represent a catalytic pose of the dipeptide substrate in the protease site. The main active site has the C atoms stained in maroon, the artificial active site has them stained in dark green, the cysteine residue from the protease site has them stained in yellow, and each substrate has them stained in a particular color. The energy profiles were created with the Matplotlib library (see reference [28]).
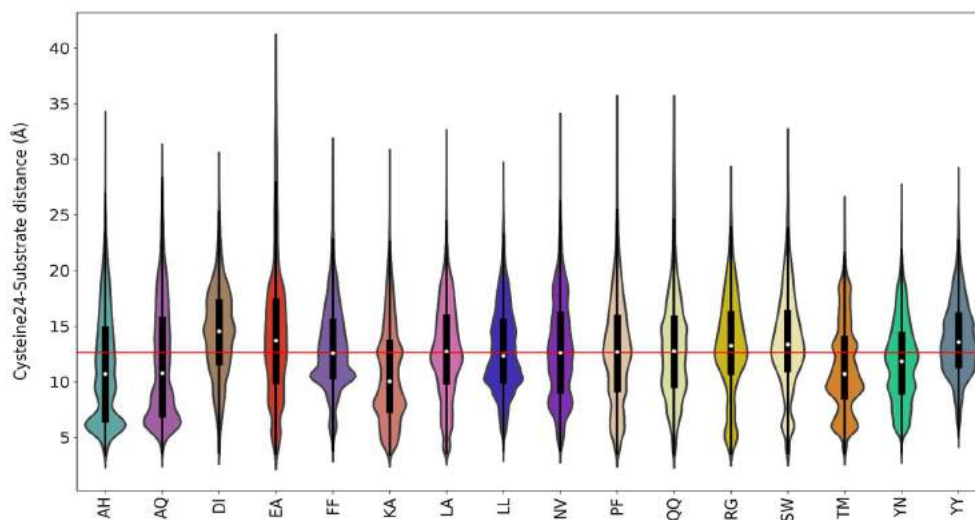
Figure S4. Violin plot representing the cysteine-substrate distance along all the accepted PELE steps from the local explorations for the different dipeptide substrates against EH$_{1AB1C}$. The red line indicates the average value of the metric in all simulations. The figure was created with the Matplotlib library (see reference [28]).
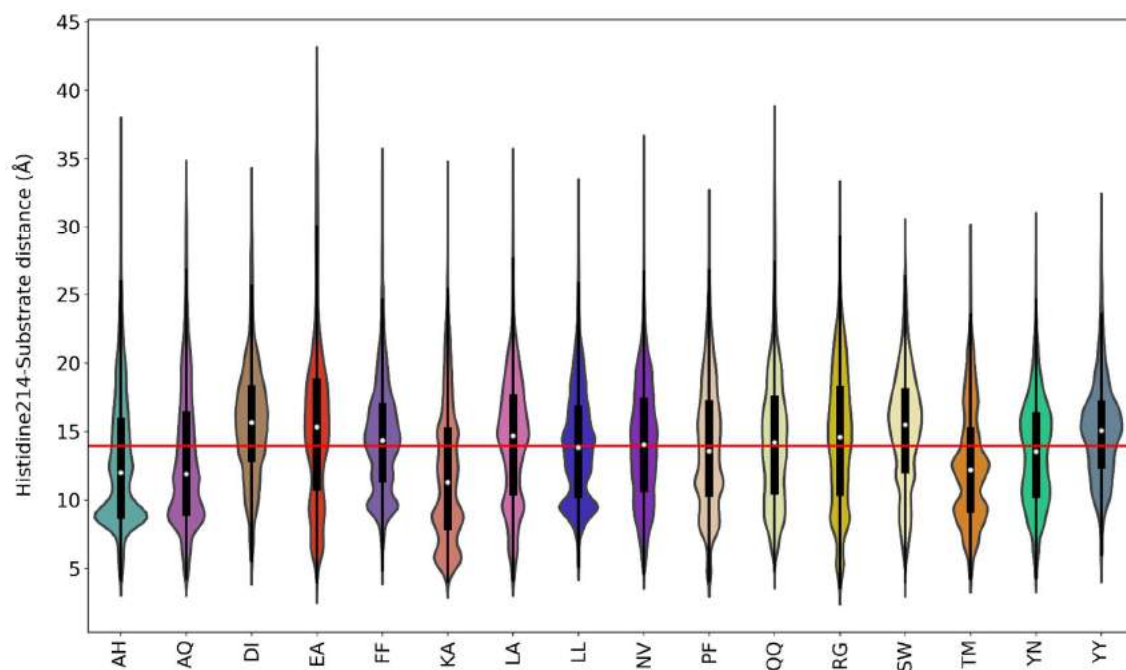
Figure S5. Violin plot representing the histidine-substrate distance along all the accepted PELE steps from the local explorations for the different dipeptide substrates against EH$_{1AB1C}$. The red line indicates the average value of the metric in all simulations. The figure was created with the Matplotlib library (see reference [28]).
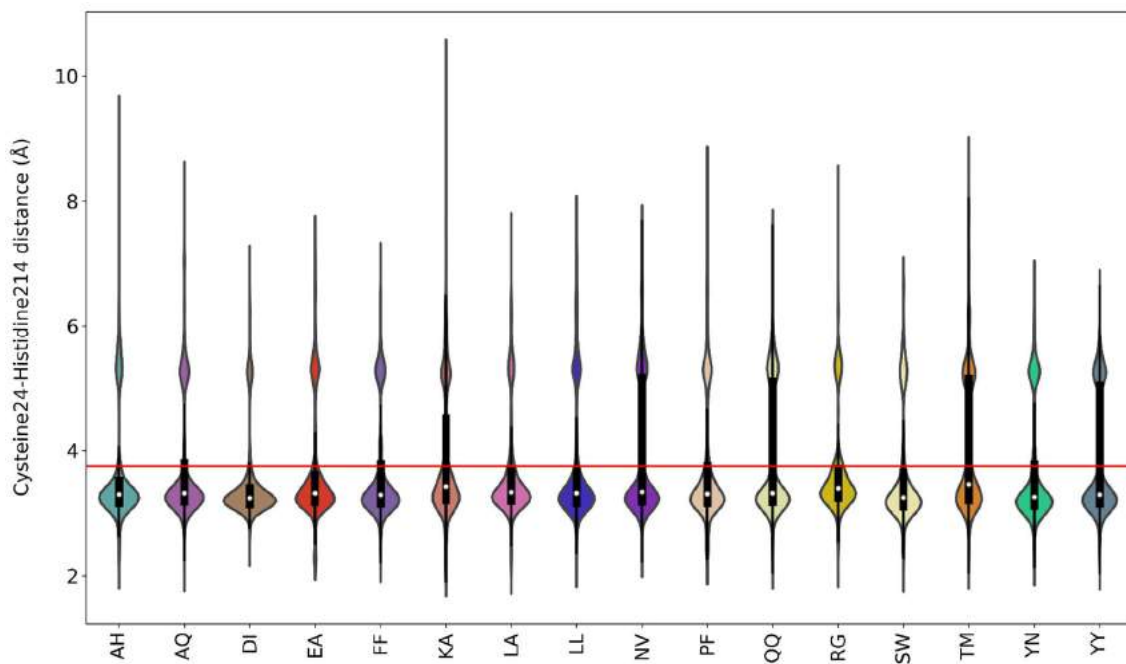
Figure S6. Violin plot representing the cysteine-histidine distance along all the accepted PELE steps from the local explorations for the different dipeptide substrates against EH$_{1AB1C}$. The red line indicates the average value of the metric in all simulations. The figure was created with the Matplotlib library (see reference [28]).
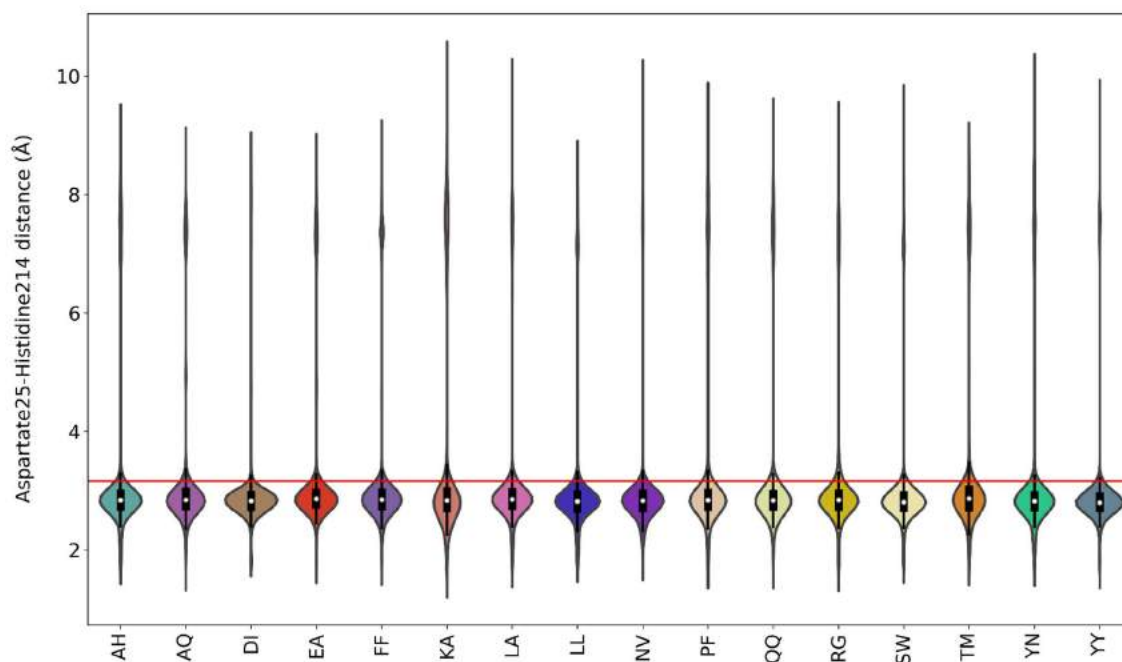
Figure S7. Violin plot representing the aspartate-histidine distance along all the accepted PELE steps from the local explorations for the different dipeptide substrates against EH_{1AB1C}. The red line indicates the average value of the metric in all simulations. The figure was created with the Matplotlib library (see reference [28]).
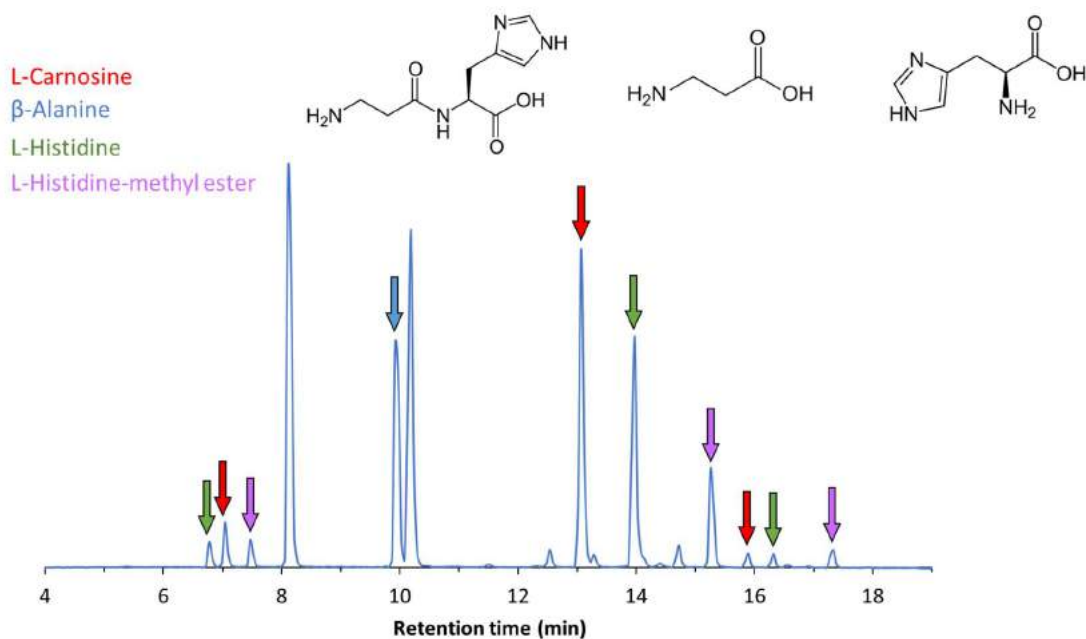
Figure S8. HPLC chromatograms representing the elution time of all substrates, intermediates and final products identified. As shown after derivatization some of the chemicals elute at different times (demonstrated by analysing each single chemical; not shown) and for the calculation of the concentration and conversion for each of them, the areas of each peak representing each chemical, were considered.
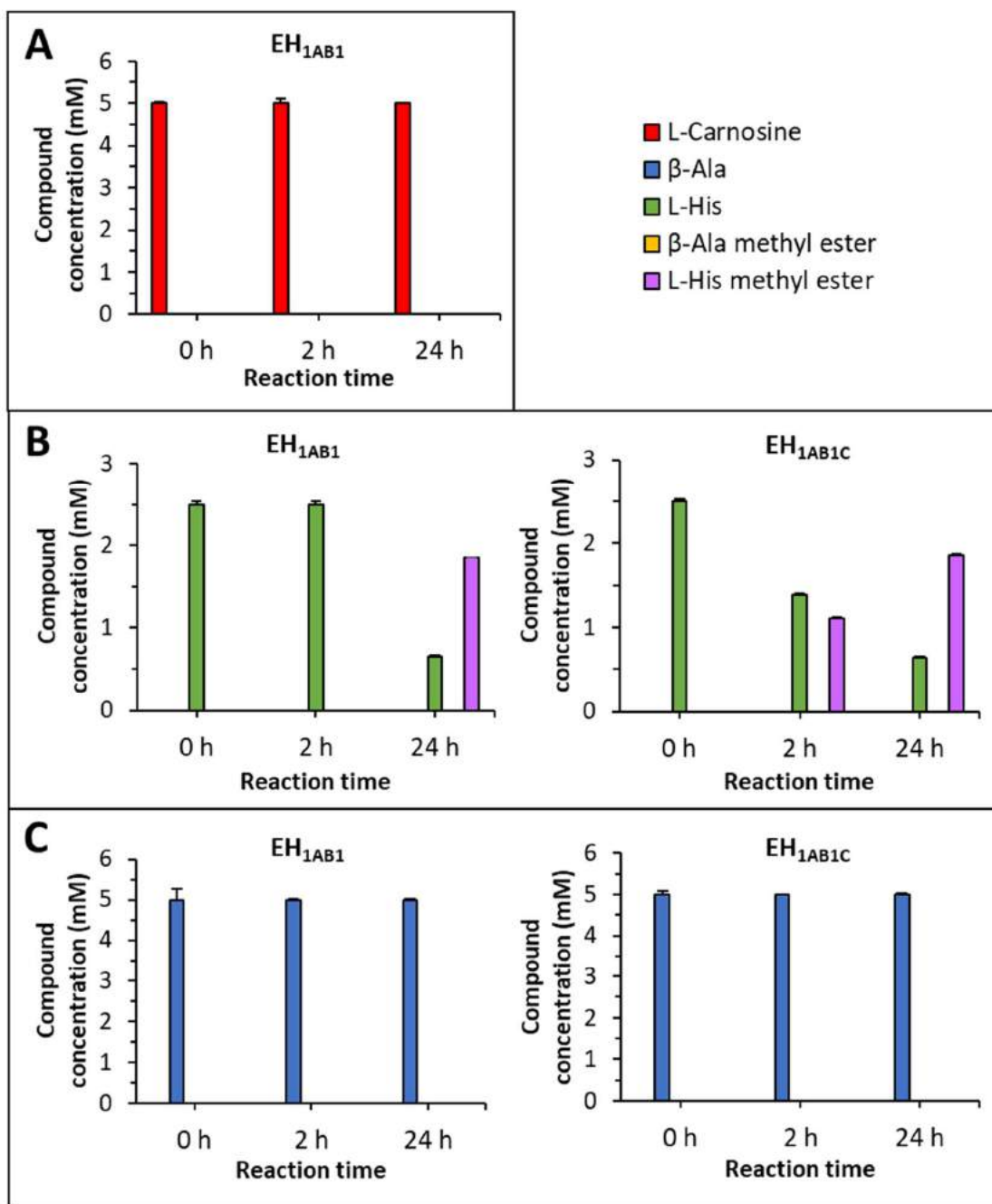
Figure S9. Concentrations of substrates and products obtained for the conversion of L-carnosine (A), L-histidine (B) and β-alanine (C) with whole cells expressing EH$_{1AB1}$ *PluriZyme* (left panels) or EH$_{1AB1C}$ *PluriZyme* (right panels) in the presence of methanol. The figure was created using SigmaPlot 14.0 software. As can be seen in panel A, the original EH$_{1AB1}$ *PluriZyme* does not hydrolyse L-carnosine oes not esterify it. The results in panels B and C demonstrate that both the original EH$_{1AB1}$ *PluriZyme* and the mutant EH$_{1AB1C}$ PluriZyme esterify L-histidine, but not β-alanine.

10
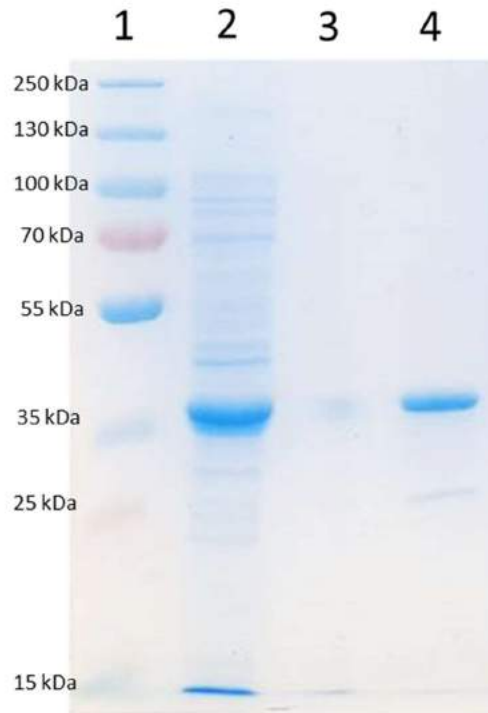
Figure S10. SDS-PAGE gel showing steps of EH$_{1AB1C}$ purification. Twelve percent SDS-PAGE gel stained with Coomassie blue. Lane 1 contains molecular weight markers. Lane 2 contains whole cell lysate of *E. coli* expressing soluble protein. Lane 3 shows the whole cell lysate of *E. coli* expressing insoluble protein. Lane 4 shows the His-tagged protein and demonstrates protein purity of >95%.

(A)

pH	Fluorescence 485/20,528/20 EH₁AB1C	
3	4	8
3,5	12	14
4	20	22
4,5	23	30
5	30	35
5,5	68	75
6	122	105
6,5	133	133
7	174	180
7,5	145	146
8	115	132
8,5	109	122

pH	Relative activity (%) EH₁AB1C (average % and SD)	
3	3,4	1,6
3,5	7,4	0,8
4	11,9	0,8
4,5	15	2,8
5	18,4	2
5,5	40,4	2,8
6	64,1	6,8
6,5	75,1	0
7	100	2,4
7,5	82,2	0,4
8	69,8	6,8
8,5	65,3	5,2

(B)

T (°C)	Absorbance at 440 nm EH₁AB1C	
20	0,024	0,027
30	0,054	0,055
40	0,076	0,067
50	0,096	0,096
60	0,135	0,134
70	0,143	0,141
75	0,137	0,147
80	0,138	0,142
85	0,128	0,124
90	0,0112	0,0112

T (°C)	Relative activity (%) (average % and SD)	
20	18	1,5
30	38,4	0,5
40	50,4	4,5
50	67,6	0
60	94,7	0,5
70	100	1
75	100	5
80	98,6	2
85	88,7	2
90	7,8	0

(C)
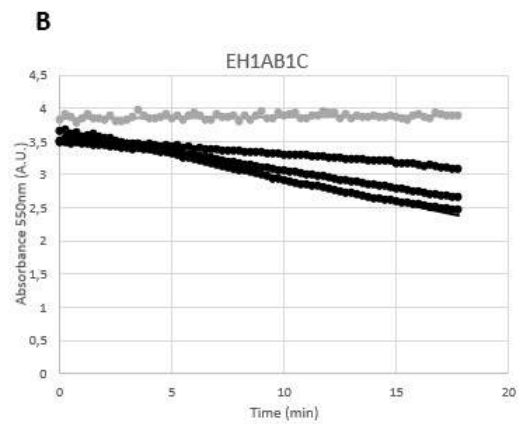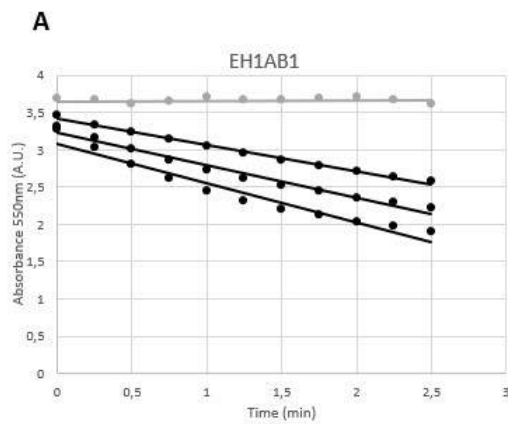
(D)

- Hydrolysis (buffer, no methanol) raw data (Figure 6A), in the presence of EH_{1AB1C}.

Area - Average				
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	66045886.33	0	0	0
1	56405937.67	9510840.333	12435133.33	0
2	20673479.67	18079872	26227694	0
24	5599927	34623204	56609214.33	0

Area - Deviation				
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	925255.158	0	0	0
1	4081705.52	129504.134	1999902.55	0
2	3855347.19	468065.626	3379406.28	0
24	1165884.48	620605.567	620605.567	0

Sample concentration (μM)						
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester	Total	Total (residues)
0	417.6392355	0	0	0	417.6392355	835.2784709
1	356.6813013	124.2711031	103.7705251	0	584.7229295	941.4042308
2	130.7281456	236.2362902	218.868709	0	585.8331447	716.5612903
24	35.41097502	452.3957509	472.4008773	0	960.2076032	995.6185782

SD (μM)				
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	5.850824	0	0	0
1	25.8105457	1.69213456	16.6890802	0
2	24.3791755	6.1158667	28.2009654	0
24	7.37243649	8.10899308	5.17892039	0

Sample concentration in reaction medium (mM) -Normalized to (5 mM carnosine)						SD mM				
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester	Total	Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0 h	5.000	0.000	0.000	0	5.000	0	0.035	0.000	0.000	0.000
1 h	3.789	1.320	1.102	0	6.211	1	0.137	0.018	0.177	0.000
2 h	1.824	3.297	3.054	0	8.176	2	0.170	0.085	0.394	0.000
24 h	0.356	4.544	4.745	0	9.644	24	0.037	0.081	0.052	0.000

13

- Cascade reaction (with metanol) raw data (Figure 6B), in the presence of EH$_{1AB1C}$.

	Area - Average			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	72916723.33	0	0	0
1	55722512.67	9406029	7916764	4090492
2	21611397	19710882	17615182.33	8929280.333
24	4923047	34745890.67	16872401.67	26284110.33

	Area - Deviation			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	11167492.4	0	0	0
1	6726467.34	34811.0622	1241768.93	844628.1463
2	4946823.13	264351.172	2486694.35	2122136.349
24	1081449.47	369638.206	369638.206	5926748.901

	Sample concentration (µM)						
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester	Total	Total (residues	Total His+ His-ME
0	461.0867728	0	0	0	461.0867728	922.1735455	0
1	352.3596832	122.9016111	66.06497376	34.13493779	575.4612058	893.6859513	100.1999115
2	136.6590385	257.5474893	146.997758	74.51436861	615.7186544	677.8633242	221.5121266
24	31.13074408	453.9988066	140.7992929	219.3395002	845.2683439	657.0595877	360.1387932

	SD (µM)			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	70.6173123	0	0	0
1	42.53462	0.45485036	10.3624956	7.048376877
2	31.2810918	3.45408088	20.7513319	17.70911476
24	6.83851415	4.82978853	3.08461113	49.45840379

Sample concentration in reaction medium (mM) -Normalized to (5 mM carnosine)					
Reaction time (h)	L- Carnosine	β-Ala	His	His-methyl ester	Total
0 h	5.000	0.000	0.000	0.000	5.000
1 h	3.943	1.229	0.810	0.419	6.401
2 h	2.016	2.575	1.709	0.866	7.167
24 h	0.474	4.540	1.775	2.765	9.554

SD mM				
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0.383	0.000	0.000	0.000
1	0.238	0.005	0.104	0.070
2	0.231	0.035	0.208	0.177
24	0.052	0.048	0.031	0.495

14

(E)

- Cascade reaction (with metanol) raw data, in the presence of EH_{1AB1} (Figure S9A).

Reaction time (h)	Area - Average			
	Carnosine	β-Ala	His	His-methyl ester
0	74291435,33	0	0	0
2	62687474,33	0	0	0
20	52139788	0	0	0

Reaction time (h)	Area - Deviation				
	Carnosine	β-Ala	His	β-Ala methyl ester	His-methyl ester
0	609950,783	0	0	0	0
2	3036089,87	0	0	0	0
20	324015,542	0	0	0	0

Reaction time (h)	Sample concentration (µM)					
	Carnosine	β-Ala	His	His-methyl ester	β-Ala methyl ester	Total
0	469,779724	0	0	0	0	469,779724
2	396,4024151	0	0	0	0	396,4024151
20	329,7044283	0	0	0	0	329,7044283

Reaction time (h)	SD (µM)				
	Carnosine	β-Ala	His	β-Ala methyl ester	His-methyl ester
0	3,85700598	0	0		0
2	19,1986257	0	0		0
20	2,04890283	0	0		0

Reaction time (h)	Sample concentration in reaction medium (mM) -Normalized to (5 mM carnosine)					
	L-Carnosine	β-Ala	L-His	L-His methyl ester	β-Ala methyl ester	Total
0	5,000	0,000	0,000	0		5,000
2	5,000	0,000	0,000	0		5,000
20	5,000	0,000	0,000	0		5,000

Reaction time (h)	SD mM				
	Carnosine	β-Ala	His	β-Ala methyl ester	His-methyl ester
0	0,021	0,000	0,000	0	0,000
2	0,121	0,000	0,000	0	0,000
20	0,016	0,000	0,000	0	0,000

- L- Histidine reaction (with metanol) raw data in the presence of EH_{1AB1} (Figure S9B).

Reaction time (h)	Area - Average			
	Carnosine	β-Ala	His	His-methyl ester
0	0	0	35535189.67	0
2	0	0	31907237	0
24	0	0	12867910.33	36645382

Reaction time (h)	Area - Deviation			
	Carnosine	β-Ala	His	His-methyl ester
0	0	0	575075.85	0
2	0	0	475937.293	0
24	0	0	392832.675	189697.3025

Reaction time (h)	Sample concentration (µM)					
	Carnosine	β-Ala	His	His-methyl ester	Total	Total (residuos)
0	0	0	296.5392644	0	296.5392644	296.5392644
2	0	0	266.2641927	0	266.2641927	266.2641927
24	0	0	107.3820261	305.8037602	413.1857863	413.1857863

Reaction time (h)	SD (µM)			
	Carnosine	β-Ala	His	His-methyl ester
0	0	0	4.79897732	0
2	0	0	3.97167135	0
24	0	0	3.27816774	1.583013881

Sample concentration in reaction medium (mM) -Normalized to (2.5 mM His)					
Reaction time (h)	L-Carnosine	β-Ala	L-His	L- His-methyl ester	Total
0 h	0.000	0.000	2.500	0.000	2.500
2 h	0.000	0.000	2.500	0.000	2.500
24 h	0.000	0.000	0.650	1.850	2.500

SD mM				
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0.000	0.000	0.040	0.000
2	0.000	0.000	0.037	0.000
24	0.000	0.000	0.020	0.010

- - Histidine reaction (with metanol) raw data in the presence of EH_{1AB1C} (Figure S9B).

	Area - Average			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0	0	35111999	0
2	0	0	25123322.67	19989914.67
24	0	0	14500647.67	41832287.33

	Area - Deviation			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0	0	401282.111	0
2	0	0	282187.094	162377.1727
24	0	0	87384.3772	297615.2006

	Sample concentration (µM)					
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester	Total	Total (residues)
0	0	0	293.0077608	0	293.0077608	293.0077608
2	0	0	209.652789	166.8147728	376.4675618	376.4675618
24	0	0	121.0071321	349.0882089	470.095341	470.095341

	SD (µM)			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0	0	3.34867783	0
2	0	0	2.35483627	1.355028854
24	0	0	0.72921797	2.483582991

Sample concentration in reaction medium (mM) -Normalized to (2.5 mM His)					
Reaction time (h)	L-Carnosine	β-Ala	L-His	L- His-methyl ester	Total
0 h	0.000	0.000	2.500	0.000	2.500
2 h	0.000	0.000	1.392	1.108	2.500
24 h	0.000	0.000	0.644	1.856	2.500

SD mM				
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0.000	0.000	0.029	0.000
2	0.000	0.000	0.016	0.009
24	0.000	0.000	0.004	0.013

- β-Alanine reaction (with metanol) raw data in the presence of EH_{1AB1} (Figure S9C).

	Area - Average			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0	38621021.33	0	0
2	0	34250592	0	0
24	0	35603589.33	0	0

	Area - Deviation			
Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
0	0	1955503.02	0	0
2	0	112858.066	0	0
24	0	33521.6922	0	0

Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester	Total	Total (residues)	Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
		Sample concentration (µM)							SD (µM)		
0	0	504.6322676	0	0	504.6322676	504.6322676	0	0	25.5511089	0	0
2	0	447.527106	0	0	447.527106	447.527106	2	0	1.47463272	0	0
24	0	465.2057195	0	0	465.2057195	465.2057195	24	0	0.43800311	0	0

Reaction time (h)	L-Carnosine	β-Ala	L-His	L- His-methyl ester	Total	Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
	Sample concentration in reaction medium (mM) -Normalized to (5 mM Ala)							SD mM		
0 h	0.000	5.000	0.000	0.000	5.000	0	0.000	0.253	0.000	0.000
2 h	0.000	5.000	0.000	0.000	5.000	2	0.000	0.016	0.000	0.000
24 h	0.000	5.000	0.000	0.000	5.000	24	0.000	0.005	0.000	0.000

- β-Alanine reaction (with metanol) raw data in the presence of EH$_{1AB1C}$ (Figure S9C).

Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester	Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
		Area - Average					Area - Deviation		
0	0	37663773.33	0	0	0	0	715050.418	0	0
2	0	34222480	0	0	2	0	67054.195	0	0
24	0	33938377.33	0	0	24	0	247115.117	0	0

Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester	Total	Total (residues)	Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
		Sample concentration (µM)							SD (µM)		
0	0	492.1246173	0	0	492.1246173	492.1246173	0	0	9.34303396	0	0
2	0	447.1597873	0	0	447.1597873	447.1597873	2	0	0.87614748	0	0
24	0	443.4476283	0	0	443.4476283	443.4476283	24	0	3.22887011	0	0

Reaction time (h)	L-Carnosine	β-Ala	L-His	L- His-methyl ester	Total	Reaction time (h)	Carnosine	β-Ala	His	His-methyl ester
	Sample concentration in reaction medium (mM) -Normalized to (5 mM Ala)							SD mM		
0 h	0.000	5.000	0.000	0.000	5.000	0	0.000	0.095	0.000	0.000
2 h	0.000	5.000	0.000	0.000	5.000	2	0.000	0.010	0.000	0.000
24 h	0.000	5.000	0.000	0.000	5.000	24	0.000	0.036	0.000	0.000

Raw Dataset. (A) Raw fluorescence data (corresponding to Figure 3A) after the hydrolysis of BODIBY FL casein by EH$_{1AB1C}$, at different pH values (corrected by background signal). Reaction conditions as described in Section 4.5. (B) Raw absorbance data (corresponding to Figure 3B) after the hydrolysis of azocasein by EH$_{1AB1C}$, at different temperatures. Reaction conditions as described in Section 4.5. (B) Representative time-course curve for the hydrolysis of glyceryl tripropionate by EH$_{1AB1C}$ (corrected by background signal). Shown are the raw data (absorbance at 550 nm over time), corresponding to calculations of specific activity. Reaction conditions as described in Section 4.5. (A) Representative time-course curve for the hydrolysis of glyceryl tripropionate by EH$_{1AB1C}$ andEH$_{1AB1}$. Shown are the raw data (absorbance at 550 nm over time), corresponding to calculations of specific activity. Reaction conditions as described in Section 4.5. (D)

Raw data and calculations corresponding to Figure 5 (cascade reaction with EH$_{1AB1C}$). (E) Raw data and calculation for the control cascade reaction in the presence of EH$_{1AB1C}$; as shown no reactions products were found.

GDCh

Research Articles

Angewandte
Chemie
International Edition
www.angewandte.org

Check for updates

Protein Engineering

A Plurizyme with Transaminase and Hydrolase Activity Catalyzes Cascade Reactions

*Sergi Roda[+], Laura Fernandez-Lopez[+], Marius Benedens, Alexander Bollinger,
Stephan Thies, Julia Schumacher, Cristina Coscolín, Masoud Kazemi, Gerard Santiago,
Christoph G. W. Gertzen, Jose L. Gonzalez-Alfonso, Francisco J. Plou, Karl-Erich Jaeger,
Sander H. J. Smits, Manuel Ferrer,* and Víctor Guallar**

Abstract: Engineering dual-function single polypeptide catalysts with two abiotic or biotic catalytic entities (or combinations of both) supporting cascade reactions is becoming an important area of enzyme engineering and catalysis. Herein we present the development of a *PluriZyme*, TR_2E_2, with efficient native transaminase (k_{cat}: 69.49 ± 1.77 min^{-1}) and artificial esterase (k_{cat}: 3908–0.41 min^{-1}) activities integrated into a single scaffold, and evaluate its utility in a cascade reaction. TR_2E_2 (pH$_{opt}$: 8.0–9.5; T_{opt}: 60–65 °C) efficiently converts methyl 3-oxo-4-(2,4,5-trifluorophenyl)butanoate into 3-(R)-amino-4-(2,4,5-trifluorophenyl)butanoic acid, a crucial intermediate for the synthesis of antidiabetic drugs. The reaction proceeds through the conversion of the β-keto ester into the β-keto acid at the hydrolytic site and subsequently into the β-amino acid (e.e. > 99 %) at the transaminase site. The catalytic power of the TR_2E_2 *PluriZyme* was proven with a set of β-keto esters, demonstrating the potential of such designs to address bioinspired cascade reactions.

[*] S. Roda,[+] Dr. M. Kazemi, Dr. G. Santiago, Prof. V. Guallar
Department of Life Sciences,
Barcelona Supercomputing Center
Carrer de Jordi Girona, 31, 08034 Barcelona (Spain)
E-mail: victor.guallar@bsc.es

L. Fernandez-Lopez,[+] Dr. C. Coscolín, J. L. Gonzalez-Alfonso,
Prof. F. J. Plou, Prof. M. Ferrer
Department of Applied Biocatalysis, ICP, CSIC
Marie Curie 2, 28049 Madrid (Spain)
E-mail: mferrer@icp.csic.es

M. Benedens, Dr. J. Schumacher, Dr. C. G. W. Gertzen,
Prof. S. H. J. Smits
Center for Structural Studies,
Heinrich-Heine-University, Building 26.44.01.62
Universitaetsstr 1, 40228 Duesseldorf (Germany)

Dr. A. Bollinger, Dr. S. Thies, Prof. K.-E. Jaeger
Institute of Molecular Enzyme Technology, Heinrich-Heine-Universität Düsseldorf, Building 26.44.01.62
Universitaetsstr 1, 40228 Duesseldorf (Germany)
and
Forschungszentrum Jülich, Building 15.8, 01/303, 52428
Wilhelm Johnen Straße, Jülich (Germany)

Prof. V. Guallar
Institució Catalana de Recerca i Estudis Avançats
Passeig de Lluís Companys, 23, 08010 Barcelona (Spain)

[+] These authors contributed equally to this work.

Introduction

Microorganisms are regarded as self-replicating containers of enzymes that catalyze enzyme cascades (or biosynthetic pathways), yielding complex products.[1] Cascade reactions are also relevant in the field of biocatalysis for the synthesis and assembly of numerous molecules.[2] In vitro cascades have been investigated with at least two enzymatic reactions combined with a chemical step[3] or purely enzymatic systems, excluding enzymes required for the regeneration of cofactors or the removal of poisonous side products.[2] Thus, multienzyme systems and biomimetic or bioinspired architectures have been built to spatially organize different enzymes with nanometer precision.[4–6] Using this approach, the transport of intermediates between enzymes in close proximity, known as proximity channeling, is efficiently controlled, thus favoring the progression of chain reactions.[7,8] Recent examples illustrating the design of microfluidic electrospray microcapsules mimicking natural subcellular compartments, such as organelles or organs,[9] or onion-like photonic spheres,[10] demonstrate the relevance of systems that minimize the diffusion of intermediates between enzymes, enhancing the overall efficiency of bio-reactions. At the nanoscale, artificial chimeric proteins in which two modules, domains, or enzymes, each being different concerning the chemistry or reaction step catalyzed, are joined via a linker have also been developed. An example is the design of enzyme chimeras in which an endoglucanase and a β-glucosidase are linked to support polysaccharide degradation.[11] Other examples include the design of two-enzyme polyethylene terephthalate depolymerization systems for plastic upcycling[12] and the fusion of enzymes integrating terpene synthases.[13]

The design of dual-function catalysts characterized by the precise positioning of two abiotic[14] or biotic[15] catalytic entities (or a combination of both),[16] that act in concert to catalyze synergetic reactions, represent complementary alternatives to the traditional two-catalysts systems (either mixed or fused[11–13]). They have provided new solutions for activating and transforming specific molecules involving several steps, or even new-to-nature reactions.[14–16] The challenge originates from the difficulty in precisely controlling the positioning of the catalytic entities in a single protein scaffold, which is different from combining the active centers of two covalently bound enzymes.[11–13] The catalytic potential of an artificial protein scaffold with two abiotic catalytic groups has been recently demonstrated by the design of a chimeric streptavidin with two adjacent Au^I complexes structurally equal. Both gold entities can work individually but, through adopting multiple poses, can work in synergy to activate an alkyne.[14,17] In a recent elegant study, the incorporation of a serine close to the haem cofactor of a P450 enzyme allowed the design of a dual-function catalyst for efficient enantioselective carbene C–H insertion reactions.[15] The serine residue contributes to controlling the orientation and activation of the reaction intermediates. Other examples of artificial protein scaffolds with combined abiotic and biotic catalytic groups supporting tandem catalysis have been reported.[16] These recent studies exemplify the interest in designing multi-function enzymes for synthetic chemistry. In this study, we went one step further and designed an enzyme with two biotic sites, one supporting hydrolase and one transaminase activity, that are independent catalytic entities but can also work in synergy (Scheme 1).

For this purpose, we used recently published computational tools to design *PluriZymes*,[18,19] the name we assigned to the concept of adding multiple active sites to the same enzyme structure. In detail, we report the addition of a catalytic triad (Ser, His, Asp/Glu) and an oxyanion hole supporting ester hydrolysis in a class III ω-transaminase (ω-TA), which is subsequently referred to as TR_2 (GenBank acc. nr. MH588437).[20] ω-TA is a well-established enzyme used to design artificial biocatalytic linear cascades to prepare organic molecules,[21] and thus, it was targeted in the present study. We designed a *PluriZyme*, hereafter referred to as TR_2E_2, able to catalyze the one-pot cascade synthesis of the β-amino acid (R)-3-amino-4-(2,4,5-trifluorophenyl)-butanoic acid (3-ATfBA), a key precursor in the synthesis of gliptins (Scheme 1).[22,23] Our results show that the design of this catalytically efficient *PluriZyme* is feasible and facilitates the generation of 3-ATfBA from the methyl 3-oxo-4-(2,4,5-trifluorophenyl)butanoate (3-OTfBE) β-keto ester with an excellent conversion rate (>99%; conc.: 14 mM) and enantioselectivity (e.e. >99%). Moreover, by using a range of β-keto esters, we further demonstrated that dual function *PluriZymes* are powerful catalysts for tandem reactions. In addition, we report the crystal structures of the TR_2 enzyme and the newly designed TR_2E_2 *PluriZyme*.

Results and Discussion

In this study, we targeted TR_2, a class III ω-TA (GenBank acc. nr. MH588437) isolated from the beach acidic pool on Vulcano Island and most likely derived from a bacterium of the *Acidihalobacter* genus.[20] TR_2 showed maximal transaminase activity at 45–55 °C, suggesting that it is moderately thermostable. It also showed a broad substrate range and efficiently converted bulky ketones and amines with (R) stereochemistry, which is rare among class III ω-TAs.

We aimed to crystalize and solve the structure of TR_2. We used sitting drop vapor diffusion and tested 2000
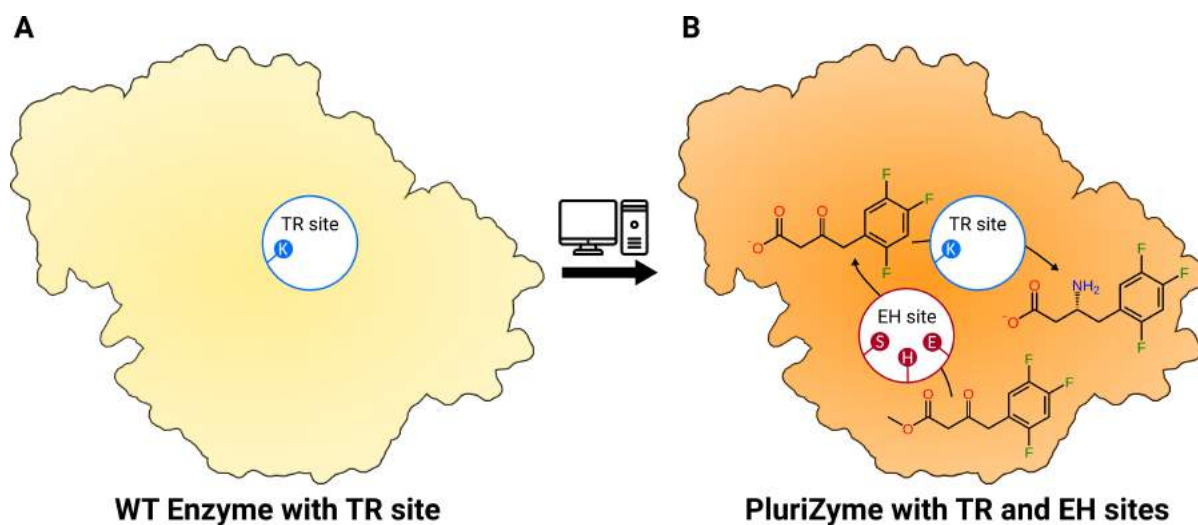
Scheme 1. The general concept for transforming a wild-type class III ω-TA (A) into an artificial *PluriZyme* with transaminase (TR) and ester hydrolase (EH) activities (B). The cascade reaction supported by TR_2E_2, namely, the hydrolysis of 3-OTfBE to methyl 3-oxo-4-(2,4,5-trifluorophenyl)-butanoic acid (3-OTfBA) and its further transamination to (R)-3-ATfBA, is shown. Methyl 3-amino-4-(2,4,5-trifluorophenyl)butanoate (3-ATfBE) was detected as a minor by-product.

conditions from commercially available screens to achieve this goal. The obtained crystals were optimized in 24-well sitting drop crystallization plates. The crystals of TR_2 diffracted moderately, and a dataset was collected at 3.6 Å resolution (statistics for the data are presented in Table S1). The TR_2 crystals were phased by molecular replacement using PHASER in the Phenix program,[24] and a model calculated with AlphaFold as a template.[25] The structure was then refined in iterative cycles of manual building and refinement in Coot, followed by software-based refinements using the Phenix program suite.[26,27]

TR_2 shows an asymmetric unit of four peptide chains building a functional dimer (Figure 1A,B). It presents the typical type I fold of class III ω-TA observed in known structures present in the PDB database (Table S1),[28] and it contains the characteristic pyridoxal-5′-phosphate (PLP)-binding domain composed of β-sheets and the catalytic base Lys289. The TR_2 monomer is subdivided into a core domain, which is the PLP-binding domain, and a C-terminal sub-domain. The core domain is composed of a central β-sheet consisting of the strands β1–β7 surrounded by helices α1–α2 and the short helical structures η1 and η2. The smaller subdomain at the C-terminus consists of the three α-helices

and β-strands 8, 9, and 10. TR_2 is a functional dimer, as represented by Figure 1B; here, the tunnel for accessing the binding pocket with the catalytic base (red circle) is composed of the two monomers (highlighted in different colors).

The protocol used to design an artificial hydrolase site has been explained previously,[18,19] including two review articles.[29,30] Briefly, the process begins with scanning the transaminase surface to identify noncatalytic ester binding sites using global Protein Energy Landscape Exploration (PELE) exploration.[31] Next, we perform local explorations of active site variants to introduce a well-positioned catalytic triad while also considering the existence of oxyanion holes that are required for efficient ester hydrolysis.[32] Finally, optimum variants designed with PELE are ranked through i) Molecular Dynamics (MD) refinement simulations, accounting for the integrity of the triad, and ii) $\Delta\Delta G$ estimations using the HotSpot Wizard.

As shown in Figure 2, several potential sites for ester binding were located with global exploration. During the local exploration, we focused on the enzyme-substrate interaction energy (see Figure S1), the catalytic triad hydrogen bond lengths (see Figures S2–S5), and the occurrence
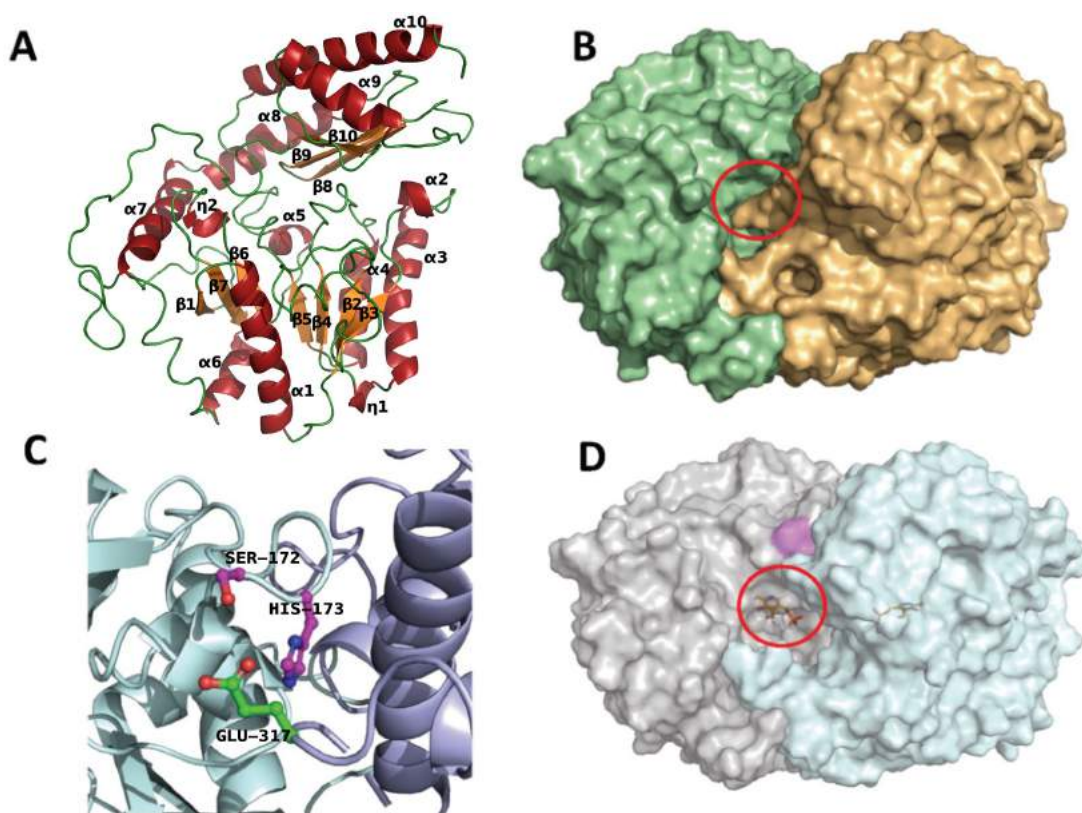

Figure 1. A) Secondary structural elements of TR_2, with α-helices (red), β-strands (orange) and loops (green) shown. B) Surface representation of the TR_2 dimer with highlighted transaminase binding site (red circle). C) Higher-magnification image of the TR_2E_2 structure at the Ser172 and His173 mutations. D) Surface representation of the TR_2E_2-PLP dimer with a highlighted transaminase binding site (red circle). PLP is located in the transaminase binding site and is shown as a ball-and-stick representation. The positions of the Ser172 and His173 mutations are colored magenta. The figure was created using PyMOL Version 2.3.2.
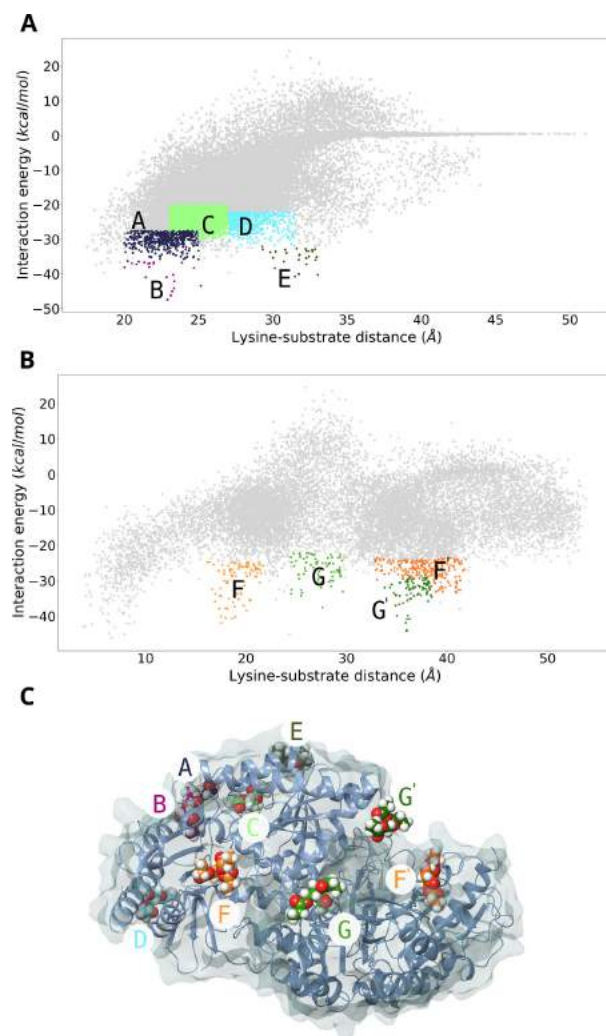
Angew. Chem. Int. Ed. **2022**, e202207344 (3 of 10)

© 2022 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

Figure 2. Global exploration of noncatalytic ester binding sites on the TR$_2$ surface with PELE. Energetic profiles for the monomer (A) and dimer (B) are shown. Representative PELE steps around the non-catalytic ester binding sites are highlighted in the energetic profiles with a particular color (site A in dark blue, site B in purple, site C in lime, site D in turquoise, site E in dark green, sites F/F′ in orange for chains A and B and sites G/G′ in green for chains A and B). C) 3D structure of TR$_2$ with representative binding poses of the probe ester in the different noncatalytic binding orientations. The surface is displayed in a transparent blue-white color, and the secondary structure of each monomer is shown as ribbons. Substrate molecules highlight C atoms using the energetic profile colors (underlying the noncatalytic ester binding sites). Energetic profiles were created with the Matplotlib library.

(density) of catalytic poses (see Table S2) to introduce the catalytic triad in each of the sites described above. Typically, we aimed for the ester carbon to be located within 3–4 Å from the serine nucleophilic oxygen and catalytic hydrogen bonds to be located within reasonable distances (3.5 Å). These distance parameters, the root mean square deviation (RMSD) of the overall structure, and the local RMSD of the main catalytic lysine and the designed triad (see Figures S6–

S8) were estimated from MD simulations. Then, a score ranking of all mutants was created from all these metrics and combined with the $\Delta\Delta G$ stability calculation (see Table S3); the average of each metric was compared to ideal values obtained from the simulations to obtain the score. Although, the $\Delta\Delta G$ stability calculation of the variants was just used to penalize or even discard the high destabilizing mutants ($\Delta\Delta G > 5$ kcal mol^{-1}), it has been proven to be crucial for enzyme design.[33–35] The final list contained 20 ranked mutants (only 6 were discarded due to drastic deviations in the metrics).

TR$_2$ and the 20 designed mutants (Table S4) were produced and purified, and their transaminase activity was quantified to evaluate whether any of the mutations were deleterious to the original transaminase activity. Of the 20 mutants examined, only six showed transaminase activity, retaining 58 to 92% of the original TR$_2$ activity (k_{cat}: 83.38 $\pm$ 2.52 min^{-1}) when using hexanal, a model aldehyde that was shown to be one of the preferred substrates for TR$_2$ (see Table S4, Figure S9).[20] These mutants contain from 1 to 3 mutations (Table S4) located at sites B, E, F, and G, as shown in Figure 2C.

All six mutants retaining transaminase activity were further examined for the presence of hydrolytic activity using model esters *p*-nitrophenyl acetate (*p*NPA), propionate (*p*NPP), and butyrate (*p*NPB). As shown in Table 1, only purified TR$_2$E$_2$ (Figure S10), which retained approximately 80% of the original transaminase activity (k_{cat}: 69.48 $\pm$ 1.77 min^{-1}), was capable of hydrolyzing *p*NPA, *p*NPP, and *p*NPB, with the latter being the preferred substrate (k_{cat}: 0.41 $\pm$ 0.05 min^{-1}). This variant contains two mutations, A172S and Q173H. We further observed that TR$_2$E$_2$ hydrolyzed 52 of a set of 96 chemically and structurally diverse esters tested (k_{cat}: 3908 to 1.2 min^{-1}) (Table S5). Under the same conditions, the activity for TR$_2$ was below the detection limit for all esters. When the residues Ser172 and His173 were replaced with Ala, the activity was completely abolished. Thus, the A172S and Q173H mutations in TR$_2$E$_2$ (located at site G in Figure 2) do not exert a deleterious effect on the original transaminase activity and, importantly, confer esterase activity.

Furthermore, we tested 3-OTfBE as a substrate. This ester is used to synthesize precursors of gliptins, namely, β-amino acids obtained through bienzymatic cascade reactions catalyzed by esterases and transaminases.[22,23] TR$_2$E$_2$ was shown to efficiently hydrolyze 3-OTfBE (k_{cat}: 32.47 $\pm$ 1.13 min^{-1}). Using 3-OTfBE and High Performance Liquid Chromatography (HPLC) determinations, TR$_2$E$_2$ showed maximal hydrolytic activity at pH ranging from 8.0 to 9.5 (Figure 3A) and temperatures ranging from 60 to 65 °C (Figure 3B). The optimal pH (ca. 8.5) and temperature (ca. 50–55 °C) for the transaminase activity of the variant TR$_2$E$_2$ were similar to those of TR$_2$ itself (Figure 4A and B). The finding that both TR$_2$ and TR$_2$E$_2$ retained high activity at 50–65 °C was consistent with their denaturation temperatures (56.23 ± 0.09 °C and 61.85 ± 0.11 °C, respectively), as revealed by circular dichroism spectroscopy (Figure 4C).

Diffracting crystals were obtained for TR$_2$E$_2$ as well. Datasets for each crystal were collected and refined to a

Table 1: Kinetic constants for the enzymatic activities of TR_2E_2.

Substrate	K_m [mM]	k_{cat} [min^{-1}]	k_{cat}/K_m [s^{-1} M^{-1}]
Hydrolytic activity[a,b]			
pNPA[a]	2.056 ± 0.239	0.323 ± 0.010	2.62 ± 0.70
pNPP[a]	2.664 ± 0.642	0.409 ± 0.051	2.57 ± 1.3
pNPB[a]	0.279 ± 0.096	0.112 ± 0.010	6.67 ± 1.75
3-OTfBE[b]	2.407 ± 0.038	32.47 ± 1.13	224.9 ± 11.4
3-ATfBE[b]	4.750 ± 0.503	0.131 ± 0.02	0.460 ± 0.12
Transaminase activity[c]			
3-OTfBE[c]	12.27 ± 0.47	0.771 ± 0.20	1.05 ± 0.07
3-OTfBA[c]	7.65 ± 0.35	92.91 ± 1.16	202.5 ± 11.8

Reaction conditions as detailed in Supporting Experimental Procedures: [a] K_m-[protein]: 4.5 µg mL^{-1}; [ester]: 0–7 mM; reaction volume: 200 µL; T: 30°C; pH: 7.0 (40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer). [b] K_m-[protein]: 4.5 µg mL^{-1}; [ester]: 0–25 mM; reaction volume: 44 µL; T: 30°C; pH: 8.0 (5 mM 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid (EPPS) buffer, plus 0.45 Phenol Red®). [c] K_m-[protein]: 4.5 µg mL^{-1}; [substrate]: 0–100 mM; [PLP]: 1 mM; [2-(4-nitrophenyl)ethan-1-amine (NPEA)]: 0–100 mM; reaction volume: 200 µl; T: 40°C; pH: 7.5 (100 mM K_2HPO_4 buffer). In all cases, k_{cat} was calculated assuming $\nu = k_{cat}$ when $\nu = $ µM product min^{-1} µM^{-1} enzyme. Raw data for the kinetic experiments are shown in Figure S11.
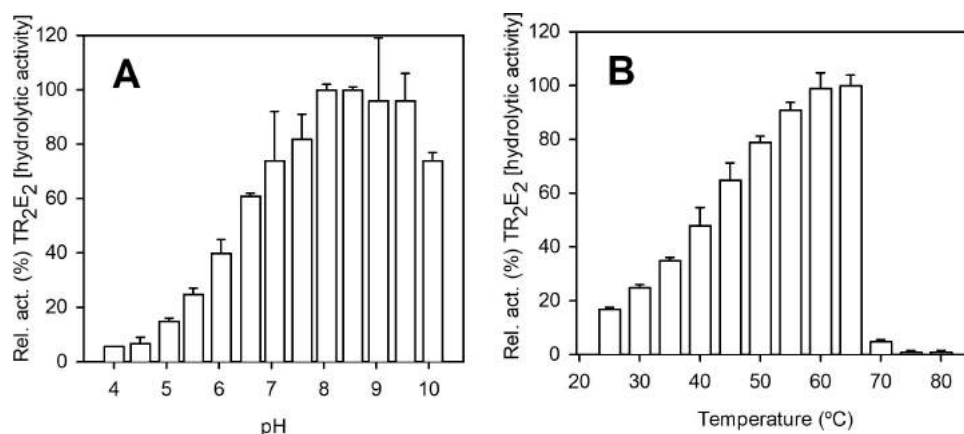

Figure 3. Optimal parameters for assessing the hydrolytic activity of purified TR_2E_2. A) TR_2E_2 activity toward 3-OTfBE at different pH values. Reaction conditions: [protein]: 31 µM; [3-OTfBE]: 7.2 mM; pH: 50 mM Britton–Robinson (BR) buffer with a pH ranging from 3 to 11; reaction volume: 100 µL; and T: 30°C. B) TR_2E_2 activity toward 3-OTfBE at different temperatures. Reaction conditions: [protein]: 31 µM; [3-OTfBE]: 7.2 mM; T: 25–80°C; pH: 100 mM K_2HPO_4 buffer, pH 7.5; and reaction volume: 100 µL. In both cases, the reactions were performed in triplicate and were incubated for 15 min at 750 rpm, after which 300 µL of a solution composed of acetonitrile, H_2O and formic acid (10:10:0.6) were added to stop the reaction. Substrate depletion was detected using HPLC. The figure was created using SigmaPlot 14.0 software.

resolution of 3.3 Å. All statistics for the data are presented in Table S1. The artificial esterase catalytic triad was observed in the TR_2E_2 crystal structures and is composed of Ser172 and His173 of chain A (Figure 1C, magenta) and Glu317 of chain B (Figure 1C, green); similar to the transaminase's active site, the esterase site is composed of a functional protein dimer. Figure 1B presents the surface of the TR_2 protein (monomer A in light green and monomer B in wheat) compared to TR_2E_2 (monomer A in light purple and monomer B in blue) (Figure 1D). The position of the mutation used to gain the esterase function is shown in magenta and features good accessibility, as it is located at the surface, whereas the transaminase catalytic base is located within a small notch, as shown by the bound PLP

(highlighted in ball and sticks in Figure 1D). The RMSD between TR_2 and TR_2E_2 was calculated to be 0.71 Å between 299 of 389 C_α atoms of the aligned residues, indicating that no large conformational change occurs due to mutations, as also confirmed by the transaminase activity assay (Tables 1 and S4). Notably, the distance between both active sites is ≈20 Å, with the esterase site being more solvent-exposed and located at the beginning of the tunnel and the transaminase site being more buried in the protein structure.

Additionally, we succeeded in growing crystals and solving the structure of TR_2E_2 bound to the PLP cofactor (Figure 5A, resolution of 3.5 Å) and in complex with the ethanolamine O-sulfate (EOS) inhibitor (Figure 5B, resolu-
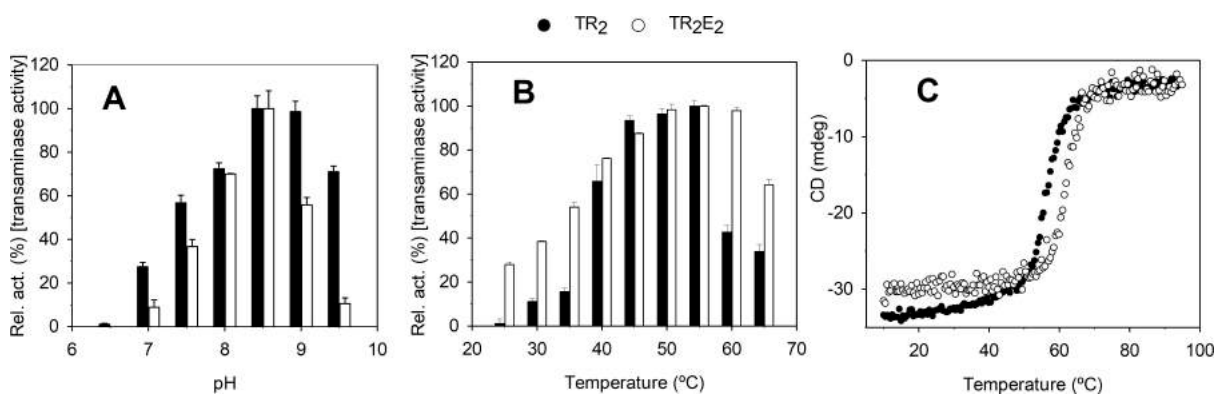
● TR₂   ○ TR₂E₂

Figure 4. Optimal parameters for the transaminase activity and stability of purified TR₂ (black) and TR₂E₂ (white). A) pH profiles of TR₂ and TR₂E₂. Reaction conditions: [protein]: 31 μM; [hexanal]: 14 mM; [PLP]: 1 mM; [NPEA]: 14 mM; *T*: 40 °C; pH: 50 mM BR buffer with a pH ranging from 3 to 11; reaction volume: 200 μL. B) Thermal profiles of TR₂ and TR₂E₂. Reaction conditions: [protein]: 31 μM; [hexanal]: 14 mM; [PLP]: 1 mM; [NPEA]: 14 mM; *T*: 22–80 °C; pH: 100 mM K₂HPO₄ buffer, pH 7.5; and reaction volume: 200 μL. C) The thermal denaturation curves of TR₂ and TR₂E₂ at pH 7.0, as measured by changes in ellipticity at 220 nm. The figure was created using SigmaPlot 14.0 software.
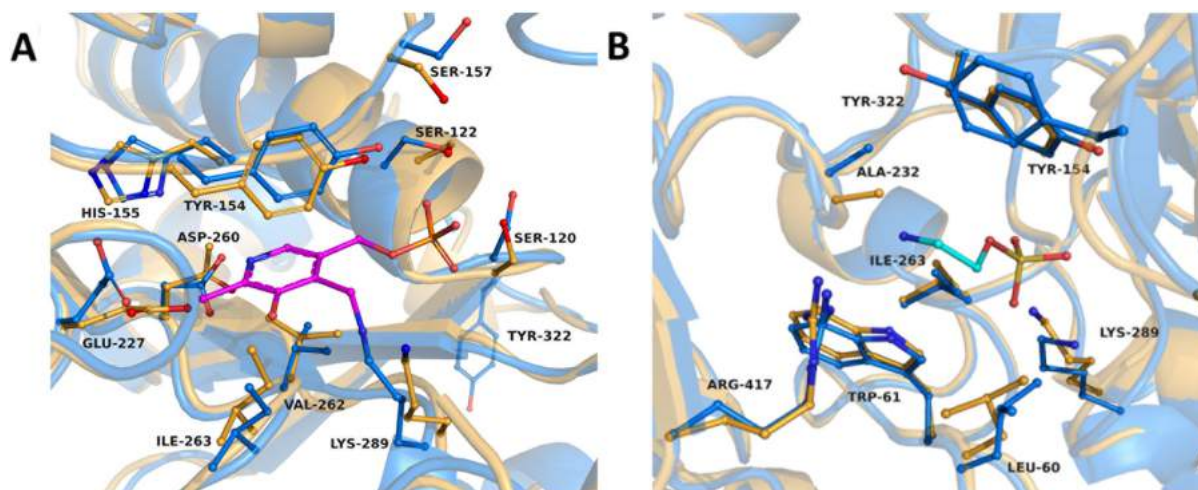

Figure 5. Comparison of the transaminase catalytic site in the TR₂ and TR₂E₂ structures. A) Representation of the catalytic center of TR₂E₂ with PLP (magenta) bound to Lys289. The TR₂ (orange) structure is overlaid with the TR₂E₂-PLP bound structure (blue). B) Overlay of the EOS inhibitor binding site. The TR₂ structure (orange) is compared with the TR₂E₂ structure (blue) within the EOS (cyan) binding site. The introduction of the mutation does not alter either the PLP or the EOS binding site. The figure was created using PyMOL Version 2.3.2 software.

tion of 3.6 Å), located in the transaminase catalytic center. All data statistics are presented in Table S1. The functional TR₂E₂ dimer shows one monomeric subunit (chain A) with PLP bound to the ε-amino group of the catalytic base Lys289, forming an internal aldimine (Figure 5A). The other monomeric subunit (chain B) shows no linkage to the cofactor but an orientation in the catalytic center with stabilization of chain A, indicating that ω-transaminase forms a functional dimer.

After the computational design of a hydrolase site in TR₂ and characterization of the successful variant TR₂E₂, we wanted to test the ability of this validated *PluriZyme* to catalyze a cascade reaction of interest. Therefore, as a model reaction, we chose to synthesize 3-ATfBA from 3-OTfBE; this molecule is an intermediate of gliptins.[22,23] Simulation

of the reaction by a local PELE exploration showed efficient 3-ATfBE and 3-OTfBE catalytic (hydrolytic) binding poses: 23 catalytic events for 3-ATfBE and 42 for 3-OTfBE when the serine-substrate distance threshold was set to 4 Å (Figure S12). Glyceryl tripropionate, an example of an ester commonly hydrolyzed by most esterases,[36] had 121 catalytic events with an associated experimental k_{cat} of 388.1 min⁻¹ (Table S5). Thus, we expected that 3-OTfBE would be preferentially hydrolyzed. This prediction was confirmed experimentally by showing that the catalytic efficiency of TR₂E₂ for the hydrolysis of 3-OTfBE was 480-fold higher than that of 3-ATfBE (Table 1).

We set up a one-pot reaction at 40 °C and pH 7.5 with all the reagents necessary for the hydrolysis and transamination reactions, i.e., 3-OTfBE, PLP (as a cofactor) and NPEA (as

amine donor). After adding the TR_2E_2 *PluriZyme*, the levels of the substrate 3-OTfBE, the intermediates 3-ATfBE and 3-OTfBA, and the product 3-ATfBA were quantified using HPLC over 21 h (Figure 6 and Figure S13), and the e.e. was analyzed using Gas Chromatography (GC). The same experimental conditions were employed with TR_2 and a control reaction with no enzymes.

We found that the β-keto ester substrate 3-OTfBE (concentration, 14 mM) was fully converted ($>99\%$) after 60 min, and that the reaction proceeded via the formation of the β-keto acid 3-OTfBA (initial rate: $0.702 \pm 0.07\,mM\,min^{-1}$), which reached the maximal concentration (9.2 ± 0.8 mM) at 30 min. Afterward, the β-amino acid (R)-3-ATfBA (initial rate: $0.479 \pm 0.08\,mM\,min^{-1}$) was finally produced ($14.3 \pm 0.21$ mM; e.e. $>99\%$). The intermediate β-amino ester 3-ATfBE was a minor product of the reaction during the early stages of the reaction (max. conc. 1.05 ± 0.03 mM at 15 min; initial rate: $0.290 \pm 0.04\,mM\,min^{-1}$), but was subsequently consumed. This result is consistent with the higher transamination efficiency (ca. 200-fold) of TR_2E_2 for 3-OTfBA compared to 3-OTfBE and the 480-fold lower capacity of TR_2E_2 to hydrolyze 3-ATfBE compared to the initial substrate 3-OTfBE (Table 1). Notably, when the reaction was established in the presence of TR_2, no β-amino acid 3-ATfBA was formed, and only a minor conversion to 3-ATfBE β-amino ester was detected (Figure S14).

In summary, of the two possible routes by which the β-keto ester substrate may yield the β-amino acid in the presence of TR_2E_2, we identified the preferential route as the conversion of the β-keto ester to the β-keto acid at the hydrolytic site, followed by its subsequent amination at the transaminase site (Scheme 2A).

In order to demonstrate the catalytic potential of the dual-function TR_2E_2 enzyme, a range of additional substrates were further tested. A series of β-keto esters were tested as potential substrates, because they are commercially available, and have various potential applications as biomedical compounds or are important synthons for natural products synthesis. Both hydrolase and transaminase activities were evaluated and quantified.

In detail, a set of 7 β-keto esters were tested, which included the aromatic substrates benzyl 3-oxobutanoate, indole-3-pyruvic acid methyl ester, methyl 2-oxo-3-phenyl-propanoate, 2-methyl-4-oxo-4H-pyran-3-yl propionate, ethyl (4-ethoxyphenyl)(oxo)acetate, and ethyl (1-hydroxy-4-oxocyclohexyl)acetate, and the linear alkyl β-keto ester dimethyl 2-oxoglutarate (Table S6; Figure S11). Among them, 6 were hydrolyzed by the biotic artificial esterase site in TR_2E_2, at different rates. In addition, 6 were also accepted as amine acceptors in transaminase reaction by the biotic native transaminase site in TR_2E_2; they were converted at a lower rate compared to hexanal.

We first set up a one-pot cascade reaction with benzyl 3-oxobutanoate, using similar conditions as those used for 3-OTfBE (see Figure 6), as it was one of the preferred β-keto esters for both hydrolase and transaminase reactions from those seven tested, and the possible end β-amino acid product of the reaction is an important intermediate for the production of biomedicals.[36] The products were quantified using HPLC. We found that the β-keto ester benzyl 3-oxobutanoate (concentration, 14 mM) was converted to a high extent ($>95\%$ after 20 min and 99% after 60 min; Figure S15). The reaction proceeded via the rapid formation of benzyl alcohol (initial rate: $5.85 \pm 0.04\,mM\,min^{-1}$) and 3-
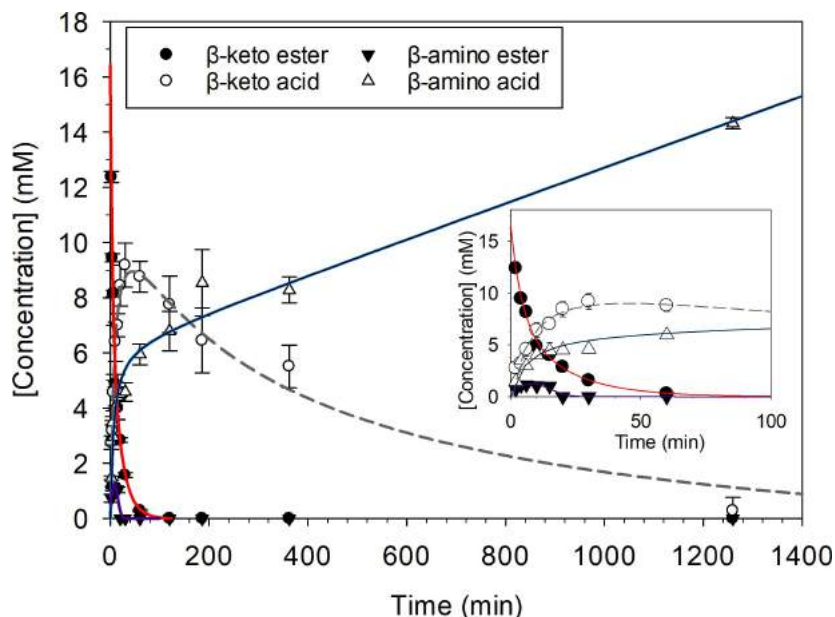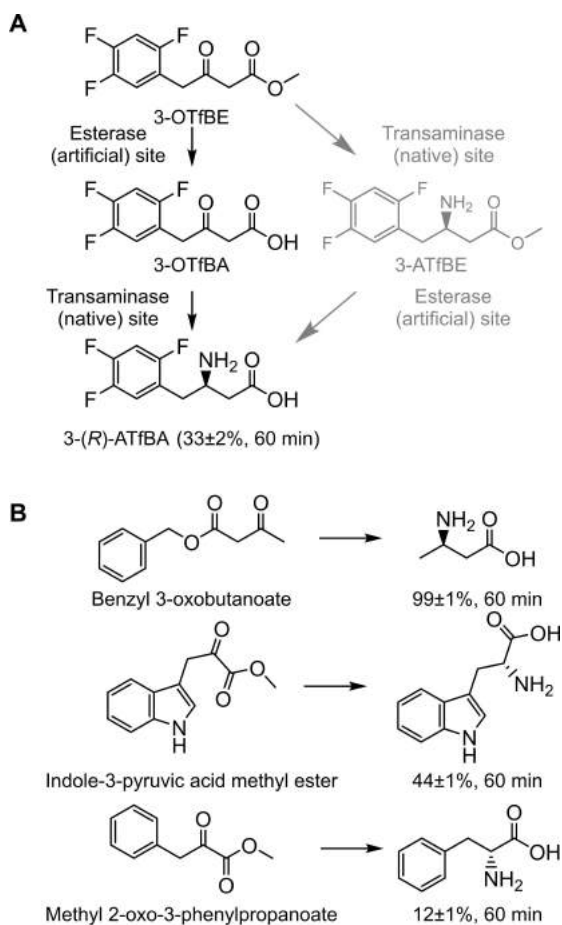

Figure 6. Concentrations of substrates, intermediates and the final product obtained for the conversion of 3-OTfBE by the *PluriZyme* TR_2E_2. The following reaction conditions were used: [protein]: 31 μM; [3-OTfBE]: 14 mM; [PLP]: 1 m; [NPEA]: 14 mM; pH: 100 mM K_2HPO_4 buffer, pH 7.5; reaction volume: 100 μL; and *T*: 40°C. The figure was created using SigmaPlot 14.0 software.

Scheme 2. Schematic representation the one-pot reaction products for converting a set of β-keto esters by TR$_2$E$_2$. A) Schematic representation of the two possible routes for converting 3-OTfBE into 3-ATfBA. As shown in the figure, the preferential route is the production of 3-ATfBA via hydrolysis of 3-OTfBE to 3-OTfBA, which then yields (R)-3-ATfBA, and not via formation of 3-ATfBE (gray color). B) Schematic representation of one-pot synthesis of amino acids from a set of β-keto esters. The conversion, determined by HPLC for all substrates, at 60 min is shown. The figure was created using ChemDraw 18.2.

oxo-butyric acid, finally yielding the expected amine-product (R-3-aminobutyric acid, e.e. >99%) (Scheme 2B), as confirmed by high-resolution mass spectrometry (HR-MS; Figure S16) analysis which determined the m/z ratio 104.0701 for the composition $C_4H_{10}NO_2$ ([M+H]) with an error of less than 5 ppm with respect to the theoretically calculated mass for that composition.

Under similar conditions and by HPLC, we further succeed in the one-pot conversion at 60 min of the aromatic β-keto ester indole-3-pyruvic acid methyl ester to tryptophan (44±1%, 60 min; Figure S17) and of methyl 2-oxo-3-phenyl-propanoate to phenylalanine (12±1%, 60 min; Figure S18) (Scheme 2B).

Structural analysis revealed a 20 Å distance between both active sites, with the esterase site being solvent-exposed and the transaminase site buried in the protein structure

(Figure 1D), thus ensuring that the transfer of the reaction intermediate within the protein structure is theoretically possible. Actually, we tested the migration of 3-OtfBA using PELE from the esterase site to the transaminase one for TR$_2$E$_2$ and for two mutants, TR$_2$E$_2$A232F/L60F and TR$_2$E$_2$A232F/F89W, designed to close the access channel to the transaminase site. As it can be seen in the energetic profiles of those simulations, the substrate can easily reach the transaminase site from the esterase site in TR$_2$E$_2$ but not in the two mutants (Figure S19). The feasibility of such a connection between sites was further proven by experimentally characterizing the TR$_2$E$_2$A232F/L60F. Indeed, this mutant retained the affinity and hydrolytic activity for 3-OTfBA (K_m of 2.063±0.046 mM and k_{cat} of 28.89± 1.52 min^{-1}), while showing lower affinity and transamination activity for 3-OTfBA (K_m of 9.98±0.32 mM and k_{cat} of 44.87±4.14 min^{-1}) (Figure S20), compared to TR$_2$E$_2$ (see Table 1). This indicates that mutations affect the substrate access to the transaminase site but not the esterase site. This was further validated by following the one-pot conversion from 3-OTfBE to 3-(R)-ATfBA by HPLC (Figure S21), which was 3-fold times lower in the mutant at a short incubation time (6 minutes); however, at 60 min an almost full conversion was achieved with TR$_2$E$_2$ and TR$_2$E$_2$A232F/L60F, most likely due to the possibility that the intermediate diffuse from one side to the other at longer incubation times.

Conclusion

We have documented the successful design and construction of a *PluriZyme* integrating biotic transaminase and esterase activities and its application in a cascade reaction. We achieved these results by applying computational techniques to a class III ω-TA and by examining a set of 20 mutants supporting presumptive catalytic triads and oxyanion holes for ester hydrolysis. Of the six variants retaining transaminase activity, one acquired additional esterase activity and was capable of hydrolyzing multiple esters at turnover numbers approaching those of highly efficient similar native esterases/lipases reported in the literature.[37] Interestingly, TR$_2$E$_2$ was designed using a site where a loop was missing in the crystal structure of the original transaminase TR$_2$; the loop contained one mutation and was modeled with Prime and refined through MD simulations (Figure S22). Thus, the insertion of the hydrolytic site helped stabilize this region by adding two internal hydrogen bonds, improving the thermostability of the enzyme. The incorporation of a second biological activity to support a cascade reaction into an enzyme may not occur at the expense of the activity and stability of the original enzyme.

We further revealed that this bioinspired design facilitates the one-step synthesis of a sitagliptin intermediate, specifically the conversion of 3-OTfBE into 3-ATfBA in a single one-pot reaction (Scheme 2). This β-amino acid is commonly synthesized through a bienzymatic cascade reaction that involves an esterase and a transaminase or a recombinant *Escherichia coli* strain coexpressing both

enzymes.[22,23,38] A series of chiral 3-substituted cyclo-hexyl-amine derivatives, capsaicin analogs and nylon-6 monomer 6-aminohexanoic acid, to cite some significant examples, have also been synthesized by combining hydrolases with ωTAs as separate catalytic entities.[39–41] It is worth mentioning that to date, only three native hydrolases, namely, those from *Candida rugosa*, *Pseudomonas avellanae* and *P. stutzeri*,[22,23,37] have been reported as capable of hydrolyzing 3-OTfBE and used in the cascade synthesis of 3-ATfBA. Thus, the computational approach used here expands the range of biocatalysts supporting this transformation. The TR_2E_2 *PluriZyme* achieved (R)-3-ATfBA production rates (ca. 0.479 ± 0.08 mM min^{-1}) in the range of the previously reported bienzymatic cascade systems (ca. 0.05–0.4 mM min^{-1}).[22,38] Having said that, further studies are needed to determine whether using a single artificial enzyme with two "bioactive" sites in cascade reactions, is more efficient than a conventional multienzyme system and natural or artificial chimeric enzymes with hybrid activities.[11–13] Among the advantages, it is worth mentioning that using one enzyme with two biotic sites catalysing different chemistry would reduce the development and costs associated with the production of several enzymes, each supporting one chemistry, or favoring reaction yields by helping the substrates channeling; however, these and other advantages, yet to be defined, may vary from enzyme to enzyme.

In this study, we used a transaminase as a scaffold to introduce esterase activity. We consider this workflow more appropriate than the other approach, namely, exploring esterases or lipases to find alternative binding sites to accommodate a new site supporting transaminase activity. This design would require cofactor binding sites or large channels to accommodate the substrates, which may be technically more challenging.

The bioinspired design of the *PluriZyme* described here clearly provides a computational and experimental framework to develop future enzyme designs integrating two biotic-catalytic centers, which, being independent, can also work in synergy for cascade reactions. In this direction, the fact that TR_2E_2 could convert multiple keto esters exemplified the catalytic potential of such designs. An important aspect of such future work will be the configuration of *PluriZymes* that, by supporting multiple biological activities through the assembly of appropriated binding and catalytic pockets, might reassemble in vitro "artificial metabolisms" as pathways for the preparation of valuable biomolecules, including medicines.

We would like to highlight that biosynthetic enzymes with different domains, in which a substrate is channelled between two biotic sites catalysing different chemistries, already exist in Nature.[13] Therefore, the dual active site strategy herein designed, namely the artificial TR_2E_2 enzyme, may represent a complementary one to that already employed by Nature for multi-domain enzymes. In addition to that, the catalytic promiscuous character of the enzyme design herein reported may complement the catalytic potential of engineered promiscuous enzymes that are known to represent attractive alternatives to conventional chemical catalysts. A recent example is the promiscuous acyltransferase activity of certain hydrolases that can catalyze not only the formation of esters, but also the formation of amides, carbonates, and carbamates in water.[42] While such catalytic promiscuous enzymes can potentially perform two different chemistries, their use in cascade reactions have been rarely investigated compared to our design.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the Supporting Information of this article.

[1] C. T. Walsh, B. S. Moore, *Angew. Chem. Int. Ed.* **2019**, *58*, 6846–6879; *Angew. Chem.* **2019**, *131*, 6918–6952.

[2] J. M. Sperl, V. Sieber, *ACS Catal.* **2018**, *8*, 2385–2396.

[3] H. Xia, N. Li, W. Huang, Y. Song, Y. Jiang, *ACS Appl. Mater. Interfaces* **2021**, *13*, 22240–22253.

[4] A. Belluati, I. Craciun, J. Liu, C. G. Palivan, *Biomacromolecules* **2018**, *19*, 4023–4033.

[5] P. B. O'Mara, P. Wilde, T. M. Benedetti, C. Andronescu, S. Cheong, J. J. Gooding, R. D. Tilley, W. Schuhmann, *J. Am. Chem. Soc.* **2019**, *141*, 14093–14097.

[6] E. T. Hwang, S. Lee, *ACS Catal.* **2019**, *9*, 4402–4425.

[7] S. Tsitkov, H. Hess, *ACS Catal.* **2019**, *9*, 2432–2439.

[8] Y. Zhang, H. Hess, *ACS Catal.* **2017**, *7*, 6018–6027.

[9] H. Wang, Z. Zhao, Y. Liu, C. Shao, F. Bian, Y. Zhao, *Sci. Adv.* **2018**, *4*, eaat2816.

[10] K. Zhou, T. Tian, C. Wang, H. Zhao, N. Gao, H. Yin, P. Wang, B. J. Ravoo, G. Li, *J. Am. Chem. Soc.* **2020**, *142*, 20605–20615.

[11] M. Rizk, G. Antranikian, S. Elleuche, *Mol. Biotechnol.* **2016**, *58*, 268–279.

[12] B. C. Knott, E. Erickson, M. D. Allen, J. E. Gado, R. Graham, F. L. Kearns, I. Pardo, E. Topuzlu, J. J. Anderson, H. P. Austin, G. Dominick, C. W. Johnson, N. A. Rorrer, C. J. Szostkiewicz, V. Copié, C. M. Payne, H. L. Woodcock, B. S. Donohoe, G. T. Beckham, J. E. McGeehan, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 25476–25485.

[13] J. L. Faylo, T. van Eeuwen, H. J. Kim, J. J. Gorbea Colón, B. A. Garcia, K. Murakami, D. W. Christianson, *Nat. Commun.* **2021**, *12*, 3487.

[14] F. Christoffel, N. V. Igareta, M. M. Pellizzoni, L. Tiessler-Sala, B. Lozhkin, D. C. Spiess, A. Lledós, J.-D. Maréchal, R. L. Peterson, T. R. Ward, *Nat. Catal.* **2021**, *4*, 643–653.

[15] Z. Liu, C. Calvó-Tusell, A. Z. Zhou, K. Chen, M. Garcia-Borràs, F. H. Arnold, *Nat. Chem.* **2021**, *13*, 1166–1172.

[16] J. M. Palomo, *Curr. Opin. Green Sustainable Chem.* **2021**, *29*, 100452.

[17] A. G. Jarvis, *Nat. Catal.* **2021**, *4*, 639–640.

[18] G. Santiago, M. Martínez-Martínez, S. Alonso, R. Bargiela, C. Coscolín, P. N. Golyshin, V. Guallar, M. Ferrer, *Biochemistry* **2018**, *57*, 2245–2255.

[19] S. Alonso, G. Santiago, I. Cea-Rama, L. Fernandez-Lopez, C. Coscolín, J. Modregger, A. K. Ressmann, M. Martínez-Martínez, H. Marrero, R. Bargiela, M. Pita, J. L. Gonzalez-Alfonso, M. L. Briand, D. Rojo, C. Barbas, F. J. Plou, P. N. Golyshin, P. Shahgaldian, J. Sanz-Aparicio, V. Guallar, M. Ferrer, *Nat. Catal.* **2020**, *3*, 319–328.

[20] C. Coscolín, N. Katzke, A. García-Moyano, J. Navarro-Fernández, D. Almendral, M. Martínez-Martínez, A. Bollinger, R. Bargiela, C. Gertler, T. N. Chernikova, D. Rojo, C. Barbas, H. Tran, O. V. Golyshina, R. Koch, M. M. Yakimov, G. E. K. Bjerga, P. N. Golyshin, K.-E. Jaeger, M. Ferrer, *Appl. Environ. Microbiol.* **2019**, *85*, e02404.

[21] J. H. Schrittwieser, S. Velikogne, M. Hall, W. Kroutil, *Chem. Rev.* **2018**, *118*, 270–234.

[22] G. H. Kim, H. Jeon, T. P. Khobragade, M. D. Patil, S. Sung, S. Yoon, Y. Won, I. S. Choi, H. Yun, *Enzyme Microb. Technol.* **2019**, *120*, 52–60.

[23] T. P. Khobragade, S. Sarak, A. D. Pagar, H. Jeon, P. Giri, H. Yun, *Front. Bioeng. Biotechnol.* **2021**, *9*, 757062.

[24] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Crystallogr. Sect. D* **2012**, *68*, 352–367.

[25] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.

[26] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126–2132.

[27] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallogr. Sect. D* **2019**, *75*, 861–877.

[28] N. V. Grishin, M. A. Phillips, E. J. Goldsmith, *Protein Sci.* **1995**, *4*, 1291–1304.

[29] S. Roda, G. Santiago, V. Guallar, *Adv. Protein Chem. Struct. Biol.* **2020**, *122*, 1–31.

[30] S. Roda, A. Robles-Martín, R. Xiang, M. Kazemi, V. Guallar, *J. Phys. Chem. B* **2021**, *125*, 6491–6500.

[31] S. Acebes, E. Fernandez-Fueyo, E. Monza, M. F. Lucas, D. Almendral, F. J. Ruiz-Dueñas, H. Lund, A. T. Martinez, V. Guallar, *ACS Catal.* **2016**, *6*, 1624–1629.

[32] A. Höppner, A. Bollinger, S. Kobus, S. Thies, C. Coscolín, M. Ferrer, K. E. Jaeger, S. H. J. Smits, *FEBS J.* **2021**, *288*, 3570–3584.

[33] H. Cui, H. Cao, H. Cai, K. E. Jaeger, M. D. Davari, U. Schwaneberg, *Chem. Eur. J.* **2020**, *26*, 643–649.

[34] H. Cui, L. Eltoukhy, L. Zhang, U. Markel, K. E. Jaeger, M. D. Davari, U. Schwaneberg, *Angew. Chem. Int. Ed.* **2021**, *60*, 11448–11456; *Angew. Chem.* **2021**, *133*, 11549–11557.

[35] H. Cui, S. Pramanik, K. E. Jaeger, M. D. Davari, U. Schwaneberg, *Green Chem.* **2021**, *23*, 3474–3486.

[36] M. V. C. B. Cortes, R. R. Menezes, E. G. Oestreicher, *Braz. J. Chem. Eng.* **2015**, *32*, 53–57.

[37] M. Martínez-Martínez, C. Coscolín, G. Santiago, J. Chow, P. J. Stogios, R. Bargiela, C. Gertler, J. Navarro-Fernández, A. Bollinger, S. Thies, C. Méndez-García, A. Popovic, G. Brown, T. N. Chernikova, A. García-Moyano, G. E. K. Bjerga, P. Pérez-García, T. Hai, M. V. Del Pozo, R. Stokke, I. H. Steen, H. Cui, X. Xu, B. P. Nocek, M. Alcaide, M. Distaso, V. Mesa, A. I. Peláez, J. Sánchez, P. C. F. Buchholz, J. Pleiss, A. Fernández-Guerra, J. O. Glöckner, O. V. Golyshina, M. M. Yakimov, A. Savchenko, K.-E. Jaeger, A. F. Yakunin, W. F. Streit, P. N. Golyshin, V. Guallar, M. Ferrer, The INMARE Consortium, *ACS Chem. Biol.* **2018**, *13*, 225–234.

[38] T. P. Khobragade, S. Yu, H. Jung, M. D. Patil, S. Sarak, A. D. Pagar, H. Jeon, S. Lee, P. Giri, G. H. Kim, S. S. Cho, S. H. Park, H. J. Park, H. M. Kang, S. R. Lee, M. S. Lee, J. H. Kim, I. S. Choi, H. Yun, *Biotechnol. Bioeng.* **2021**, *118*, 3263–3268.

[39] R. C. Simon, N. Richter, E. Busto, W. Kroutil, *ACS Catal.* **2014**, *4*, 129–143.

[40] J. D. Carballeira, M. A. Quezada, P. Hoyos, Y. Simeo, M. J. Hernaiz, A. R. Alcantara, J. V. Sinisterra, *Biotechnol. Adv.* **2009**, *27*, 686–714.

[41] F. Guo, P. Berglund, *Green Chem.* **2017**, *19*, 333–360.

[42] H. Müller, S. P. Godehard, G. J. Palm, L. Berndt, C. P. S. Badenhorst, A. K. Becker, M. Lammers, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2020**, *59*, 2013–2017; *Angew. Chem.* **2020**, *132*, 2029–2033.

Research Articles

Protein Engineering

S. Roda, L. Fernandez-Lopez, M. Benedens,
A. Bollinger, S. Thies, J. Schumacher,
C. Coscolín, M. Kazemi, G. Santiago,
C. G. W. Gertzen, J. L. Gonzalez-Alfonso,
F. J. Plou, K.-E. Jaeger, S. H. J. Smits,
M. Ferrer,* V. Guallar* ———— **e202207344**

A Plurizyme with Transaminase and Hydrolase Activity Catalyzes Cascade Reactions
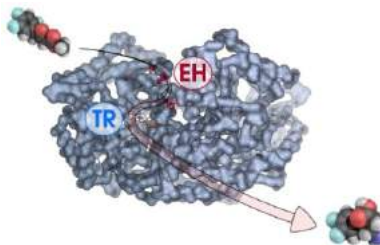
A *PluriZyme* with transaminase and esterase activities was constructed by combining computational and laboratory design methods. The enzyme can perform cascade reactions in a single protein scaffold. Experimental evidence is presented for the catalytic efficiency of both the native transaminase and the artificial esterase active site and their synergistic action to transform β-keto esters into enantiopure β-amino acids.

WILEY-VCH

A PluriZyme with Transaminase and Hydrolase Activity Catalyzes Cascade Reactions

Sergi Roda,[§][a] Laura Fernandez-Lopez,[§][b] Marius Benedens,[c] Alexander Bollinger,[d] Stephan Thies,[d] Julia Schumacher,[c] Cristina Coscolín,[b] Masoud Kazemi,[a] Gerard Santiago,[a] Christoph G. W. Gertzen, [c] Jose L. Gonzalez-Alfonso,[b] Francisco J. Plou,[b] Karl-Erich Jaeger,[d] Sander H. J. Smits,[c] Manuel Ferrer,*[b] and Víctor Guallar,*[a,e]

Abstract: *Engineering dual-function single polypeptide catalysts with two abiotic or biotic catalytic entities (or combinations of both) supporting cascade reactions is becoming an important area of enzyme engineering and catalysis. Here we present the development of a PluriZyme, TR_2E_2, with efficient native transaminase (k_{cat}: 69.49±1.77 min^{-1}) and artificial esterase (k_{cat}: 3908 to 0.41 min^{-1}) activities, integrated into a single scaffold and evaluate its utility in a cascade reaction. TR_2E_2 (pH_{opt} 8.0-9.5; T_{opt}: 60-65 °C) efficiently converts methyl 3-oxo-4-(2,4,5-trifluorophenyl)butanoate into 3-(R)-amino-4-(2,4,5-trifluorophenyl)butanoic acid, a crucial intermediate for the synthesis of antidiabetic drugs. The reaction proceeds through the conversion of the β-keto ester to the β-keto acid at the hydrolytic site and subsequently to the β-amino acid (e.e. >99%) at the transaminase site. The catalytic power of the TR_2E_2 PluriZyme was proved with a set of β-keto esters, demonstrating the potential of such designs to address bioinspired cascade reactions.*

Table of Contents

Experimental Procedures

Chemicals. All esters used here were of the highest purity and were obtained from the manufacturers described in a previous study.[1] Hexanal (ref. 115606), PLP (ref. P9255), and NPEA (ref. 184802-5G) were ordered from Merck Life Science S.L.U. (Madrid, Spain). 3-OTfBE, 3-OTfBA, 3-ATfBA, and 3-ATfBE were purchased from FluoroChem (Hadfield, UK). 2-Methyl-4-oxo-4h-pyran-3-yl propionate (ref. 28245775), ethyl (4-ethoxyphenyl)(oxo)acetate (100873341), ethyl (1-hydroxy-4-oxocyclohexyl)acetate (311298422), and benzyl 3-oxobutanoate (28792564) were from AldrichCPR, Merck Life Science S.L.U. Madrid, Spain. Indole-3-pyruvic acid methyl ester was ordered from Biosynth Carbosynth, Compton, United Kingdom (ID FI30480), methyl 2-oxo-3-phenylpropanoate from Biosynth Carbosynth, Compton, United Kingdom (ID GAA36258), and dimethyl 2-oxoglutarate from Merck Life Science S.L.U., Madrid, Spain (ID 349631). The intermediates indole-3-pyruvic acid (I7017), 2-oxo-3-phenylpropionic acid (ID 286958), and 2-oxoglutaric acid (ID 1051940025) were ordered from Merck Life Science S.L.U., Madrid, Spain.

Enzyme source, production, and purification. The sequences encoding wild-type TR_2 (GenBank acc. nr. MH588437),[1] and its mutants (TR_2E_2, and $TR_2E_2A232F/L60F$) were used as templates for gene synthesis (GenScript, US), and genes were codon-optimized to maximize expression in *E. coli*. Genes were flanked by BamHI and HindIII (stop codon) restriction sites and inserted in a pET-45b(+) expression vector with an ampicillin selection marker (GenScript, US). This plasmid, which was introduced into *E. coli* BL21(DE3), supports the expression of N-terminal histidine (His) fusion proteins. The soluble His-tagged proteins were produced and purified at 4 °C after binding to a Ni-NTA His-Bind resin (Merck Life Science S.L.U., Madrid, Spain), as previously described.[1] The purity was assessed to be >98% by performing an SDS–PAGE analysis using a Mini PROTEAN electrophoresis system (Bio-Rad, Madrid, Spain). Purified proteins were stored at -86 °C until use at a concentration of 10 mg ml⁻¹ in 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 7.0). Approximately 32-45 mg of purified recombinant proteins (TR_2, TR_2E_2, and $TR_2E_2A232F/L60F$) were obtained on average from a 1-liter cultures.

Hydrolytic activity assessments. The activity toward the model esters *p*NPA, *p*NPP and *p*NPB was assessed as described previously[2] by monitoring the production of 4-nitrophenol at 348 nm (pH-independent isosbestic point, ε = 4147 M⁻¹ cm⁻¹) in 96-well plates (ref. 655801, Greiner Bio-One GmbH, Kremsmünster, Austria). Conditions for determining K_m were as follows - [protein]: 4.5 µg ml⁻¹; [*p*NP ester]: 0-7 mM; reaction volume: 200 µl; T: 30 °C; pH: 7.0 (40 mM HEPES buffer). The activity was calculated by determining the absorbance per minute from the generated slopes and applying the following formula (Equation 1):

$$Rate\left(\frac{\mu mol}{min\, mg\, protein}\right) = \frac{\dfrac{\Delta\frac{|Abs|}{min}}{4147M-1cm-1}*1}{1M} * 10^{6}\mu M}{1M} * 0.0002\frac{L*1}{mg\, protein}$$

Equation 1. Formula to calculate hydrolase rate with *p*NP esters.

Hydrolysis of a set of 96 additional esters, plus 3-OTfBE, 3-ATfBE, 2-methyl-4-oxo-4h-pyran-3-yl propionate, ethyl (4-ethoxyphenyl)(oxo)acetate, ethyl (1-hydroxy-4-oxocyclohexyl)acetate, and benzyl 3-oxobutanoate, was also assayed using a pH indicator (Phenol Red®) assay in 384-well plates (ref. 781162, Greiner Bio-One GmbH, Kremsmünster, Austria), at 550 nm.[1] If otherwise not indicated, esters and conditions were described in detail in a previous study.[1] Conditions for determining K_m were as follows - [protein]: 4.5 µg ml⁻¹; [ester]: 0-25 mM (range depending on the substrate); reaction volume: 44 µl; T: 30 °C; pH: 8.0 (5 mM EPPS buffer, plus 0.45 Phenol Red®). The activity was calculated by determining the absorbance per minute from the generated slopes and applying the following formula (Equation 2):

$$Rate\left(\frac{\mu mol}{min\,mg\,protein}\right) = \frac{\frac{\frac{\Delta\frac{|Abs|}{min}}{8450M-1cm-1}*1}{0.4cm}*10^6\mu M}{1M} * 0.000044 \frac{L*1}{mg\,prot.}$$

Equation 2. Formula to calculate hydrolyse rate with a pH indicator (Phenol Red®) assay.

In all cases, k_{cat} was calculated assuming v = k_{cat} when v = µM product min⁻¹ µM⁻¹ enzyme. In case of esters others that pNP esters, 3-OTfBE, 3-ATfBE, 2-methyl-4-oxo-4h-pyran-3-yl propionate, ethyl (4-ethoxyphenyl)(oxo)acetate, ethyl (1-hydroxy-4-oxocyclohexyl)acetate, benzyl 3-oxobutanoate, and methyl 2-oxo-3-phenylpropanoate, k_{cat} were calculated under substrate saturation conditions (50 mM substrates[1]). In all cases, all values, which were determined in triplicate, were corrected for nonenzymatic transformation, and the absence of activity was defined as an at least two-fold increase compared with the background signal, as described previously.[1,2] A Synergy HT Multi-Mode Microplate Reader (Agilent, Madrid, Spain) was used for continuous monitoring of the signal.

Transaminase activity assay. Transaminase activity was assayed using NPEA as an amine donor and hexanal, 3-OTfBA, 3-OTfBE, 2-methyl-4-oxo-4h-pyran-3-yl propionate, ethyl (4-ethoxyphenyl)(oxo)acetate, ethyl (1-hydroxy-4-oxocyclohexyl)acetate, benzyl 3-oxobutanoate, indole-3-pyruvic acid methyl ester, methyl 2-oxo-3-phenylpropanoate, and dimethyl 2-oxoglutarate as acceptors, as previously described with some modifications.[3] Briefly, assay reactions were conducted in a Synergy HT Multi-Mode Microplate Reader as described below. Before the assay, a solution of 14 mM NPEA, the amine donor, and 1.0 mM PLP, the cofactor, was prepared in 100 mM K₂HPO₄ buffer, pH 7.5 (40 ml). A 224 mM acceptor stock solution was prepared in acetonitrile. If otherwise not indicated, reaction assays in 96-well microtiter plates were started by adding 10 µl of a protein solution (stock solution, 10.0 mg ml⁻¹ in 100 mM K₂HPO₄ buffer, pH 7.5) to an assay mixture containing 185 µl of the PLP/NPEA solution and 12.5 µl of the acceptor stock solution. The final volume of the assay mixture was 200 µl, and the amine donor and acceptor concentrations were 14 mM each. Unless indicated otherwise, all measurements were performed at 40 °C, and the absorbance was continuously recorded at 600 nm for a total time of 180 min. One unit (U) of enzyme activity was defined as the amount of protein required to transform 1 µmol of substrate in 1 min under the assay conditions using a reaction product extinction coefficient of 537 M⁻¹ cm⁻¹.[3] Conditions for the determination of kinetic constants for 3-OTfBE and 3-ATfBE were as follows: 1) for K_m - [protein]: 4.5 µg ml⁻¹; [substrate]: 0-100 mM (range depending on the substrate); [PLP]: 1 mM; [NPEA]: 0-100 mM; reaction volume: 200 µl; T: 40 °C; pH: 7.5. 2) k_{cat} was calculated assuming v = k_{cat} when v = µM product min⁻¹ µM⁻¹ enzyme. All values, which were recorded in triplicate, were corrected for nonenzymatic transformation. The activity was calculated by determining the absorbance per minute from the generated slopes and applying the following formula (Equation 3):

$$Rate\left(\frac{\mu mol}{min\,mg\,protein}\right) = \frac{\frac{\frac{\Delta\frac{|Abs|}{min}}{567M-1cm-1}*1}{0.4cm}*10^6\mu M}{1M} * 0.0002 \frac{L*1}{mg\,prot.}$$

Equation 3. Formula to calculate transaminase rate.

pH and temperature profiles. The reaction conditions used to determine the optimal pH for the hydrolysis of 3-OTfBE were as follows: [protein]: 31 µM; [3-OTfBE]: 7.2 mM; pH: 50 mM Britton–Robinson (BR) buffer with a pH ranging from 3 to 11; reaction volume: 100 µl; and T: 30 °C. The reaction conditions used to determine the optimal temperature for 3-OTfBE hydrolysis were as follows: [protein]: 31 µM; [3-OTfBE]: 7.2 mM; T: 25-80 °C; pH: 50 mM BR buffer pH 8.0; reaction volume: 100 µl In both cases, triplicate reactions were incubated for 15 min at 750 rpm, after which 300 µl of a solution composed of acetonitrile, H₂O and formic acid (10:10:0.6) were added to stop the reaction. Substrate depletion in triplicate reactions was detected using HPLC (see a description of the analytical conditions below).

The reaction conditions used to determine the optimal pH for the transamination of hexanal were as follows: [protein]: 31 µM; [hexanal]: 14 mM; [PLP]: 1 mM; [NPEA]: 14 mM; T: 40 °C; pH: 50 mM BR buffer with a pH ranging from 3 to 11; and reaction volume: 200 µl. The reaction conditions used to determine the optimal temperature for the transamination of hexanal were as follows: [protein]: 31 µM; [hexanal]: 14 mM; [PLP]: 1 mM; [NPEA]: 14 mM; T: 20-80 °C; pH: 100 mM K₂HPO₄ buffer, pH 7.5; and reaction volume: 200 µl. In all cases, the absorbance was continuously recorded at 600 nm for a total time of 180 min, and all values were corrected for nonenzymatic transformation.

Circular dichroism spectroscopy to determine the thermal denaturation profile. 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 the denaturation temperature (T_d) values were determined at 220 nm between 10 and 85 °C at a rate of 30 °C per hour in 50 mM BR buffer at pH 7.0. A protein concentration of 1.0 mg ml⁻¹ was used. The T_d (and standard deviation of the linear fit) was calculated by fitting the ellipticity (mdeg) at 220 nm at each of the different temperatures using a 5-parameter sigmoid fit with SigmaPlot 14.0 software.

One-pot Conversion of 3-OTfBE to 3-ATfBA, and Other β-Keto Esters. Prior to the assay, a solution containing 14 mM NPEA, the amine donor, and 1.0 mM PLP was first prepared in 100 mM K_2HPO_4 buffer, and a 224 mM 3-OTfBE stock solution was prepared in acetonitrile. Reactions were prepared in 5-ml penicillin vials that were closed tightly and started by adding the PLP/NPEA solution (to a final volume of 1 ml), 62.5 µl of the 3-OTfBE stock solution and 125 µl of a protein solution (stock solution, 12.5 mg ml^{-1} in 100 mM K_2HPO_4 buffer, pH 7.5). The final volume of the assay mixture was 1000 µl, and the amine donor and 3-OTfBE concentrations were 14 mM each. Aliquots (100 µl) were removed at different time points, and reactions were stopped by adding 300 µl of a mixture containing acetonitrile, H_2O and formic acid (ratio 10:10:0.6). Both the substrate and products were monitored using HPLC (see below). Enantiomeric excess (e.e.) in the reaction mixture was determined using GC as reported previously.[3] Similar conditions were used for the reactions with benzyl 3-oxobutanoate, indole-3-pyruvic acid methyl ester, and methyl 2-oxo-3-phenylpropanoate.

HPLC analysis. The substrates and reaction products were quantified using HPLC with a quaternary pump (Model 600, Waters) coupled to an autosampler (Varian ProStar, Model 420) and a photodiode array detector (ProStar, Varian). The injection volume was 20 µL (for reactions with 3-OTfBE, benzyl 3-oxobutanoate, and methyl 2-oxo-3-phenylpropanoate), or 1 µL (for reactions with indole-3-pyruvic acid, methyl ester). The temperature of the column was kept constant at 40 °C. Quantification was performed at a wavelength of 265 nm (for reactions with 3-OTfBE), 220 nm (for reactions with benzyl 3-oxobutanoate), 280 nm (for reactions with indole-3-pyruvic acid, methyl ester), and 256 nm and 210 nm (for reactions with methyl 2-oxo-3-phenylpropanoate: 256 nm for detecting methyl 2-oxo-3-phenylpropanoate and the intermediate phenylpyruvic acid; and 210 nm for detecting phenylalanine), and integration was carried out using the Varian Star LC workstation 6.41. The column was a Gemini C6-Phenyl 110A (4.6 x 150 mm, 5 µm, Phenomenex) for reactions with 3-OTfBE and methyl 2-oxo-3-phenylpropanoate, and Zorbax Eclipse Plus C-18 (4.6 x 100 mm, 3.5 µm, Agilent) for benzyl 3-oxobutanoate and indole-3-pyruvic acid, methyl ester reactions. For the analysis of the reaction products of 3-OTfBE, the mobile phase was a gradient of acidified acetonitrile containing formic acid at 0.2% (v/v), and water, both degassed with helium. The flow rate was 0.8 ml min^{-1}, and the total analysis time was 25 min. The mobile phase (acetonitrile with 0.2% formic acid) was continuously fed as follows: 0-4.5 min, 12%; from 4.5 to 15 min, 60% → 70%; from 15 to 25 min. 12%. For the analysis of the reaction products of benzyl 3-oxobutanoate the mobile phase was ACN/H_2O degassed with helium and acidified with a 0.1 % (v/v) of formic acid with a flow rate fixed at 1.0 ml min^{-1}. The gradient consisted of an initial 15 % (v/v) of acetonitrile to 85 % (v/v) in 20 min, followed by additional 10 minutes to re-equilibrate the column at initial conditions. For the analysis of the reaction products of indole-3-pyruvic acid, methyl ester, the mobile phase was methanol and water degassed with helium with a flow rate fixed at 1.0 ml min^{-1}. The gradient consisted of: 0-2 min, 4% methanol; from 2.0 to 6.0 min, 4% → 20%; from 6.0 to 6.5 min, 20% → 60%; from 6.5 to 10.9 min, 60%; from 10.9 to 11.0 min, 60% → 4%; and from 11 to 20 min, 4%. For the analysis of the reaction products of methyl 2-oxo-3-phenylpropanoate, the mobile phase was a gradient with acidified acetonitrile containing formic acid at 0.2 % (v/v), and water degassed with helium. The flow rate was 0.8 ml min^{-1} and the total analysis time was 25 min. The gradient consisted of: from 0 to 4.5 min, 6% → 15%; from 4.5 to 15 min, 60% → 70%; from 15 to 25 min, 6%. In all cases, a calibration curve with standards was made between 0 and 5 mM for 3-OTfBE reaction, between 0 and 10 mM for benzyl 3-oxobutanoate reaction, and between 0 and 2 mM for indole-3-pyruvic acid, methyl ester, and methyl 2-oxo-3-phenylpropanoate reactions.

Mass spectrometry. The samples were analyzed in a mass spectrometer with a hybrid analyzer QTOF model MAXIS II from Bruker. Samples were introduced into the spectrometer by direct infusion using a syringe pump at a flow rate of 3 ul min^{-1}. Methanol with 0.1% formic acid was used as the ionizing phase for positive mode analysis (ESI+) and methanol for negative ion detection mode analysis (ESI-). The assays were performed with exact mass determination using a mixture of phosphazenes (Agilent Technologies) as a calibrant. The Data Analysis v.4.0 spectrometer software allowed studying the probable elemental compositions for each m/z value obtained. The analyses (analysis ID 120-12561) were performed at the Servicio Interdepartamental de Investigación (SIDI) from the Autonomous University of Madrid, Spain.

Protein preparation for *in silico* analysis. The apo TR_2 crystal structure was prepared and protonated at pH 7.5, the pH at which the experimental assays were performed, using the Protein Preparation Wizard.[4] This process includes fixing side chains and loops missing in the crystal structure using Prime software.[5] The ester compound used as a probe to identify noncatalytic binding sites was glyceryl tripropionate; activity simulations of the proposed variants used 3-ATfBE and 3-OTfBE as substrates; migration simulations of the functional variant and the two mutants enclosing the channel used 3-OTfBA as substrate. All substrates were modeled using the OPLS2005 force field,[6] except for the atomic charges, which were calculated with Jaguar using density functional theory[7] with a B3LYP-D3 exchange-correlation functional and the polarized triple-zeta (pVTZ) basis set.

Protein Energy Landscape Exploration (PELE) simulations. PELE was used to identify the noncatalytic ester binding sites in TR2 and to assess if catalytic poses were achieved in the functionalized site variants.[2,8-10] PELE is a Monte Carlo (MC)-based algorithm coupled with protein structure prediction methods.[11] The heuristic MC approach begins with the sampling of different microstates by initially applying small perturbations (translations and rotations) to the ligand. Then, the flexibility of the protein is considered by applying normal modes through the anisotropic network model (ANM) approach. Once the system has been perturbed, side chains of the residues near the ligand are sampled with a library of rotamers to avoid steric clashes. Finally, a truncated Newton minimization with the OPLS2005 force field is performed,[6] and the new microstate is accepted or rejected according to the Metropolis criterion. The Surface Generalized Born Non-Polar (SGBNP) implicit solvent was applied to mimic the effect of water molecules around the protein.[12,13]

Molecular dynamics (MD) simulations. Four replicates of 500 ns of apo MD simulations were performed with OPENMM to assess the integrity of the designed catalytic triad.[14] Likewise, 4 replicates of 250 ns of apo MD simulations were performed to further refine the modeled loops. A water cubic box (distance of 8 Å between the closest protein atom and the edge of the box) was created around the system using the TIP3P water model,[15] and the charge of the system was stabilized using monovalent ions (Na$^+$ and Cl$^-$). The protein system was parameterized with the AMBER99SB force field.[16] The Andersen thermostat[17] and Monte Carlo barostat[18,19] were applied to the NPT ensemble (constant pressure and temperature of 1 bar and 300 K, respectively). The NVT equilibration lasted 400 ps, and a constraint of 10 kcal/(mol·Å^2) was applied to the system, while the NPT equilibration lasted 1 ns, and a milder constraint of 5 kcal/(mol·Å^2) was used. The Verlet integrator,[20] which has a 2 fs time step, was used with constraints between H and heavy atoms. For the nonbonded long-range interactions, a radius of 8 Å was used.

Prediction of $\Delta\Delta G$ for the TR_2 variants. The stability ($\Delta\Delta G$(mut-WT)) of the experimentally tested variants was predicted using the module of thermodynamic stability from HotSpot Wizard, which uses FoldX to repair possible problems in the protein structure and Rosetta to perform the energy minimization and $\Delta\Delta G$ calculation (according to protocol 3 from Rosetta).[21]

Crystallization and structure determination. TR_2 was crystallized using hanging-drop vapor diffusion at 12 °C. A total of 1.5 µl of 10 mg ml^{-1} TR_2 in 20 mM HEPES, pH 7.5, 300 mM NaCl, and 160 mM imidazole, 1.5 µl of reservoir solution consisting of 4% (w/v)

SUPPORTING INFORMATION

PEG 6000 and 100 mM citric acid, pH 4, and 0.3 µl of 200 mM NDSB 211 as a crystallization additive were mixed. This drop was equilibrated against the reservoir solution, and crystals formed after 1-2 d. Crystals were harvested, immersed in mineral oil for cryoprotection, and flash frozen, and diffraction data were collected at beamline ID30 B (ESRF, Grenoble, France). The TR2 crystals diffracted to 3.8 Å resolution. Crystallization of TR_2E_2 was achieved using hanging-drop vapor diffusion at 4 °C. Two microliters of 7-9 mg ml^{-1} TR_2E_2 in 300 mM NaCl and 20 mM HEPES, pH 7.5, were mixed with 1 µl of reservoir solution consisting of 200 mM calcium acetate, 100 mM HEPES, pH 7.5, and 10% (w/v) PEG 8000. A total of 0.2 µl of seeding solution consisting of crushed TR_2E_2 crystals in reservoir solution was added to the crystallization drops, which were then equilibrated against the reservoir solution, and crystals were observed after 1-3 days. The crystals were cryoprotected by immersing them in reservoir solution with 25% glycerol and flash frozen. Diffraction data were collected at beamline ID 30A-3 (ESRF, Grenoble, France). TR_2E_2 crystallized in space group P212121, and the crystals diffracted to a 3.3 Å resolution. Crystals were prepared as described above to determine the structure of TR_2E_2 bound to PLP. A total of 0.2 µl of 30 mM PLP in meso 2,3-dimercaptosuccinic acid (DMSA) was added to the crystallization drops, and the crystals were soaked for 16 h. The crystals were cryoprotected by immersing them in the reservoir solution containing 25% ethylene glycol and 1 mM PLP and flash frozen, and diffraction data were collected at beamline P13 (EMBL, Hamburg, Germany). Crystals of TR2E2 bound to PLP diffracted to a 3.5 Å resolution. The structure of TR_2E_2 bound to ethanolamine O-sulfate (EOS) was determined by preincubating the protein with 1 mM EOS at 4 °C for at least 1 h. Crystals were produced as described for TR_2E_2 without the inhibitor and cryoprotected by immersing the crystals in a reservoir solution containing 1 mM EOS and 25% ethylene glycol and flash freezing. Diffraction data were collected at beamline P13 (EMBL, Hamburg, Germany). The structure of TR_2E_2 bound to EOS crystallized in space group P212121, and the crystals diffracted to 3.6 Å resolution.

Results and Discussion

Table S1. A) Data collection and refinement statistics. B) Similarity between the TR_2 structure and other known structures was calculated using Ebi-fold (https://www.ebi.ac.uk/msd-srv/ssm/). The ten structures are highlighted.

A)

	TR2 (7QYG)[1]	TR2E2 (7QYF)[1]	TR2E2 EOS (7QX3)[1]	TR2E2 PLP (7QX0)[1]
Wavelength	0.9762	0.9762	0.9762	0.9762
Resolution range	49.09 - 3.6 (3.729 - 3.6)	39.65 -3.3 (3.418 - 3.3)	48.94 - 3.6 (3.729 - 3.6)	48.62 - 3.5 (3.625 - 3.5)
Space group	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	60.776 166.626 209.821 90 90 90	100.358 108.988 212.178 90 90 90	100.61 109.502 211.455 90 90 90	99.921 107.627 211.637 90 90 90
Total reflections	48221 (3588)	176427 (18057)	360952 (37338)	189998 (18912)
Unique reflections	24144 (1812)	34886 (3500)	27774 (2733)	29483 (2892)
Multiplicity	2.0 (2.0)	5.1 (5.2)	13.0 (13.7)	6.4 (6.5)
Completeness (%)	94.21 (71.93)	98.45 (98.01)	99.85 (99.96)	99.80 (99.93)
Mean I/sigma(I)	10.23 (1.95)	7.83 (1.43)	15.38 (7.71)	15.40 (3.97)
Wilson B-factor	113.55	109.49	78.95	107.15
R-merge	0.04228 (0.4077)	0.1872 (1.212)	0.1731 (0.6662)	0.08209 (0.5359)
CC1/2	1 (0.887)	0.994 (0.643)	0.997 (0.977)	0.999 (0.915)
Reflections used in refinement	24087 (1794)	35219 (3442)	27750 (2733)	29459 (2892)
Reflections used for R-free	1995 (149)	1933 (191)	965 (89)	1267 (128)
R-work	0.3451 (0.3999)	0.2260 (0.3489)	0.1606 (0.1998)	0.2000 (0.2451)
R-free	0.4222 (0.4890)	0.2797 (0.3600)	0.2183 (0.2962)	0.2866 (0.3519)
CC(work)	0.883 (0.711)	0.944 (0.513)	0.966 (0.923)	0.953 (0.883)
CC(free)	0.831 (0.555)	0.862 (0.386)	0.963 (0.761)	0.899 (0.779)
Number of non-hydrogen atoms	12076	12929	13069	13408
macromolecules	12076	12929	13068	13344
ligands				64
solvent			1	
Protein residues	1558	1672	1687	1724
RMS(bonds)	0.007	0.012	0.008	0.007
RMS(angles)	1.04	1.48	1.25	1.20

Ramachandran favored (%)	72.97	86.79	90.27	87.40
Ramachandran allowed (%)	20.06	11.28	7.46	9.91
Ramachandran outliers (%)	6.97	1.93	2.27	2.70
Rotamer outliers (%)	0.08	4.61	0.00	0.00
Clashscore	34.60	15.38	17.11	25.21
Average B-factor	130.73	96.81	78.21	99.65
macromolecules	130.73	96.81	78.21	99.68
ligands				93.77
solvent			46.13	

[1]Statistics for the highest-resolution shell are shown in parentheses.

B)

PDB code	Rmsd (Å)	% seq identity
4a6r	1.29	52
4ba5	1.57	52
5ti8	1.12	61
5ghg	1.47	38
6hx9	1.17	60
4a6u	1.20	51
5ghf	1.66	38
5ztx	1.72	32
3nui	1.8	32
4ysv	1.99	26

Table S2. Number of catalytic events of TR_2 variants determined from the local exploration.

PluriZyme variant	Number of catalytic events	Site
TR_2E_7	377	A
TR_2E_{16}	326	A
TR_2E_9	1	D
TR_2E_2	3	G
TR_2E_{18}	17	E
TR_2E_{10}	1	B
TR_2E_{19}	27	B
TR_2E_3	0	C
TR_2E_{11}	13	B
TR_2E_{12}	0	B
TR_2E_8	742	A
TR_2E_{13}	14	G
TR_2E_1	0	E
TR_2E_{17}	0	C
TR_2E_{20}	19	B
TR_2E_5	4	F
TR_2E_{14}	13	C
TR_2E_4	17	E
TR_2E_{15}	0	C
TR_2E_6	0	E

Table S3. Computational score (from 0 to 10) of TR_2 variants based on PELE & MD simulations and calculations of $\Delta\Delta G$ for thermo-stability.

PluriZyme variant	Computational score	Site
TR_2E_7	8.0	A
TR_2E_{16}	8.0	A
TR_2E_9	7.5	D
TR_2E_2	7.25	G
TR_2E_{18}	7.25	E
TR_2E_{10}	7.0	B
TR_2E_{19}	7.0	B
TR_2E_3	6.875	C

TR$_2$E$_{11}$	6.75	B
TR$_2$E$_{12}$	6.75	B
TR$_2$E$_8$	6.75	A
TR$_2$E$_{13}$	6.75	G
TR$_2$E$_1$	6.5	E
TR$_2$E$_{17}$	6.5	C
TR$_2$E$_{20}$	6.375	B
TR$_2$E$_5$	5.5	F
TR$_2$E$_{14}$	5.5	C
TR$_2$E$_4$	5.25	E
TR$_2$E$_{15}$	4.25	C
TR$_2$E$_6$	4.0	E

Table S4. Effect of mutations on the transaminase (TR) activity of TR$_2$.

ID	Mutations	Nr. of mutations	k_{cat} [min^{-1}] (hexanal)
TR$_2$	WT	-	83.38±4.04
TR$_2$E$_1$	K110D	1	76.71±2.78
TR$_2$E$_2$	A172S Q173H	2	69.49±1.77
TR$_2$E$_3$	W272S V107H	2	-
TR$_2$E$_4$	M112H M302D	2	-
TR$_2$E$_5$	G420S G399H	2	68.73±0.86
TR$_2$E$_6$	M112H T108D	2	-
TR$_2$E$_7$	V377S Y277H T271D	3	-
TR$_2$E$_8$	V377S Y277H Y278D	3	-
TR$_2$E$_9$	Q52S Y44H Q435E	3	-
TR$_2$E$_{10}$	F190S Y188H I236D	3	-
TR$_2$E$_{11}$	F190S Y188H G191D	3	55.05±2.22
TR$_2$E$_{12}$	F190S Y188H R192E	3	48.22±1.67
TR$_2$E$_{13}$	A169S M167H M170D	3	-
TR$_2$E$_{14}$	W272S I347H E348D	3	-
TR$_2$E$_{15}$	V107S W272H G351K	3	-
TR$_2$E$_{16}$	V377S Y277H T271D Q276E	4	-
TR$_2$E$_{17}$	V107S W272H I347D G351K	4	-
TR$_2$E$_{18}$	A251S D283H R306D	3	67.92±12.93
TR$_2$E$_{19}$	F190S E384H T422D R375A	4	-
TR$_2$E$_{20}$	F190S E373H E384D R375G R394G	5	-
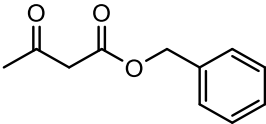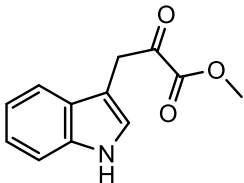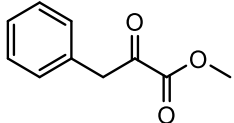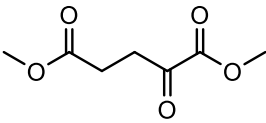
Table S5. k_{cat} of TR$_2$E$_2$ with different chemically diverse esters.

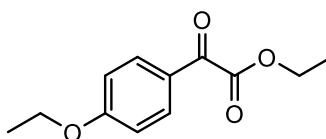| Ester	k_{cat} [min^{-1}]
1-Napthyl acetate	2104±48
1-Naphthyl butyrate	53.45±11.52
Glyceryl triacetate	44.85±2.15
Glyceryl tripropionate	388.1±17.5
Glyceryl tributyrate	284.9±12.4
Methyl-2-bromobutyrate	82.52±18.64
Hexyl acetate	4.08±0.12
Octyl acetate	10.75±0.98
Dodecanoyl acetate	2.23±0.33
Pentadecyl acetate	8.37±0.83
Ethyl propionate	93.04±1.69
Ethyl butyrate	5.92±2.1
Ethyl hexanoate	85.91±1.35
(1R)-(-)-Menthyl acetate	4.43±0.87
Methyl (R)-(-)-mandelate	53.02±5.37
Methyl (S)-(+)-mandelate	351.4±10.59
Ethyl (R)-(+)-4-chloro-3-hydroxybutyrate	16.98±0.83
Ethyl (S)-(−)-4-chloro-3-hydroxybutyrate	11.30±1.07
(+)-Ethyl D-Lactate	9.04±0.36
(−)-Ethyl L-lactate	17.73±0.59
(+)-Methyl (S)-3-hydroxybutyrate	8.89±0.75
(-)-Methyl (R)-3-hydroxybutyrate	5.14±0.76
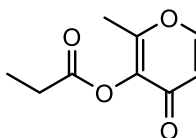
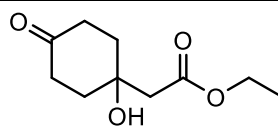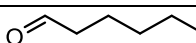|	
(−)-Methyl (R)-3-hydroxyvalerate	12.11±0.68
(+)-Methyl (S)-3-hydroxyvalerate	124.86±15.54
Benzyl (R)-(+)-2-hydroxy-3-phenylpropionate	4.66±0.87
Methyl benzoate	1.2±0.28
Methyl butyrate	7.12±2.93
Propyl hexanoate	3.17±0.17
Isobutyl cinnamate	5.18±0.7
Methyl 2,5-dihydroxycinnamate	6.18±0.4
Methyl cinnamate	1.76±0.25
Methyl ferulate	18±2.3
Vinyl acetate	17.89±1.78
Vinyl propionate	1310±11
Vinyl butyrate	523.18±129
Vinyl benzoate	17.6±3.39
Geranyl acetate	48.44±6.78
3-Methyl-3-buten-1-yl acetate	8.49±0.16
Ethyl 2-ethylacetoacetate	18.22±1.38
Ethyl 2-methylacetoacetate	206.84±0.95
Ethyl 3-oxohexanoate	313.47±76.82
Ethyl acetoacetate	233.54±31.04
Ethyl propionylacetate	434.29±13.86
Y-Valerolactone	20.11±0.83
Methyl glycolate	3.37±0.81
Cyclohexyl butyrate	15.23±1.04
(+)-Methyl D-Lactate	13.22±0.86
(-)-Methyl L-Lactate	18.75±1.65
Propyl acetate	1.46±0.31
Butyl acetate	5.24±2.8
Phenyl acetate	2915±93
Phenyl propionate	3908±66

Table S6. Specific activity of TR_2E_2 towards a series of chemically and structurally diverse β-keto esters.

Substrate	Structure	Hydrolase activity [U/g][a]	Transaminase activity [U/g][b]
Benzyl 3-oxobutanoate		391.5±5.8	12.36±0.05
Indole-3-pyruvic acid methyl ester		350.8±5.4	6.16±0.29
Methyl 2-oxo-3-phenylpropanoate		313.41±0.59	13.91±1.10
Dimethyl 2-oxoglutarate		253.38±3.04	25.74±1.83

2-Methyl-4-oxo-4H-pyran-3-yl propionate		187.2±1.43	n.d.
Ethyl (4-ethoxyphenyl)(oxo)acetate		5.70±0.50	1.43±0.01
Ethyl (1-hydroxy-4-oxocyclohexyl)acetate		n.d.	1.86±0.01
Hexanal		n.d.	36.05±0.04

[a] Reactions conditions as follows: - [protein]: 0.045-0.45 mg ml^{-1} (depending on the substrate); [substrate]: 20 mM; reaction volume: 44 μl; T: 40 °C; pH: 8.0 (5 mM EPPS buffer, plus 0.45 Phenol Red®). [b] Reactions conditions as follows: - [protein]: 0.5 mg ml^{-1}; [substrate]: 20 mM; [PLP]: 1 mM; [NPEA]: 20 mM; reaction volume: 200 μl; T: 40 °C; pH: 7.5 (100 mM K_2HPO_4 buffer). As control, hexanal was included. Raw data in Figure S11. n.d. activity not detected under our assay conditions.
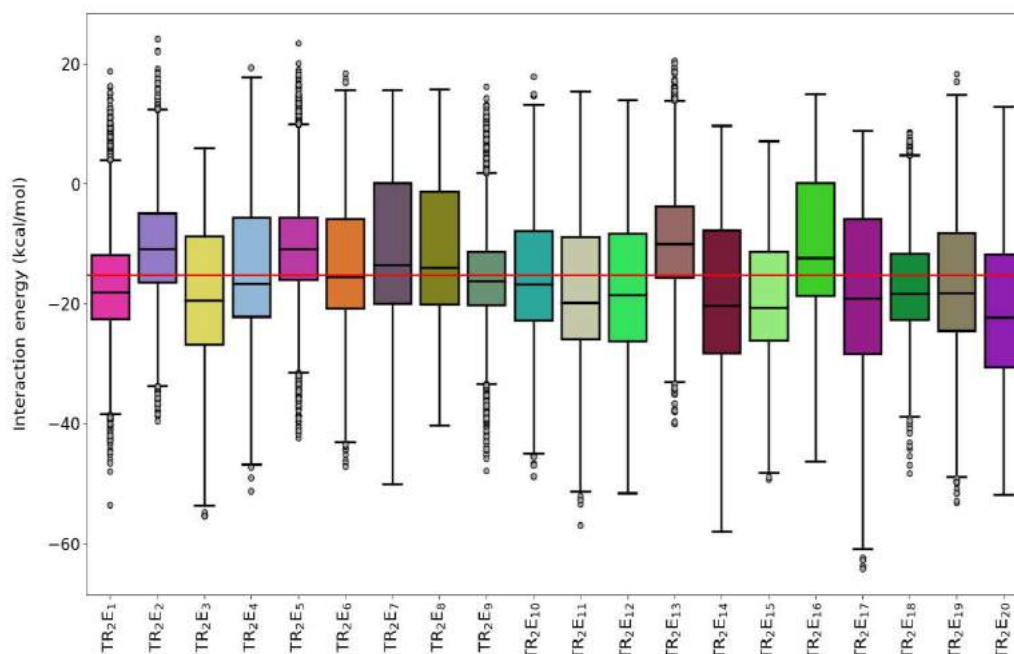
Figures

Figure S1. Box plot representing the interaction energy along all the accepted PELE steps from the local explorations of all variants. The red line indicates the average value of the metric for all variants. The figure was created with the Matplotlib library.[22]

Figure S2. Box plot representing the serine-histidine distance along all the accepted PELE steps from the local explorations of all variants. The red line indicates the average value of the metric for all variants. The figure was created with the Matplotlib library.[22]
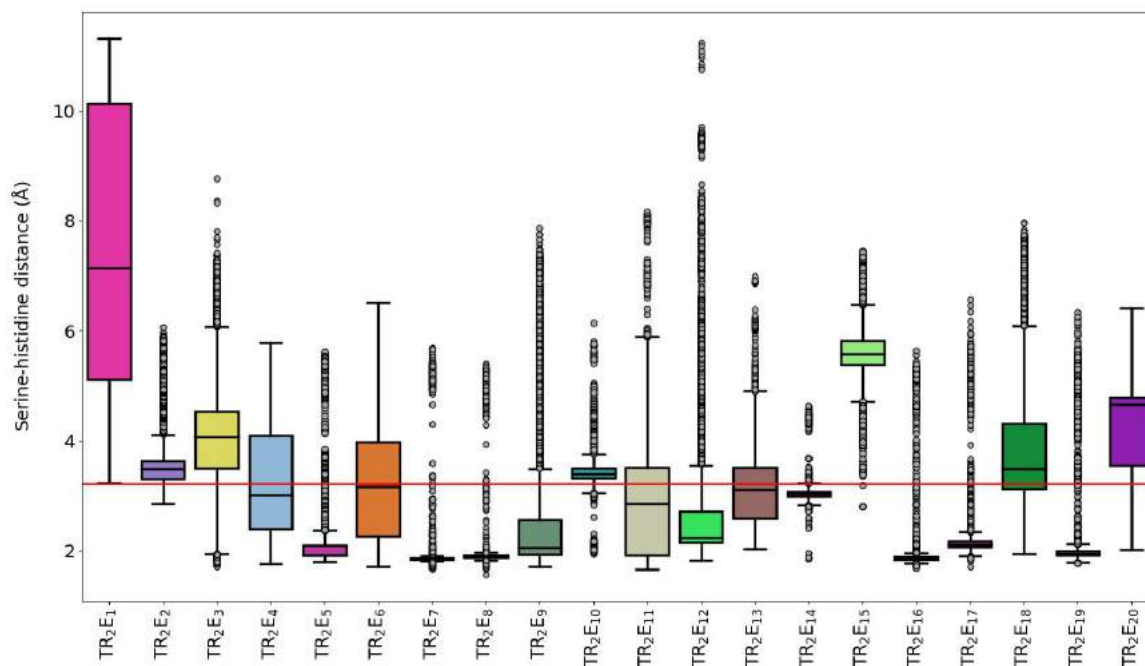

Figure S3. Box plot representing the acid-histidine distance along all the accepted PELE steps from the local explorations of all variants. The red line indicates the average value of the metric for all variants. The figure was created with the Matplotlib library.[22]
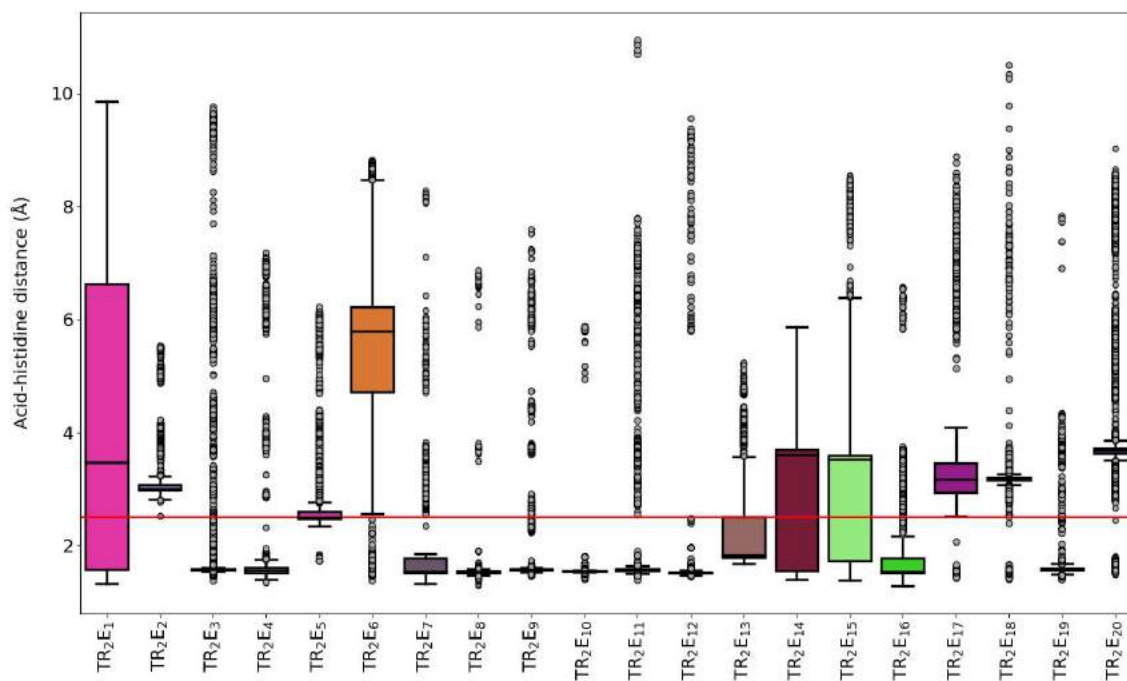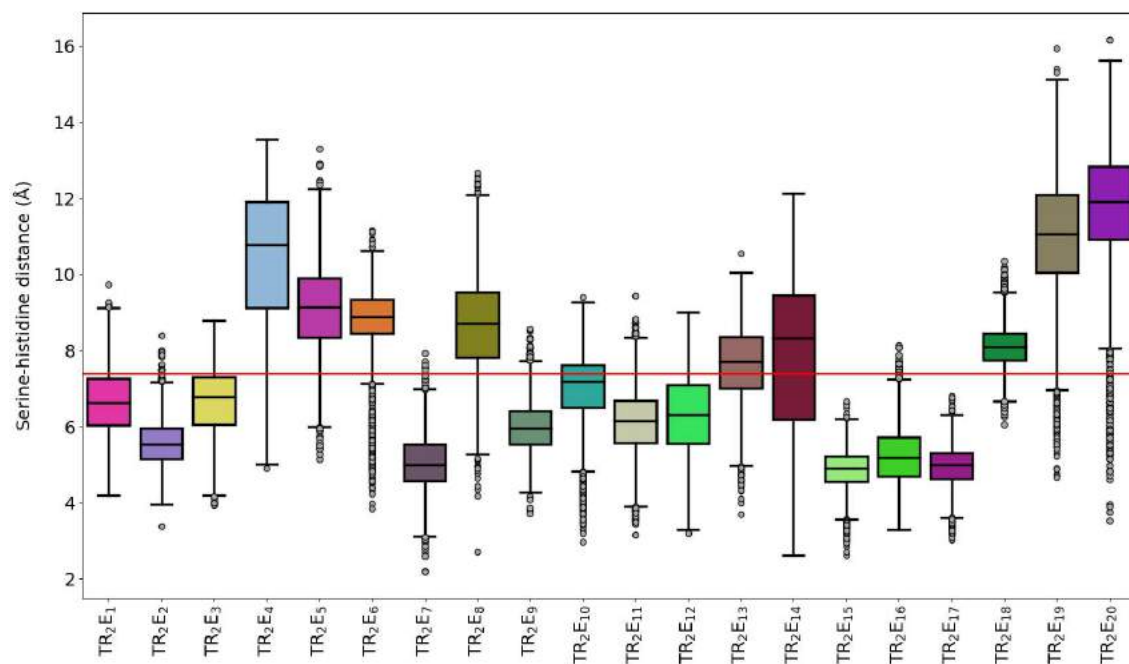
Figure S4. Box plot representing the serine-histidine distance ($d_{H_\gamma^{Ser} - N_\varepsilon^{His}}$) along the 500 ns of the 4 MD replicates performed for all variants. The figure was created with the Matplotlib library.[22]
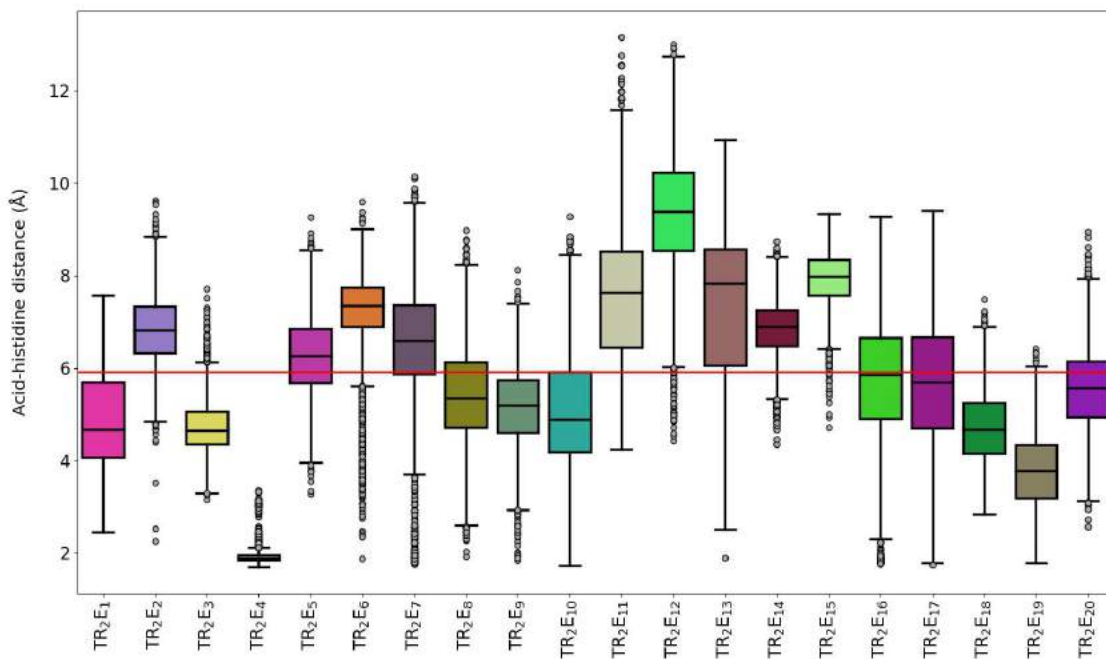
Figure S5. Box plot representing the acid-histidine distance distance ($d_{O_\delta^{Asp} - H_{\delta N}^{His}}$) along the 500 ns of the 4 MD replicates performed for all variants. The figure was created with the Matplotlib library.[22]
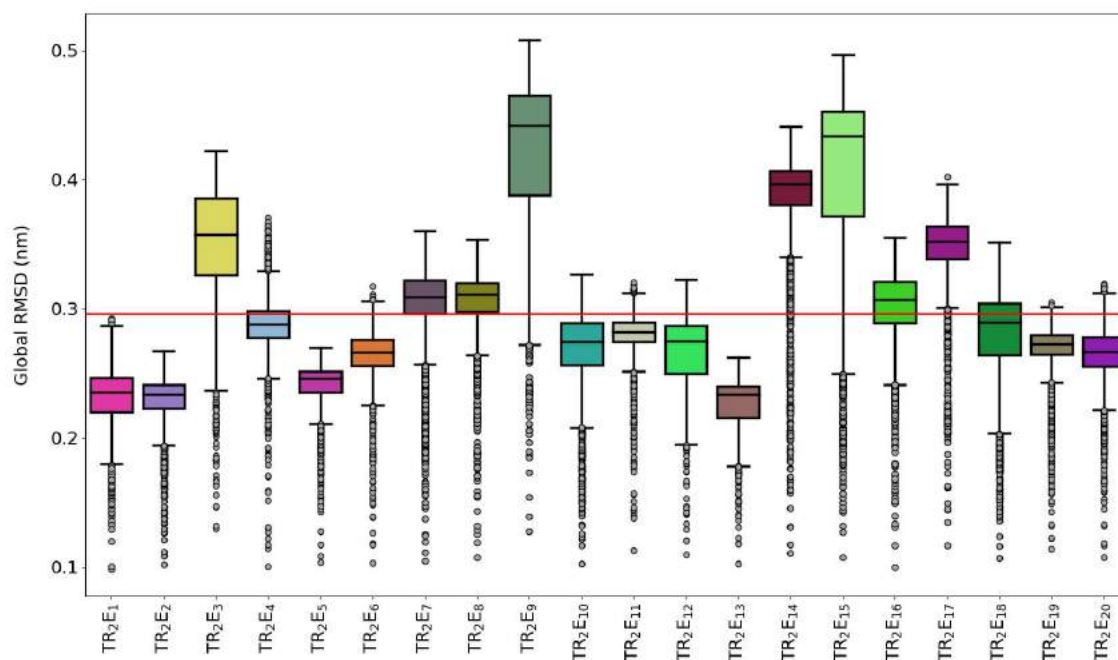
Figure S6. Box plot representing the overall RMSD along the 500 ns of the 4 MD replicates (excluding the initial frame) performed for all variants. The figure was created with the Matplotlib library.[22]
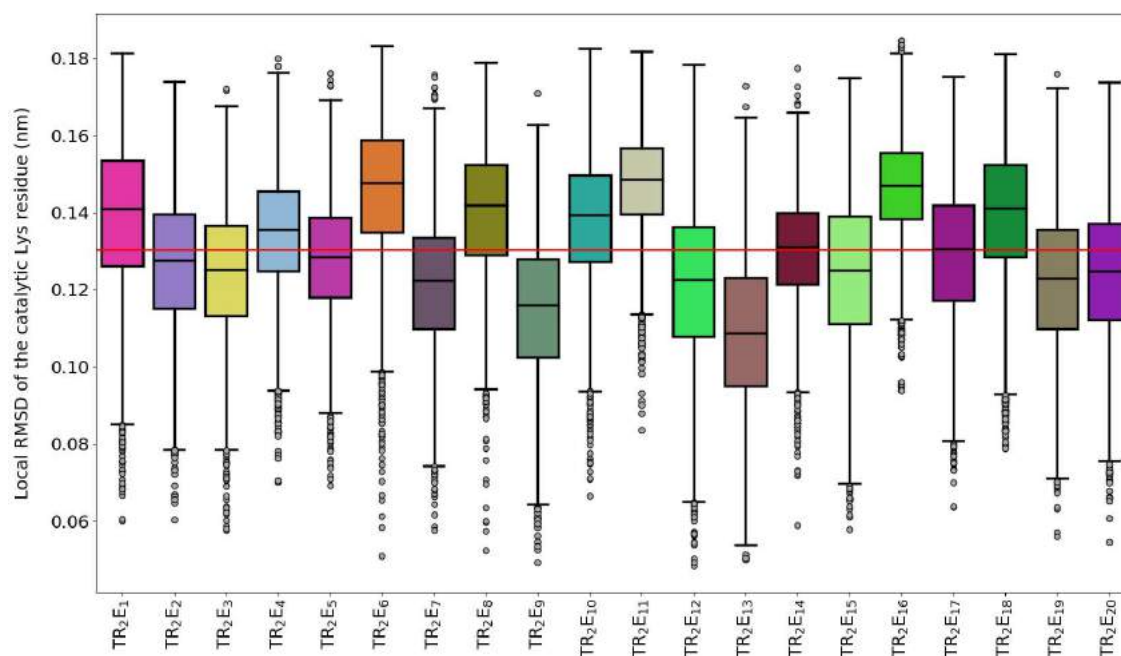
Figure S7. Box plot representing the local RMSD of the catalytic lysine residue in the transaminase site along the 500 ns of the 4 MD replicates (excluding the initial frame) performed for all variants. The figure was created with the Matplotlib library.[22]
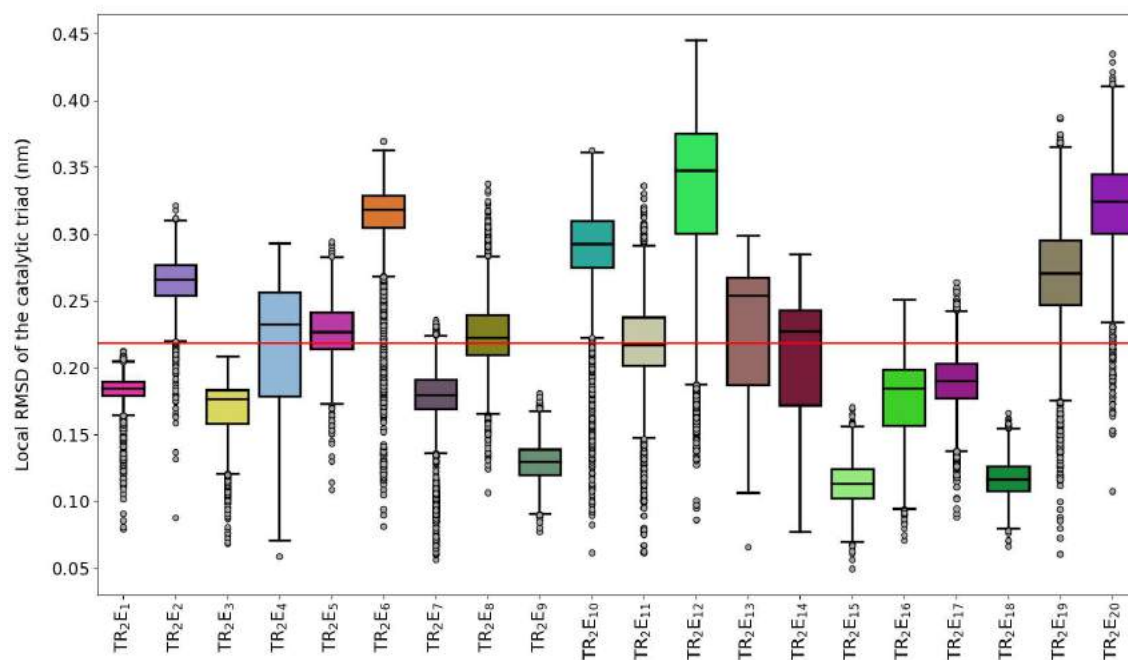
Figure S8. Box plot representing the local RMSD of the catalytic triad in the added hydrolase site along the 500 ns of the 4 MD replicates (excluding the initial frame) performed for all variants. The figure was created with the Matplotlib library.[22]

Figure S9. Visualization of the transaminase activity assays of the purified TR_2 mutants (reaction: 8 h) using hexanal as the substrate.
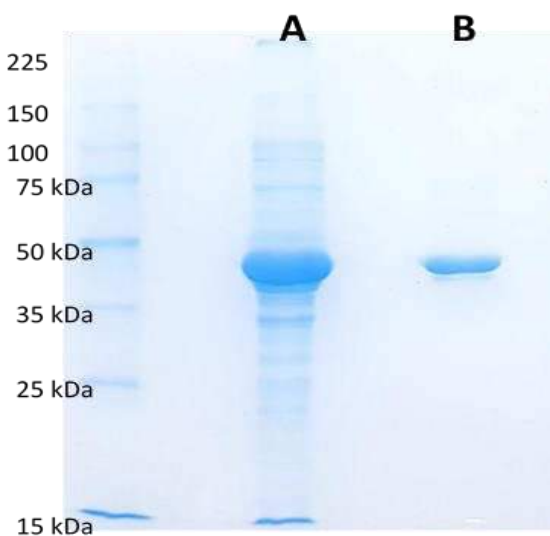
Figure S10. SDS–PAGE analysis of the TR_2E_2 mutant using a 12% Tris-glycine SDS-polyacrylamide gel. Lane A: Soluble proteins in culture. Lane B: 0.005 mg of TR_2E_2.

13
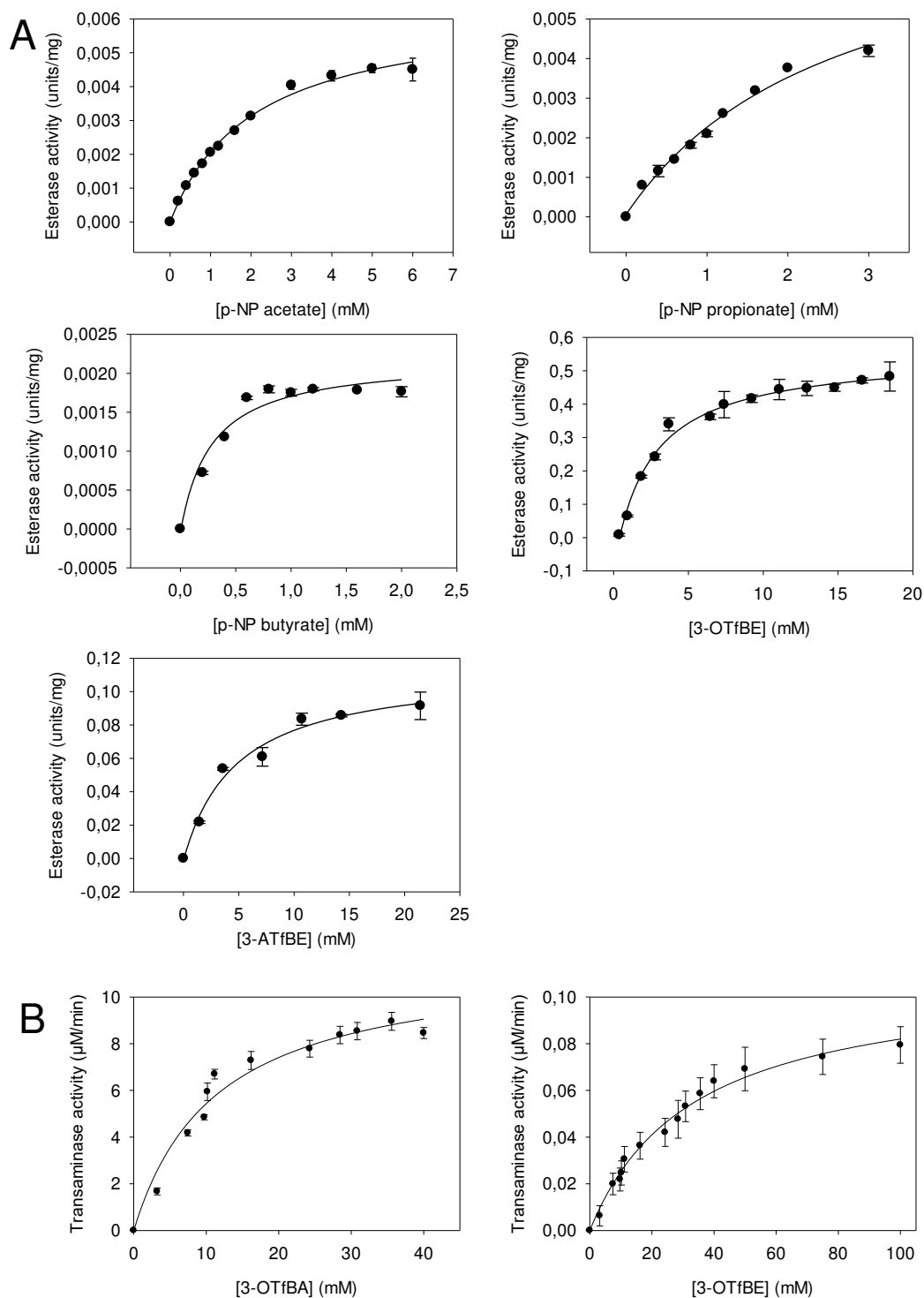
Figure S11. Determination of K_m and k_{cat} for TR$_2$E$_2$. Curves representing the esterase (A) and transaminase (B) activities vs substrate concentration are plotted. The figure was made with SigmaPlot 14.0, and the K_m values reported in the text were obtained from the best linear fit using SigmaPlot 14.0. k_{cat} was calculated assuming v = k_{cat} when v = µM product/min/µM enzyme. Reaction conditions as described in Table 1. Activities were calculated according to Equations 1-3.

A)
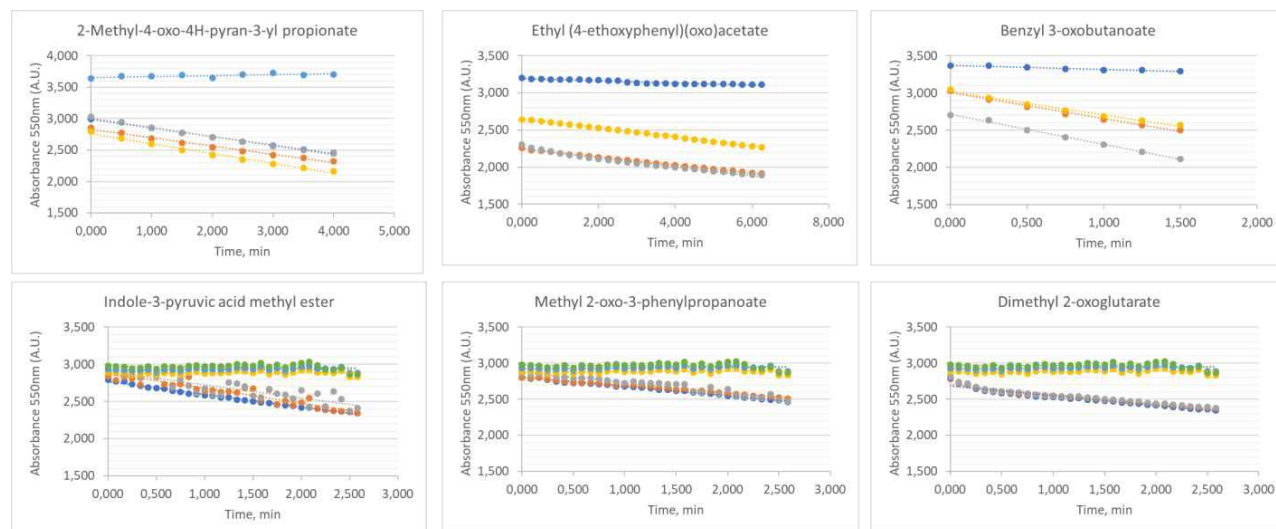
B)
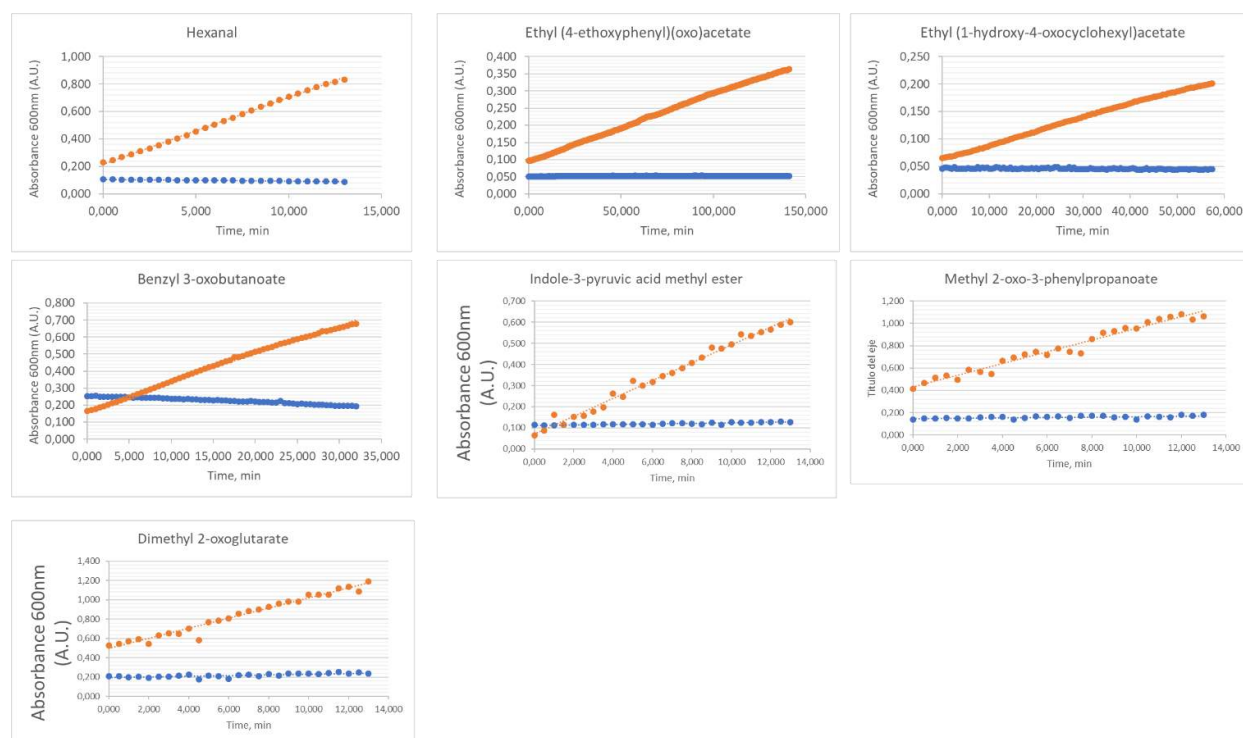
Figure S11 cont. Representative time course curve for the hydrolysis (A) and transamination (B) of β-keto esters. Reaction conditions as described in Table 2 and Experimental Section. In A, the figure representing the variation of absorbance at 550 nm (in arbitrary units) vs time is shown (controls are represented in the upper points/lines in each panel). In B, the figure representing the variation of absorbance at 600 nm (in arbitrary units) vs time is shown (control in the blue color of each panel). The figure was made with Excel 2019. The time course for a control reaction (no enzyme added) is also shown in the upper part of panel A, or with blue color in panel B. Activities were calculated according to Equations 2 and 3.
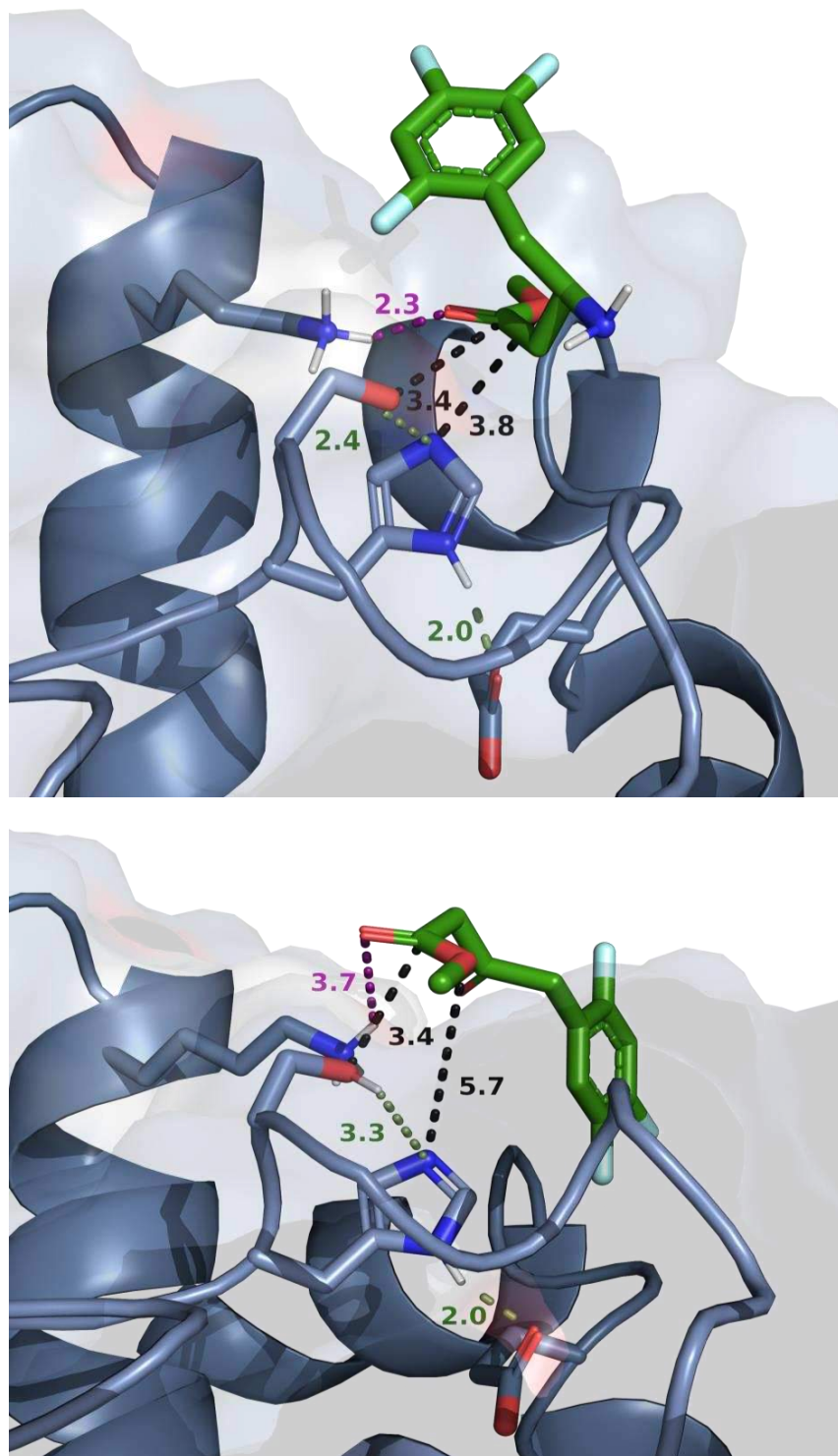
Figure S12. Representative catalytic poses of 3-ATfBE and 3-OTfBE in the hydrolase site of TR$_2$E$_2$. Both the ligand and the catalytic residues are shown as ball-and-stick models. Catalytic triad distances (in Å) are shown in green, the substrate-serine/histidine distances are shown in black, and the oxyan-ion-hole distances are shown in purple.
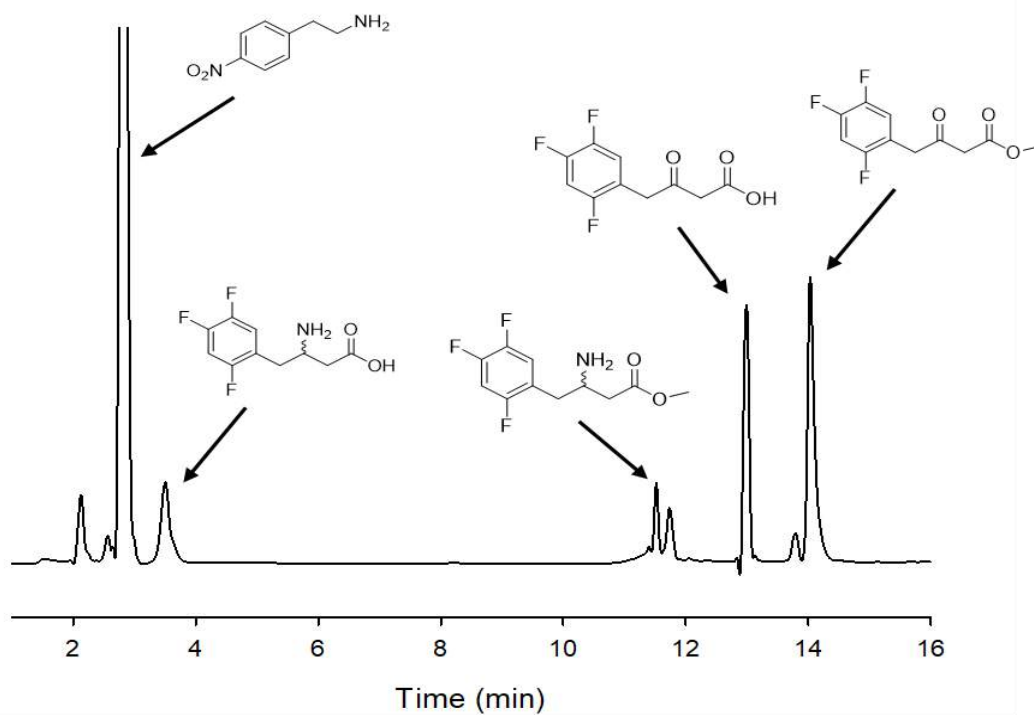
Figure S13. HPLC chromatograms representing the substrates, intermediates, and final products obtained when converting 3-OTfBE in the presence of TR_2E_2.
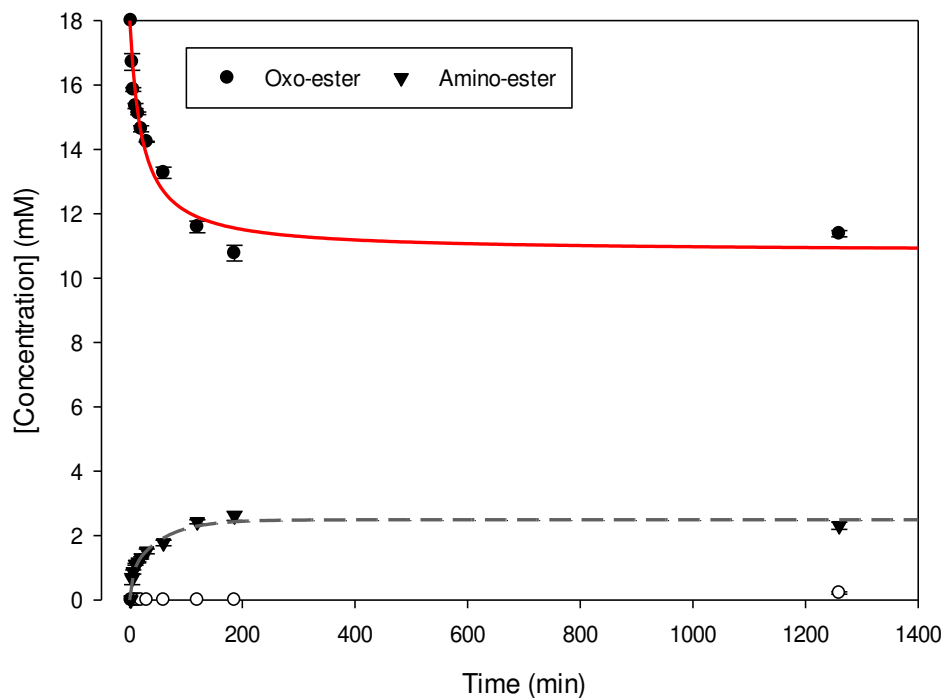
Figure S14. Concentrations of substrates, intermediates, and final products obtained when converting 3-OTfBE in the presence of TR_2.
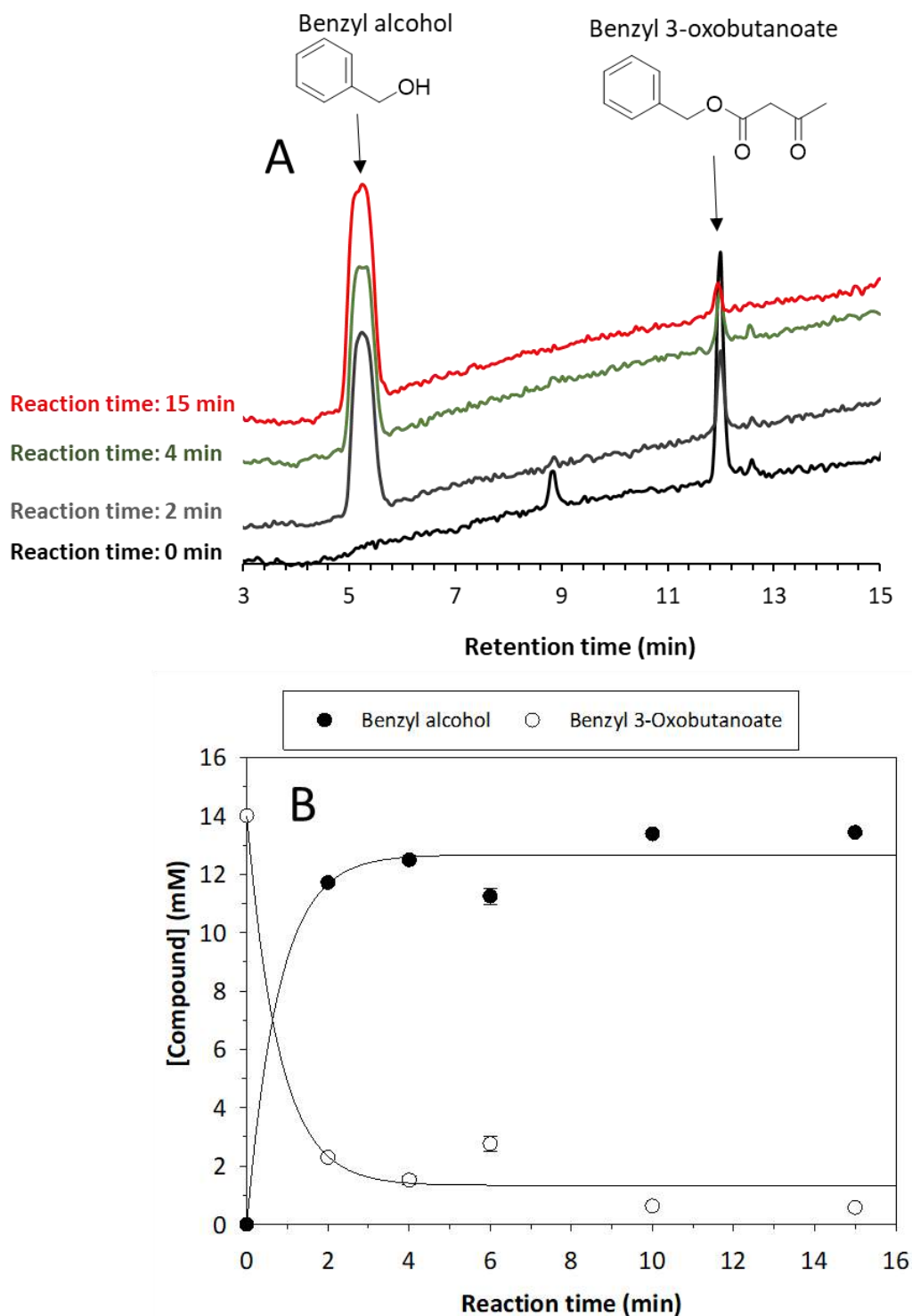
Figure S15. HPLC chromatograms representing the follow-up (A) and concentration (B) of reaction substrate and product obtained when converting benzyl 3-oxobutanoate in the presence of TR_2E_2. Conditions as in Figure 6. Note that detection was performed using a photodiode array detector that allows detecting at 220 nm the substrate (β-ketoester, benzyl 3-oxobutanoate) and the reaction product benzyl alcohol. However, it does not allow the detection of the possible reaction products 3-oxobutyric acid and 3-aminobutyric acid, whose presence was confirmed by HR-MS obtained by ESI-MS analysis (Figure S16).

18

A)

B)

D)
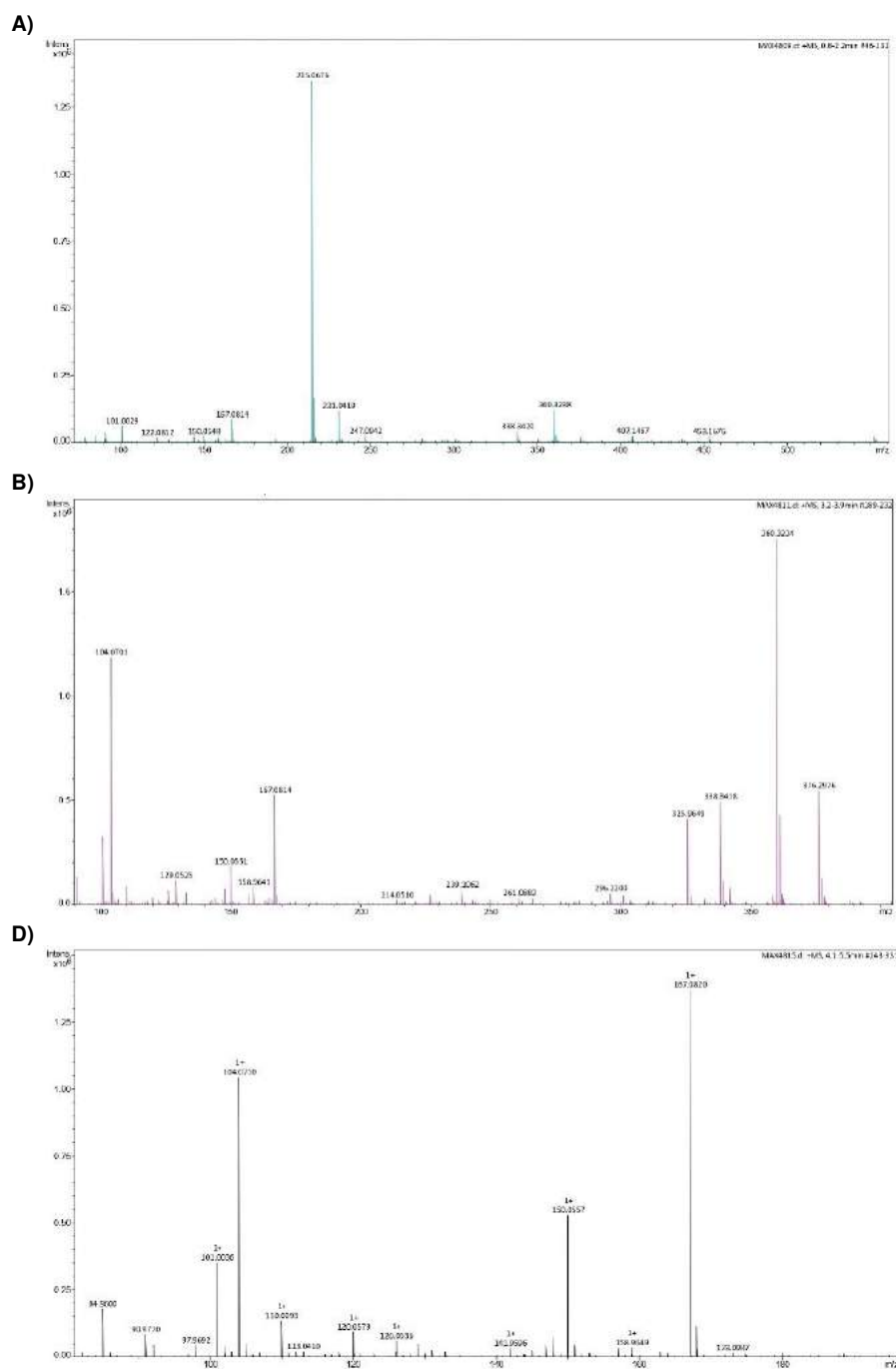
Figure S16. HR-MS spectra (obtained by ESI-MS analysis) of reaction mixtures containing the β-ketoester BOB (initial concentration 14 mM) with TR$_2$E$_2$. Shown are the ESI-MS of samples at times 0 (A), 2 (B,C) and 60 (D) min. Note: original figure provided by the Servicio Interdepartamental de Investigación (SIDI) from the Autonomous University of Madrid, Spain; raw data (Analysis ID 120-12561) can be provided upon request to María Jesús Vicente Arana (mjesus.vicente@uam.es), responsible for the Molecular Mass Spectrometry Unit. A) In the samples at time 0, the expected product is observed, which ionizes preferentially by sodium capture, giving the [M+Na]$^+$ type ion at *m/z* 215.0676, this being the base peak of the spectrum. The [M+K]$^+$ at m/z 231.0419 is also observed. The determination of the exact mass is consistent with the composition C$_{11}$H$_{12}$O$_3$ with an error of less than 5 ppm with respect to the theoretically calculated mass. B) Using methanol+0.1% formic acid (C) as ionizing phase, the species at *m/z* 104.0701 is observed, which seems to correspond by its exact mass to the composition C$_4$H$_{10}$NO$_2$ ([M+H]), which could be the 3-aminobutanoic compound (the 10 most referenced structures found in the Scifinder database for this composition). C) In the sample at the time 20 min, the compound amino butanoic acid (*m/z* 104.0710) found in the sample at the time 2 min is obtained again. The complete analysis and raw datasets are available upon request to the authors.

19
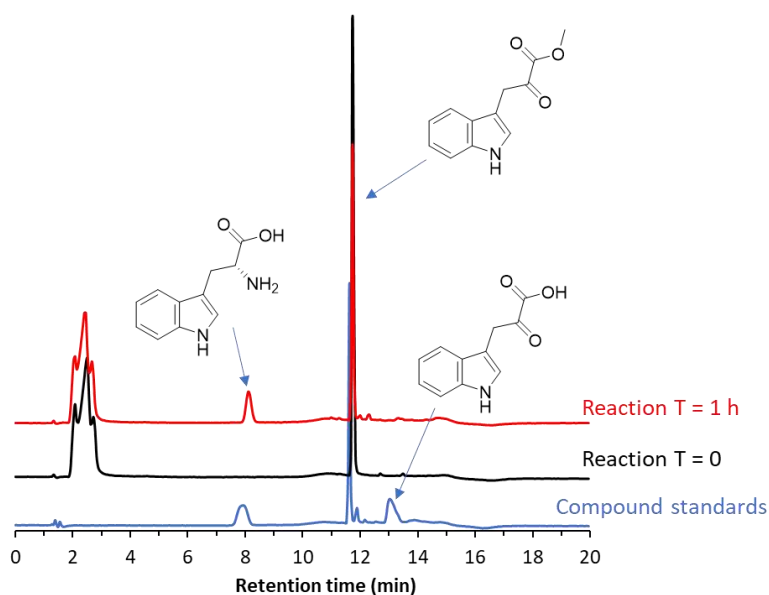
Figure S17. HPLC chromatograms representing the one-pot reaction substrate and product obtained when converting indole-3-pyruvic acid methyl ester to tryptophan in the presence of TR_2E_2. Conditions as in Figure 6. Note that detection was performed using a photodiode array detector that allows detecting at 280 nm the substrate and the reaction product. Note: the peaks at time 2-3 minutes include the reactions constituents, PLP, and amine donor. Standards at concentrations of 500 µM are shown.
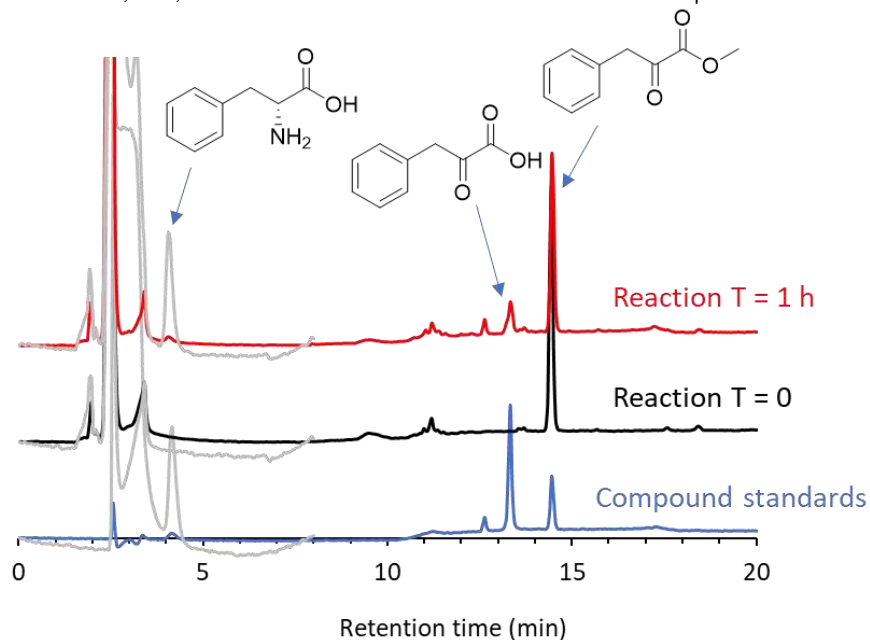
Figure S18. HPLC chromatograms representing the one-pot reaction substrate and product obtained when converting methyl 2-oxo-3-phenylpropanoate to phenylalanine in the presence of TR_2E_2. Conditions as in Figure 6. The substrate and the intermediate were quantified at 256 nm, and the product at 210 nm (superimposed in grey color).
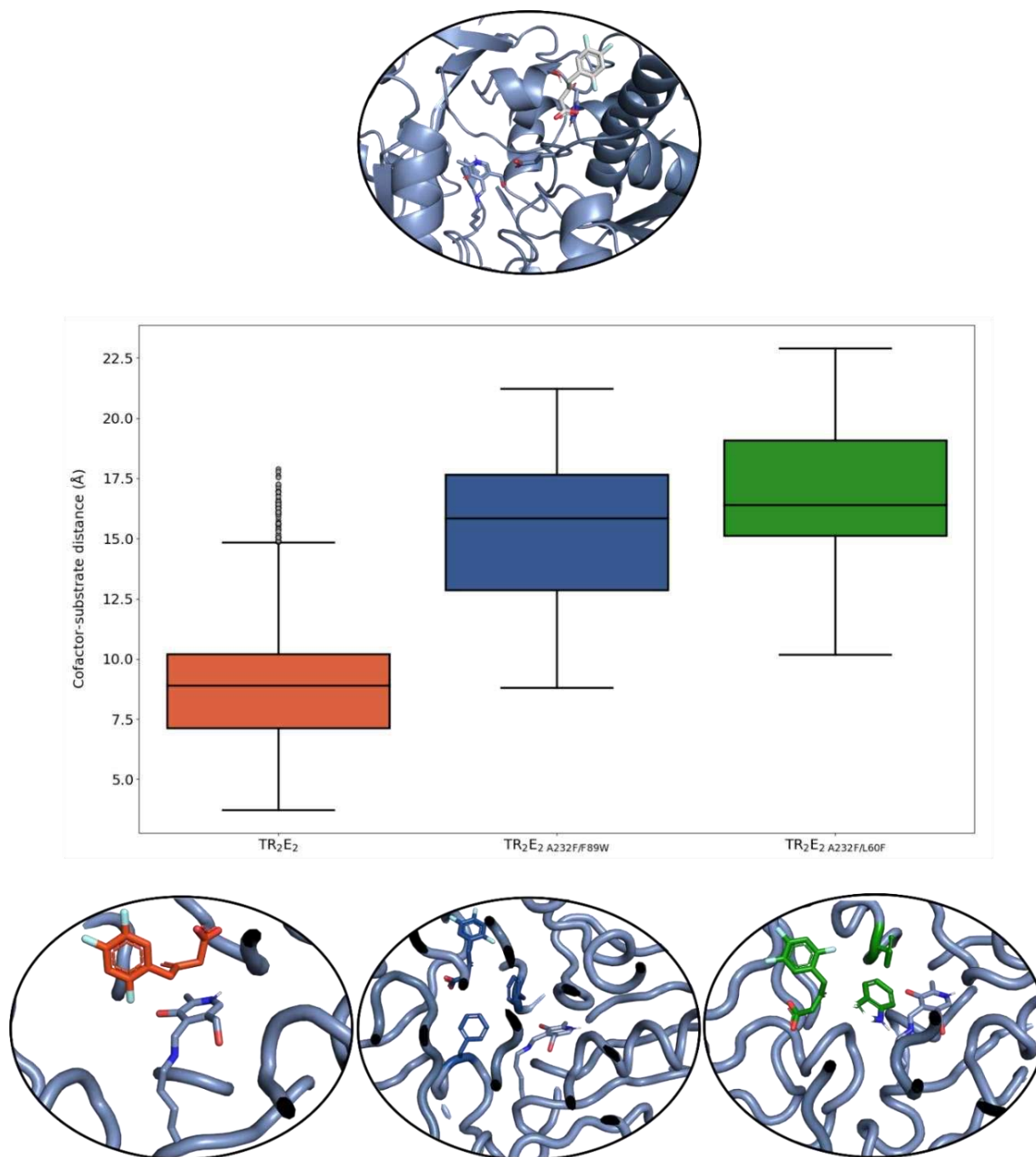
Figure S19. Initial pose from where PELE's migration simulations started in TR_2E_2, $TR_2E_2A232F/L60F$, and $TR_2E_2A232F/F89W$ being near the esterase site (top). Box plot representing the cofactor-substrate distance (distance to the transaminase site) of all the accepted PELE steps with a relative solvent-accessible surface area of 0.1 or lower from the 3-OtfBA simulations (middle). Bottom, representative 3-OtfBA catalytic poses in the transaminase site of TR_2E_2, $TR_2E_2A232F/L60F$, and $TR_2E_2A232F/F89W$ (bottom). Both the ligand and the catalytic cofactor are shown as ball-and-stick models. The figure was created with the Matplotlib library.
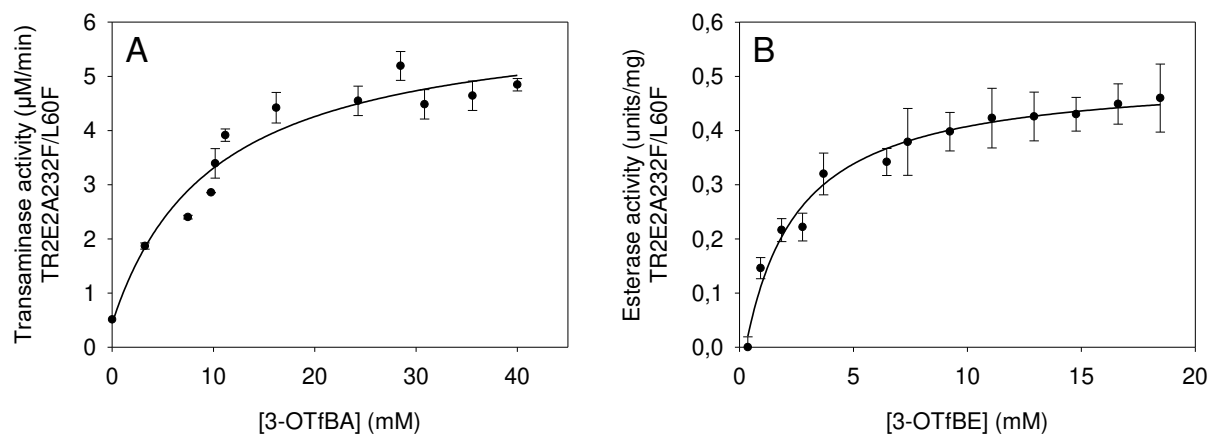
Figure S20. Determination of K_m and k_{cat} of TR$_2$E$_2$A232F/L60F for 3-OtfBA (representing transaminase activity) and 3-OtfBE (representing esterase activity). Curves representing the transaminase (A) and esterase (B) activities vs substrate concentration are plotted. The figure was made with SigmaPlot 14.0, and the K_m values reported in the text were obtained from the best linear fit using SigmaPlot 14.0. k_{cat} was calculated assuming v = k_{cat} when v = μM product/min/μM enzyme. Reaction conditions as described in Table 1 and Fig. S11. Activities were calculated according to Equations 1-3.
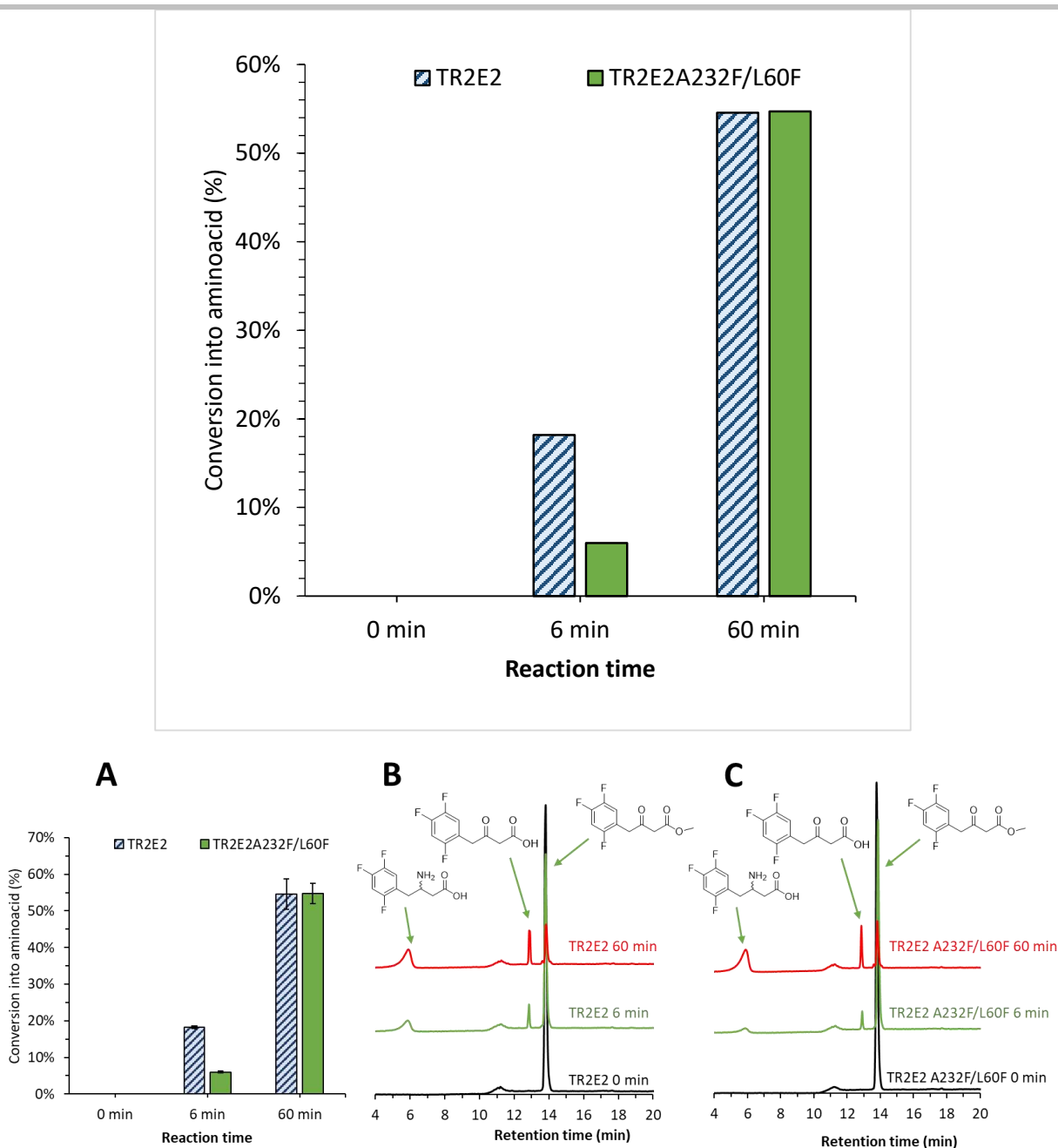
Figure S21. A) Conversion of 3-OtfBE to 3-(R)AtfBA by TR_2E_2 and $TR_2E_2A232F/L60F$ at short (6 min) and longer (60 min) incubation time. The following reaction conditions were used: [protein]: 31 µM; [3-OTfBE]: 14 mM; [PLP] : 1 mM ; [NPEA]: 14 mM; pH: 100 mM K_2HPO_4 buffer, pH 7.5; reaction volume: 100 µl; and T: 40 °C. Reactions were made in triplicates with standard deviation shown. The conversion was calculated as in Figure 6. HPLC chromatograms representing the one-pot reaction products by TR_2E_2 (B) and $TR_2E_2A232F/L60F$ (C) are shown. The figure was created using SigmaPlot 14.0 software.
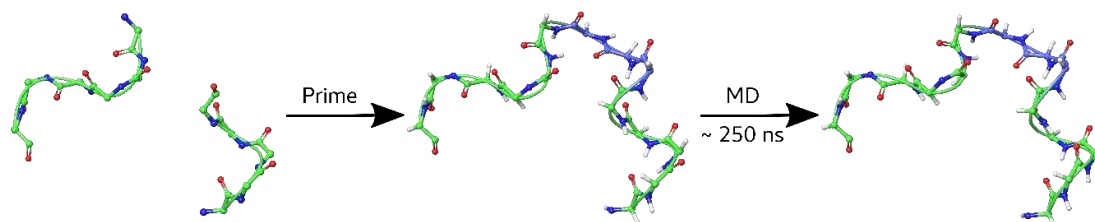
Figure S22. Modeling of the missing loop where the catalytic triad of the TR_2E_2 variant resides. First, the initial crystal structure was prepared using Protein Preparation Wizard,[23] followed by the modeling of loops using Prime.[24] Then, the modeled loops were further refined with 250 ns (4 replicas) of apo MD of the dimer structure to obtain the final structure used in the analysis. Both the secondary structure and the C atoms are shown in light green for the residues present in the crystal structure and in blue for the missing residues.

References

[1]   C. I. Giunta, I. Cea-Rama, S. Alonso, M. L. Briand, R. Bargiela, C. Coscolín, P. F. Corvini, M. Ferrer, J. Sanz-Aparicio, P. Shahgaldian, *ACS Nano* **2020**, *14*, 17652–17664.
[2]   G. Santiago, M. Martínez-Martínez, S. Alonso, R. Bargiela, C. Coscolín, P. N. Golyshin, V. Guallar, M. Ferrer, *Biochemistry* **2018**, *57*, 2245–2255.
[3]   C. Coscolín, N. Katzke, A. García-Moyano, J. Navarro-Fernández, D. Almendral, M. Martínez-Martínez, A. Bollinger, R. Bargiela, C. Gertler, T. N. Chernikova, D. Rojo, C. Barbas, H. Tran, O. V. Golyshina, R. Koch, M. M. Yakimov, G. E. K. Bjerga, P. N. Golyshin, K. -E. Jaeger, M. Ferrer, *Appl. Environ. Microbiol.* **2019**, *85*, e02404–18.
[4]   G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput. Aided Mol. Des.* **2013**, *27*, 221-234.
[5]   M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. Day, B. Honig, D. E. Shaw, R. A. Friesner, *Proteins* **2004**, *55*, 351-367.
[6]   J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, R. M. Levy, *J. Comput. Chem.* **2005**, *26*, 1752–1780.
[7]   A. D. Bochevarov, E. Harder, T. F. Hughes, J.R. Greenwood, D. A. Braden, D.M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang, R. A. Friesner, *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.
[8]   S. Alonso, G. Santiago, I. Cea-Rama, L. Fernandez-Lopez, C. Coscolín, J. Modregger, A. K. Ressmann, M. Martínez-Martínez, H. Marrero, R. Bargiela, M. Pita, J. L. Gonzalez-Alfonso, M. L. Briand, D. Rojo, C. Barbas, F. J. Plou, P. N. Golyshin, P. Shahgaldian, J. Sanz-Aparicio, V. Guallar, M. Ferrer, *Nat. Catal.* **2020**, *3*, 319–328.
[9]   S. Roda, G. Santiago, V. Guallar, *Adv. Protein Chem. Struct. Biol.* **2020**, *122*, 1–31.
[10]   S. Roda, A. Robles-Martín, R. Xiang, M. Kazemi, V. Guallar, *J. Phys. Chem. B.* **2021**, *125*, 6491–6500.
[11]   K. W. Borrelli, A. Vitalis, R. Alcantara, V. Guallar, *J. Chem. Theory Comput.* **2005**, *1*, 1304–1311.
[12]   E. Gallicchio, L. Y. Zhang, R. M. Levy, *J. Comput. Chem.* **2002**, *23*, 517–529.
[13]   D. Bashford, D. A. Case, *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
[14]   P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, V. S. Pande, *PLoS Comput Biol.* **2017**, *13*, e1005659.
[15]   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926-935.
[16]   V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712–725.
[17]   H.C. Andersen, *J. Chem. Phys.* **1980**, *72*, 2384–2393.
[18]   K.-H. Chow, D. M. Ferguson, *Comput. Phys. Commun.* **1995**, *91*, 283–289.
[19]   J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, B. O. Brandsdal, *Chem. Phys. Lett.* **2004**, *384*, 288–294.
[20]   L. Verlet, *Phys. Rev.* **1967**, *159*, 98.
[21]   L. Sumbalova, J. Stourac, T. Martinek, D. Bednar, J. Damborsky, *Nucleic Acids Res.* 2018, 46, W356–W362.
[22]   J. D. Hunter, *Comput. Sci. Eng.* **2007**, *9*, 90–95.
[23]   M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. Day, B. Honig, D. E. Shaw, R. A. Friesner, *Proteins* **2004**, *55*, 351-367.
[24]   G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput. Aided Mol. Des.* **2013**, *27*, 221-234.

Author Contributions

Computationally Driven Rational Design of Substrate Promiscuity on Serine Ester Hydrolases

Sergi Roda,[⊥] Laura Fernandez-Lopez,[⊥] Rubén Cañadas,[⊥] Gerard Santiago, Manuel Ferrer,* and Victor Guallar*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Enzymes with a broad substrate specificity are of great interest both at the basic and applied level. Understanding the main parameters that make an enzyme substrate ambiguous could be thus important not only for their selection from the ever-increasing amount of sequencing data but also for engineering a more substrate promiscuous variant. This issue, which remains unresolved, was herein investigated by targeting a serine ester hydrolase (EH102), which exhibits a narrow substrate spectrum, being only capable of hydrolyzing 16 out of 96 esters tested. By using a modeling approach, we demonstrated that one can rationalize active site parameters defining substrate promiscuity, and that based on them the substrate specificity can be significantly altered. This was accomplished by designing two variants, EH102$_{DM2}$ and EH102$_{TM2}$, that hydrolyze 51 and 63 esters, respectively, while maintaining similar or higher turnover rates compared to the original enzyme. We hypothesized that the parameters identified here (the volume, size, exposure, enclosure, hydrophobicity, and hydrophilicity of the active site cavity and its tightness) can serve in the future to expand the substrate spectra of esterases and thus expand their use in biotechnology and synthetic chemistry.

KEYWORDS: *enzymology, esterase, protein engineering, substrate promiscuity, computational chemistry*
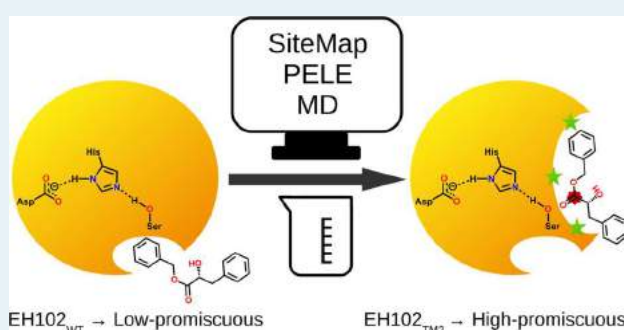
INTRODUCTION

The current needs of most biotechnology areas, both in the research and industrial sectors, demand the engineering of enzymes; new variants are continuously designed, leading to a wide variety of applications.[1] To this purpose, two main approaches have arisen.[2] The first one is rational design, which uses available knowledge to predict mutations.[3,4] On the other hand, directed evolution applies several random mutations to the system, and then selects those that enhance the desired property/ies.[4,5] Albeit both methods have benefited from important technological advances, including computational tools, the modification of the properties of a protein remains a challenging task.[6,7]

Substrate ambiguity, also called substrate promiscuity, the ability of a certain enzyme to catalyze a particular reaction for a wide range of substrates, is an appealing characteristic from the environmental and biotechnological points of view.[8] However, the properties that define whether an enzyme will have or not a broad substrate scope are not yet fully comprehended. Some investigations have addressed this issue,[9−12] but more precise information is still needed to engineer enzymes for increasing/decreasing their range of substrates with ease. In a recent article, by computing the substrate specificity datasets from more than one hundred diverse serine ester hydrolases, hereinafter referred to as esterases, when tested against a

wide range of esters,[13] we introduced a new descriptor of substrate promiscuity. It involves the solvent accessibility and the volume of the catalytic cavity, which could give an idea of the amount of substrates an enzyme can accept. Notice, however, that the active site volume alone does not produce any correlation with promiscuity ($r^2 \sim 0$), requiring additional descriptors for an efficient classification. In addition, the volume cannot be well defined in really exposed cavities, and this is why we consider the need for a deeper analysis, which should take into account the fact that enzymes with similar size cavities could accept significantly different number of substrates because of the physical/chemical properties of the amino acids conforming the active sites, and the role they may have in allowing substrate docking freedom.

Esterases (EC 3.1) are a subfamily of hydrolytic enzymes capable of breaking ester bonds (with the help of a water molecule), resulting in the alcohol and acid derived from the ester. Although this family of enzymes includes multiple

different protein folds and structures, substrate specificities, and biological functions, a substantial fraction of esterases shares the α/β fold.[14−16] Likewise, the vast majority of them have the archetypical Ser−His−Asp/Glu catalytic triad in the active site, which is a motif that enables the nucleophilic attack of the oxygen by the side chain of the serine on the electrophilic carbon of the ester bond. Regarding the industrial interest of esterases, several applications exist like their use for flavor development in food and beverages,[17] depolymerization of plastic polymers such as polyethylene terephthalate (PET) or polyhydroxyalkanoate (PHA),[18,19] production of fatty acids,[20] and more. In fact, together with lipases, most active toward insoluble esters, the market is projected to grow rapidly with new products and applications.[17,21] Thus, they have been commonly used in several applications of different sectors, which is expected to increase in the near future.[17,20,22]

When selecting an esterase to be of interest for pharmaceutical or industrial processes, substrate promiscuity becomes an important parameter, as an enzyme with broad substrate specificity opens the application range. Hence, understanding what makes an esterase promiscuous or not, and the further transformation of an esterase with a low substrate spectrum into one with a broad substrate spectrum would be a compelling accomplishment. Some studies have exemplified that influencing and also slightly expanding substrate specificity of enzymes is feasible by providing key substitutions in the proximity of the active and in the access tunnels.[23−31] However, in most cases, the specificity was established on the basis of a limited set of structurally similar substrates, which a priori limits to what extent the substrate specificity can be significantly altered. In this direction, there are no works, to the best of our knowledge, that demonstrate that significantly altering substrate specificity by rational design is feasible. In other words: to transform an enzyme with a low substrate spectrum with, a priori, low biotechnological potential, into a highly substrate promiscuous one with higher applied potential.

In this work, we present an in-depth analysis of the cavity and of the enzyme−substrate migration pathways of four serine ester hydrolases previously published,[13] two being highly substrate ambiguous (EH1 and CalB) and two being highly specific (EH88 and EH102). The comparative analysis allowed extracting the information that defines substrate promiscuity, from which we successfully approached the alteration of substrate specificity of the latter (EH102) by site-directed mutagenesis of residues near the active site. The enhancement of the substrate spectrum of an enzyme, through modifying residues near the active site, can compromise its activity, referring to the ability to increase the turnover rate of a certain reaction against a concrete substrate. Thus, we tried to find those mutations that gave us a tradeoff, increasing the range of substrates hydrolyzed without affecting the maximum specific activity of the native enzyme. We achieved our goals with two variants, capable of hydrolyzing a larger number of substrates while maintaining activity.

■ METHODOLOGY

Protein and Ester Preparation for In Silico Analysis. Three serine ester hydrolases, EH1, EH88, and EH102, isolated from the metagenomic DNA of microbial communities inhabiting the Lake Arreo, an evaporite karst lake in Spain,[13] and the commercial lipase CalB from *Pseudozyma aphidis* (formerly *Candida antarctica*), were used in the present

study. According to experimental information of substrate specificity, evaluated against a customized (diverse) library of 96 different esters,[13] EH1 and CalB could be considered as substrate promiscuous given their capacity to hydrolyze 72 and 68 out of the 96 esters. By contrast, EH102 and EH88 capable of hydrolyzing only 16 and 13 esters, were considered as low substrate ambiguous.[13] Crystal structures of EH1 (5JD4), Ca3lB (4K6G), and EH102 (5JD3) are available, and that of EH88 was modeled using homology modeling with Prime[32] (the template structure used was the PDB code: 1FXW, with 33% sequence identity and 93% coverage). For in silico analysis, proteins were prepared and protonated at pH 8.0 (the pH at which the experimental assays were performed) using Protein Preparation Wizard[33] and PROPKA,[34] including fixing side-chains and missing loops using Prime.[32] A later user's check was done by mainly inspecting whether the catalytic His residue was δ-protonated or not and the catalytic Asp residue was deprotonated or not, ensuring the proper hydrogen bond network of the catalytic triad. A final restrained minimization of root mean square deviation (*rmsd*) of 0.30 Å was carried out.

The ester compounds were modeled using the OPLS2005 force field,[35] except for the charges, which were calculated with Jaguar[36] using the density functional theory, with a B3LYP-D3 exchange−correlation functional, and the polarized triple-zeta (pVTZ) basis set. Finally, electrostatic potential (ESP) charges were fitted on the force field file to obtain the final atomic charges of the ligand.

Protein Energy Landscape Exploration Simulations. Protein energy landscape exploration (PELE) was used to study ligand migration and protein−ligand interactions. PELE is a Monte Carlo-based algorithm coupled with protein structure prediction methods.[37] The basic idea of this approach is to sample different microstates by initially applying small perturbations (translations and rotations) on the ligand. Also, the flexibility of the protein is taken into account by applying normal modes through the Anisotropic Network Model (ANM) approach or from a Principal Component Analysis (PCA). Once the system (protein and ligand) has been perturbed, side chains of the residues near the ligand are sampled with a library of rotamers to avoid steric clashes. Finally, a truncated Newton minimization with the OPLS2005 force field[35] is performed and the new microstate is accepted or rejected according to the Metropolis criterion, what we call a PELE step.

Ligand perturbation in all PELE simulations was constrained around a spherical box of 15 Å of radius around the active site. Moreover, rotations and translations were tuned smoothly as the ligand increased its contacts with the protein [the solvent-accessible surface area (SASA), of the substrate decreased] to enhance the exploration around the active site. Thus, the maximum translation allowed was 1.5 Å when the SASA was bigger than 0.15, otherwise the translation was restricted to 0.5 Å. Concerning the maximum rotation, it was 20° when the SASA was bigger than 0.15, otherwise it was reduced to 5°. The six lowest ANM eigenvectors were linearly combined at random to move the protein. The side chain phase included all residues within 6 Å of the ligand. The variable dielectric generalized-born nonpolar (VDGBNP) implicit solvent[38] was applied to mimic the influence of waters around the protein. The PELE simulations were run in the MareNostrum IV cluster from the Barcelona Supercomputing Center (BSC) with 64 cores and each core performed 1000 PELE steps.

The main variables studied in these simulations were the enzyme–substrate interaction energies, the SASA of the ligand, or distances between the oxygen of the catalytic Ser and the electrophilic carbon of the ester, referred to as the "serine-substrate distance".

Molecular Dynamics. Four replicas of 500 ns of molecular dynamics (MD) simulations with OPENMM[39] were performed to analyze the flexibility of the protein and particularly of the catalytic triad of selected systems. A water cubic box (a distance of 8 Å between the closest protein atom and the edge of the box) was created around the system using the TIP3P water model[40] and the charge of the system was stabilized using monovalent ions (Na$^+$ and Cl$^-$). The protein system was parameterized with the AMBER99SB force field.[41] The Andersen thermostat[42] and the Monte Carlo barostat[43,44] were applied for the NPT ensemble (constant pressure and temperature, being 1 bar and 300 K, respectively). The *NVT* equilibration lasted for 400 ps and a constraint of 10 kcal/(mol·Å^2) was applied to the system, while the NPT equilibration lasted for 1 ns and a milder constraint of 5 kcal/(mol·Å^2) was used. The Verlet integrator[45] with a 2 fs time step was used, using constraints between H and heavy atoms. For the nonbonded long–range interactions, a radius of 8 Å was used.

Molecular Docking Calculations. Prior to the PELE simulations, esters were docked at the active site of the studied enzymes using Glide.[46] First, the grid of each protein was generated with the center being located at the center of masses of the residues defining the catalytic triad, and the inner box was limited to a cube with an edge of 10 Å. The ligand was sampled as flexible and standard precision was used. 10 poses were extracted and all of them minimized after the molecular docking with the OPLS2005 force field.[35] All docking results were visually inspected and those with better catalytic positions, typically the top-ranked Glide score, were used to perform the PELE simulation.

Active Site Cavity Analysis. To infer those properties that defined promiscuity, SiteMap[47,48] was used. It was also used to infer the properties of the active site cavity in the different mutants. This software enables finding binding sites in a protein surface and ranks them according to several chemical and physical properties. These properties include: volume, size, exposure, enclosure, contact (tightness), hydrophobicity, and hydrophilicity. The volume is calculated by first identifying all points on the cubic mapping grid that lie within 4 Å of any site point and are outside the protein surface, then the volume is computed from the number of remaining volume points and the grid-box volume, which is $(0.7 \text{ Å})^3$ in the default case. The number of site points, where typically 2−3 site points correspond to each atom of the bound ligand, including hydrogens, is equivalent to the size of the cavity. The exposure is calculated by using the ratio of the number of extension points to the number of original points plus extension points. "Extension" site points are points that must lie within a given distance in *x*, *y*, or *z* from an original site point (by default 3 Å), and must make good contact with the receptor or lie at least 4 Å from the nearest protein atom. Enclosure is the fraction of radial rays drawn from the site points that strike the receptor surface within a distance of 10 Å over the original and "extension" site points calculated in the exposure property. Contact is computed by averaging the ligand–receptor vdW interaction energies (with nominal vdW parameters) over the original and "extension" site points. Hydrophobicity and hydrophilicity are computed by averaging the Grid$_{phobic}$ (Grid$_{phobic}$ = E_{vdW} − 0.3·$E_{oriented-dipole}$) or Grid$_{philic}$ (Grid$_{philic}$ = E_{vdW} + $E_{oriented-dipole}$) potential over the original and "extension" site points. The ligand used to probe the active site region (which included a box around the ligand of 6 Å) was ethyl-3-oxohexanoate and it was docked prior to SiteMap with the docking protocol explained in the Methodology.

Prediction of ΔΔG in the EH102 Variants. ΔΔG$_{(mut-WT)}$ of stability in the experimentally tested variants was calculated using the module of thermodynamic stability from HotSpot Wizard, which uses FoldX to repair possible problems in the protein structure and Rosetta to perform the energy minimization and ΔΔG calculation (according to protocol 3 from Rosetta).[49]

Chemicals, Oligonucleotides, Source of Enzyme, and Strains. The sources of all chemicals (of the purest grade available), oligonucleotides for DNA amplification and serine ester hydrolases EH102 (available in the expression vector pET46 Ek/LIC plasmid in *Escherichia coli* BL21 as a host) used in the present study, were as reported in ref 13.

Site-Directed Mutagenesis in EH102. To obtain EH102 variants containing mutations, the pET46 Ek/LIC plasmid containing EH102 DNA insert was used.[13] Mutagenic PCR was developed using the QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies, Cheadle, UK) and conditions described previously.[50] The following mutations were introduced, individually or in combination: Ile16Val, Ile16Ala, Ile16Gly, Ile92Gly, Ile92Ala, and Trp96Gly. We produced mutants where single (Ile16Val, Ile16Ala, Ile16Gly) or multiple (Ile92Gly Trp96Gly, Ile92Ala Trp96Gly, Ile92Gly Trp96Gly Ile16Val, Ile92Gly Trp96Gly Ile16Gly, Ile92Ala Trp 96Gly Ile16Val, Ile92Ala Trp96Gly Ile16Ala, Ile92Ala Trp96Gly Ile16Gly, Ile92Gly Trp96Gly Ile16Ala) mutations were introduced. In all cases, the forward primers used to generate the EH102 variants are as follows: Ile16ValFwd: ATC ATC GGC GAC TCG gTC ACG GAC GCG GGA C; Ile16AlaFwd: ATC ATC GGC GAC TCG gcC ACG GAC GCG GGA CG; Ile16GlyFwd: ATC ATC GGC GAC TCG ggC ACG GAC GCG GGA CG; Ile92GlyTrp96-GlyFwd: GCG ATG ATG ATC GGC ggC AAC GAC GTC gGG CGC CAG TTC GAC CTG; Ile92AlaTrp96GlyFwd: GCG ATG ATG ATC GGC gcC AAC GAC GTC gGG CGC CAG TTC GAC CTG.

Protein Production and Purification. Seven native serine ester hydrolases, EH1 (protein data bank acc. nr. 5JD4), EH3 (GenBank acc. nr. KY483645), EH5 (GenBank acc. nr. KR107271), EH7 (GenBank acc. nr. KY483644), EH12 (GenBank acc. nr. KR107263), EH37 (GenBank acc. nr. KR107248), and EH102 (protein data bank acc. nr. 5JD3) from metagenomic origin, and four mutants derived from EH102 (EH102$_{TM1}$, EH102$_{SM1}$, EH102$_{DM2}$, and EH102$_{TM2}$), were used to perform substrate fingerprint and kinetic determination (k_{cat} and K_M). The vector pET46 Ek/LIC and the host *E. coli* MC1061 were the source of the His6-tag EH1, EH5, EH12, EH17, EH37, EH102, EH102$_{TM1}$, EH102$_{SM1}$, EH102$_{DM2}$, and EH102$_{TM2}$, and the vector pBXNH3 and the host *E. coli* MC1061 were the source of the His6-tag EH3. For enzyme production, a single colony, previously grown at 37 °C on solid Luria Bertani (LB) agar medium supplemented with 100 μg mL^{-1} ampicillin (Amp), was picked and used to inoculate 50 mL of LB-Amp medium in a 0.25 L flask, followed by cultivation at 37 °C and 200 rpm overnight. Afterward, 50 mL of this culture was used to inoculate 1 L of LB-Amp
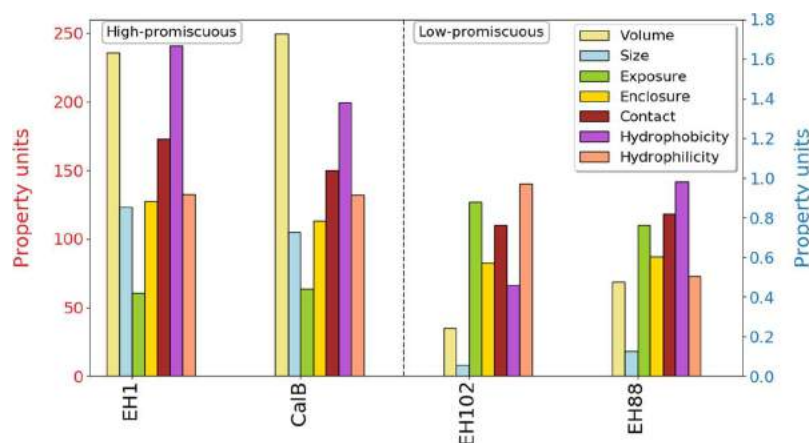
Figure 1. Bar plot representing the volume, size, exposure, enclosure, contact, hydrophobicity, and hydrophilicity properties of the active site cavity of two high substrate promiscuous and two low substrate promiscuous esterases with SiteMap.[47,48] The *Y* red axis on the left represents the scale for the volume (in Å^3) and the size properties of the active site cavity, while the *Y* blue axis on the right represents the remaining ones. The figure was created with the Matplotlib library.[51]

medium in a 2.5 L flask, which was then incubated at 37 °C to an OD$_{600nm}$ of approximately 0.8 (ranging from 0.7 to 0.9). Protein expression was induced by adding IPTG to a final concentration of approx. 1%, followed by incubation for 16 h at 16 °C and 220 rpm. The cells were harvested by centrifugation at 8000*g* for 15 min to yield a pellet of 2−3 g (wet weight). The wet cell pellet was frozen at −86 °C overnight, thawed, and resuspended in 15 mL of 50 mM sodium phosphate, pH 8.0, 10 mM imidazole, and 300 mM NaCl. Lysonase Bioprocessing Reagent (Novagen, Darmstadt, Germany) was added (4 μL g^{-1} wet cells) and incubated for 1 h on ice with rotating mixing. The cell suspension was sonicated for a total of 5 min and centrifuged at 15,000*g* for 15 min at 4 °C, and the supernatant was retained. The soluble His-tagged protein was purified at 4 °C after binding to a Ni-NTA His-Bind resin (Sigma-Aldrich, MO, US) and eluting with 50 mM sodium phosphate, pH 8.0, 250 mM imidazole, and 300 mM NaCl. Eluted protein was subjected to ultrafiltration through low adsorption hydrophilic 10,000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) to concentrate the protein solution. An extensive dialysis of protein solutions against 40 mM (4-(2-hydroxyethyl)-1-piperazinee-thanesulfonic acid) (HEPES) buffer (pH 7.0) was then performed using the Pur-A-LyzerTM Maxi 1200 dialysis kit (Sigma-Aldrich, MO, US) as follows: 2 mL concentrated protein solution was dialyzed against the 2 L buffer for 1 h at room temperature, after which the buffer was replaced with another 2 L buffer and maintained for 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. Purity was assessed as >98% using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis in a Mini PROTEAN electrophoresis system (Bio-Rad, Madrid, Spain). On average, a total of about 10−20 mg total purified recombinant proteins were obtained from 1 L culture.

Substrate Fingerprint and Kinetic Parameter Determination. Hydrolytic activity was assayed using a pH indicator assay at 550 nm using 96 structurally diverse esters in 384-well plates as previously described,[13] with slight modifications. Briefly, to 20 μL of 5 mM 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid (EPPS) buffer (pH 8.0), 2 μL of a stock ester solution was added to achieve the desired

concentration of each ester. Then, 20 μL of 5 mM EPPS buffer pH 8.0 containing 0.95 mM Phenol Red was added. Buffer was dispensed with a QFill3 Microplate Filler (Genetix, CA, USA) and the buffers with a PRIMADIAG Demo liquid handling robot (EYOWN TECHNOLOGIES S.L., Madrid, Spain). Finally, 2 μL of the stock protein solution (from stock solutions at different concentrations, in 40 mM HEPES buffer pH 7.0) was immediately added to each well, to achieve the desired protein concentration, using an Eppendorf Repeater M4 pipet (Eppendorf, Hamburg, Germany). The total reaction volume was 44 μL. Ester hydrolysis was measured at 30 °C using a Synergy HT Multi-Mode Microplate Reader (Izasa Scientific, Madrid, Spain) in continuous mode at 550 nm over 24 h, although initial rates were only considered for calculations. One unit (U) of enzyme activity was defined as the amount of free enzyme or enzyme bound to the carrier required to transform 1 μmol of substrate in 1 min under the assay conditions using the reported extinction coefficient (Phenol red at 550 nm = 8450 M^{-1} cm^{-1}). All values, in triplicates, were corrected for nonenzymatic transformation; the absence of activity was defined as at least a two-fold background signal. For K_M determination, [protein]: 4.5 μg mL^{-1}; [ester]: 0−100 mM; reaction volume: 44 μL; T: 30 °C; and pH: 8.0. For k_{cat} determination, [protein]: 0−270 μg mL^{-1}; [ester]: 50 mM; reaction volume: 44 μL; T: 30 °C; and pH: 8.0.

■ RESULTS AND DISCUSSION

Computational Study of Substrate Promiscuity. In order to extract important features describing the substrate promiscuity of the four different esterases selected as targets, their active sites were studied using the SiteMap software.[47,48] Seven properties were analyzed: volume, size, exposure, enclosure, contact (tightness), hydrophobicity, and hydrophilicity. Size, which accounts for the number of cavity points in SiteMap's procedure, is a similar measure to the one corresponding to the volume. Exposure and enclosure properties provide different measures of how opened is the site to the solvent. Low exposure/high enclosure values mean that the cavity is not very solvent-exposed. The contact feature describes the degree of tightness of the cavity. Therefore, the higher this value the more compact the cavity will be.
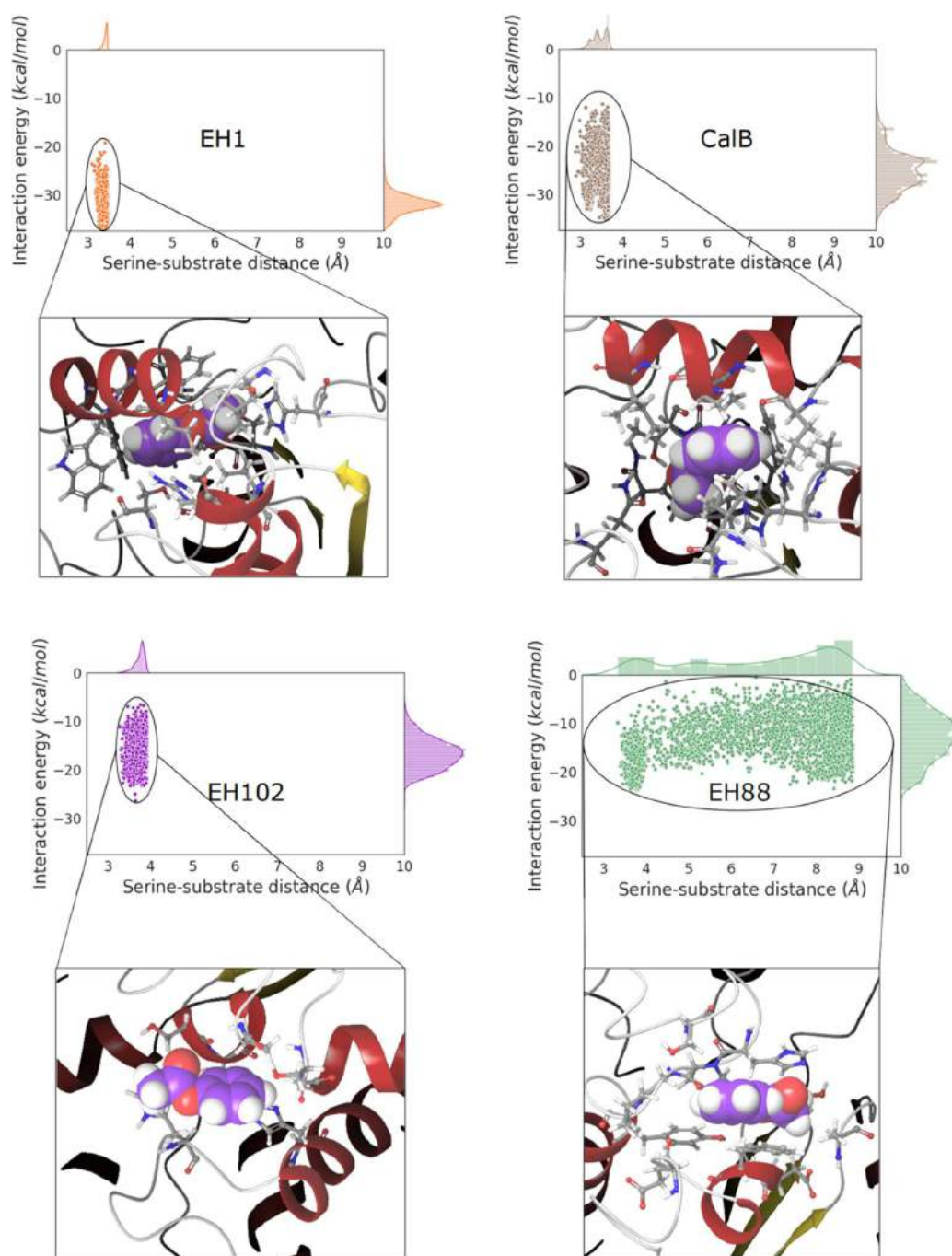
Figure 2. Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulations for the different studied esterases with phenyl acetate as the substrate (3226 data points for CalB, 3095 for EH1, 2647 for EH102, and 2734 for EH88), accompanied by one representing binding pose obtained during the simulation. The colors in the density plots represent each esterase: CalB (brown), EH1 (orange), EH102 (purple), and EH88 (green). The density plots were created with the Matplotlib library.[51] In the shown binding poses, the ligand has the C atoms stained in lilac and the overall structure represented in the CPK model, followed by the ball-and-stick representation of the residues 4 Å far from it. Protein ribbon is colored according to the secondary structure (ruby: α-helix, golden: β-sheet, and gray: loops).

As it can be seen in Figure 1, some properties are well correlated with the degree of substrate promiscuity of the enzyme. In the two high-promiscuous esterases, we can observe that the volume of the cavity is 235.64 Å³ for EH1 and 249.70 Å³ for CalB, while the low-promiscuous esterases have volumes of 68.94 Å³ for EH88 and 34.99 Å³ for EH102. Thus, the esterases with a broad substrate range have active

site cavities 3- to 7-fold higher than those with narrow substrate spectra. We see a similar trend regarding the size property, meaning that the active site has to be big enough to accommodate a wide variety of substrates. Regarding enclosure, hydrophobicity, and contact properties, they also are higher in esterases with the highest substrate ambiguity, whereas the exposure is significantly lower. These results

clearly give us a qualitative idea of the shape and the chemical properties (type of residues) of the active sites. Esterases accepting a higher number of esters have large enough, well defined, hydrophobic, and compact cavities, sheltered from the solvent, whereas low substrate promiscuous esterases possess active sites not so large nor well-defined, less hydrophobic, and significantly more exposed to the solvent.

So far, we have gained insights into the properties of the cavities. Now, we will turn to simulate substrate–enzyme interactions to see how much they correlate with their substrate range, with the goal to infer which residues could be mutated to enhance the substrate range, and thus, substrate promiscuity in an esterase only hydrolyzing a few esters.

PELE simulations were carried out with two esters; phenyl acetate (128.57 Å^3), which is catalyzed by the four esterases and it is an ester commonly hydrolyzed by most such enzymes,[13] and ethyl-3-oxohexanoate (159.92 Å^3), which is only catalyzed by the highly promiscuous ones.[13] Still, the activity against phenyl acetate is much higher in the EH1 and CalB because of the better overall properties of the active site and cavity. Looking at Figure 2 for the phenyl acetate compound, it can be seen that in EH1 and CalB, which are much more promiscuous and active than EH102 and EH88, the catalytic distances and the interaction energies are lower, which means that the ligand is more stable and in a closer interaction for the covalent addition. Besides, Figure S1 shows that SASA values are smaller in EH1 and CalB esterases, meaning that the substrate is more buried in the cavity of those esterases in comparison with the low-promiscuous ones. Concerning the simulations with ethyl-3-oxohexanoate, we observe similar results as it can be seen in Figures S2 and S3.

Therefore, our results indicate that in order to turn an esterase into a more promiscuous one, the cavity has to be enlarged but protected from large solvent exposure. Importantly, besides a quick calculation of the (change in) cavity properties, enzyme–substrate simulations should provide a clear indication of the change in substrate activity.

Rational Design of a Low-Promiscuous Esterase to Increase Its Substrate Range. We select as target the esterase EH102 because the crystal structure is available (5JD3), and that, despite its cavity size, is capable of efficiently hydrolyzing substrates such as glucose pentaacetate. Regarding the surroundings of its catalytic triad, many bulky residues can be found, hiding a possible cavity where the substrate could bind; the overall active site presents an excess of solvent-exposure. Based on a preliminary visual inspection, we focused on the Ile16, Ile92, and Trp96 residues. Taking into account the information extracted from the computational studies of the previous section, we first attempted a drastic cavity enlargement by combining three mutations, I16G/I92A/W96G, where the I92A mutation was performed to keep some hydrophobicity. Hereinafter, this triple mutant was designated as TM1 (Figure 3; see Table 1 to check the nomenclature).

To understand in depth the properties of this mutant, we first compute the seven properties used to analyze promiscuity (Figure 1). As expected, the volume of the new cavity, its size, its tightness, its enclosure, and hydrophobicity have increased in comparison with the wild type (WT) solvent-exposed active site, whereas the exposure has decreased (Figure 4).

Molecular sampling using PELE was also performed for two ester substrates that were not hydrolyzed by EH102$_{WT}$: the previously used ethyl-3-oxohexanoate and the small (112.31
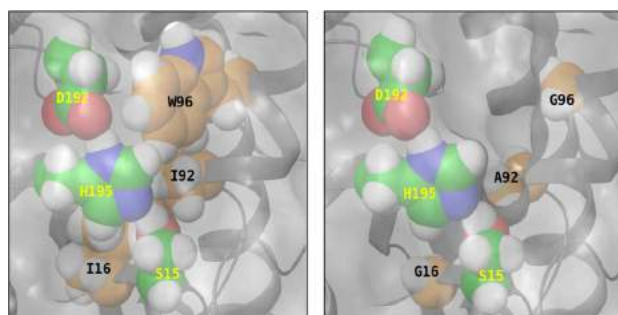
Figure 3. 3D representation of the surface of EH102$_{WT}$ (left) and EH102$_{TM1}$ (right). The mutated residues have the C atoms stained in orange and their labels in black, while the residues of the catalytic triad have them stained in green and the labels in yellow. All shown residues are displayed with the CPK model.

Å^3) vinyl crotonate.[13] As it can be seen in Figures 5 and S4, EH102$_{TM1}$ shows better accommodation of substrates with smaller serine-substrate distances and interaction energies. Moreover, SASA also decreased, confirming that the engineered cavity has the ligand more buried in the protein.

Current state of the art enzyme engineering efforts combining in silico and in vitro techniques are mostly based on early experimental validation. As in other molecular engineering fields, such as drug discovery, the most efficient route is to find an early (but weak) activity, followed by additional rounds of refinement designs. Moreover, our goal when introducing high-performance computing and sophisticated simulation techniques, such as PELE, is to drastically eliminate experimental efforts in such "lead optimization" processes; we aim at finding significantly enhancing variants within ∼10 proposed mutants. Within this in mind, we proceed to experimentally test our initial in silico validated mutant. Experimental tests, consisting in quantifying k_{cat} and K_M against the set of 96 structurally diverse esters (Figure 6), corroborated that the substrate range was enhanced, going from hydrolyzing only 16 esters in EH102$_{WT}$ to 35 esters in EH102$_{TM1}$, with glucose pentaacetate (k_{cat} of ca. 152.124 min^{-1}) and phenyl propionate (k_{cat} of ca. 9.966 min^{-1}) serving as the best substrates, respectively (Figure 6 and Table S1). Nonetheless, the average k_{cat} of the variant for all converted substrates decreased considerably by 70-fold (maximum: 450-fold for glyceryl triacetate; minimum: 5-fold for phenyl propionate), although no major effect in substrate affinity was observed (average K_M fold change of ca. 1.2). We have shown in a recent study, involving the design of artificial active sites with esterase activity, that the lack of a tight catalytic triad, with short hydrogen bond distances, may be mainly responsible for low activity.[50,52] Thus, the substitution of one of the big hydrophobic residues for smaller ones could have introduced more flexibility to the catalytic triad, disrupting the catalytic mechanism.

From the three different mutations, I16G is right beside the catalytic Ser residue (S15) and could be directly affecting the hydrogen bond between the catalytic His and Ser residues. To confirm this hypothesis, we ran classical MD simulations for EH102$_{WT}$ and EH102$_{TM1}$ throughout 500 ns with 4 replicas each. Figure 7 illustrates the catalytic serine–histidine distance where we clearly observe a significant increase for EH102$_{TM1}$ in comparison with the EH102$_{WT}$ enzyme. Likewise, the aspartic-histidine distance and the local *rmsd* of the residues in the

Table 1. Recompilation of all the Experimental EH102 Variants That Have Been Designed and Assayed and Their Nomenclature along the Document

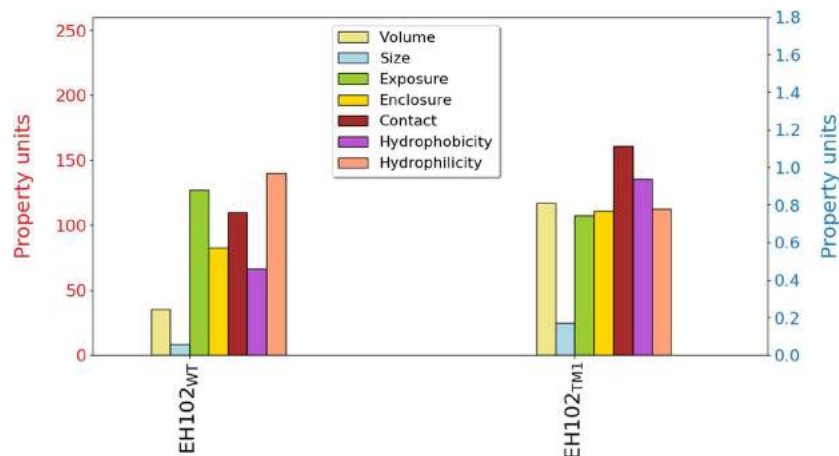| WT	TM1	TM2	TM3	TM4	TM5	TM6	DM1	DM2	SM1	SM2	SM3
Ile16	Gly16	Val16	Val16	Ala16	Ala16	Gly16			Val16	Ala16	Gly16
Ile92	Ala92	Ala92	Gly92	Gly92	Ala92	Gly92	Gly92	Ala92			
Trp96	Gly96	Gly96	Gly96	Gly96	Gly96	Gly96	Gly96	Gly96			

Figure 4. Bar plot representing the volume, size, exposure, enclosure, contact, hydrophobicity, and hydrophilicity properties of the active site cavity of the EH102$_{WT}$ in comparison with EH102$_{TM1}$ obtained with SiteMap.[47,48] The Y red axis on the left represents the scale for the volume (in Å^3) and the size properties of the active site cavity, while the Y blue axis on the right represents the remaining ones. The figure was created with the Matplotlib library.[51]
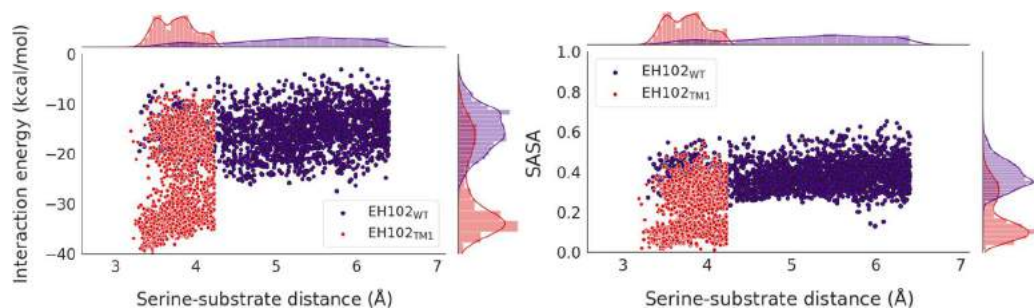
Figure 5. Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy (left) and SASA (right) from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulation for EH102$_{WT}$ and EH102$_{TM1}$ with ethyl-3-oxohexanoate as the substrate (2332 data points for EH102$_{WT}$ and 2225 for EH102$_{TM1}$). The density plots were created with the Matplotlib library.[51]

catalytic triad have significant larger values as well in the proposed variant (Figures S5 and S6). However, the *rmsd* of the protein backbone converges to similar values (Figure S7), indicating that the mutations do not destabilize considerably. Overall, these results clearly point to significant larger fluctuations of the catalytic triad in the variant, most likely as a result of side chain flexibility, compromising the hydrolytic activity of the enzyme.

Rational Design to Balance the Increase in Substrate Promiscuity with the Conversion Rate: from Prediction to Experimental Validation. From all the information gathered from the EH102$_{TM1}$ variant, we proceed to the second mutant-refinement stage. As a reminder, our goal is not to design an extensive library of mutants but to narrow down the list of mutants to only a few, by means of using molecular modeling. Single mutants at the position 16, the closest one to the catalytic serine, were assayed to see the progressive effect of substituting for a smaller and less-hydrophobic residue in

activity and promiscuity. Moreover, double mutants were performed at the two other residues (92 and 96) in order to reduce the increase of flexibility of the catalytic serine and to check whether substrate promiscuity can be increased without changing residue 16. Finally, we tried six TM variants that were created from the permutations of mutating I92 to A92 or G92, I16 to (not so small residues) V16, A16, or G16, and W96 to G96; W96G mutation was always kept, as it is the most solvent-exposed residue and the biggest one, significantly increasing the cavity (and its access) once it is mutated. On the other hand, I92 was substituted by both A and G because it is a more buried residue. Table 1 lists all variants studied.

We quickly checked the cavity properties of these variants with SiteMap,[47,48] seeing similar results for the triple mutants and double mutants to those of the first TM1 mutant (Figure S8); smaller changes with the single mutants were observed, as expected. Thus, changes in the cavity are also predicted in the double mutant and triple mutant variants, where we would
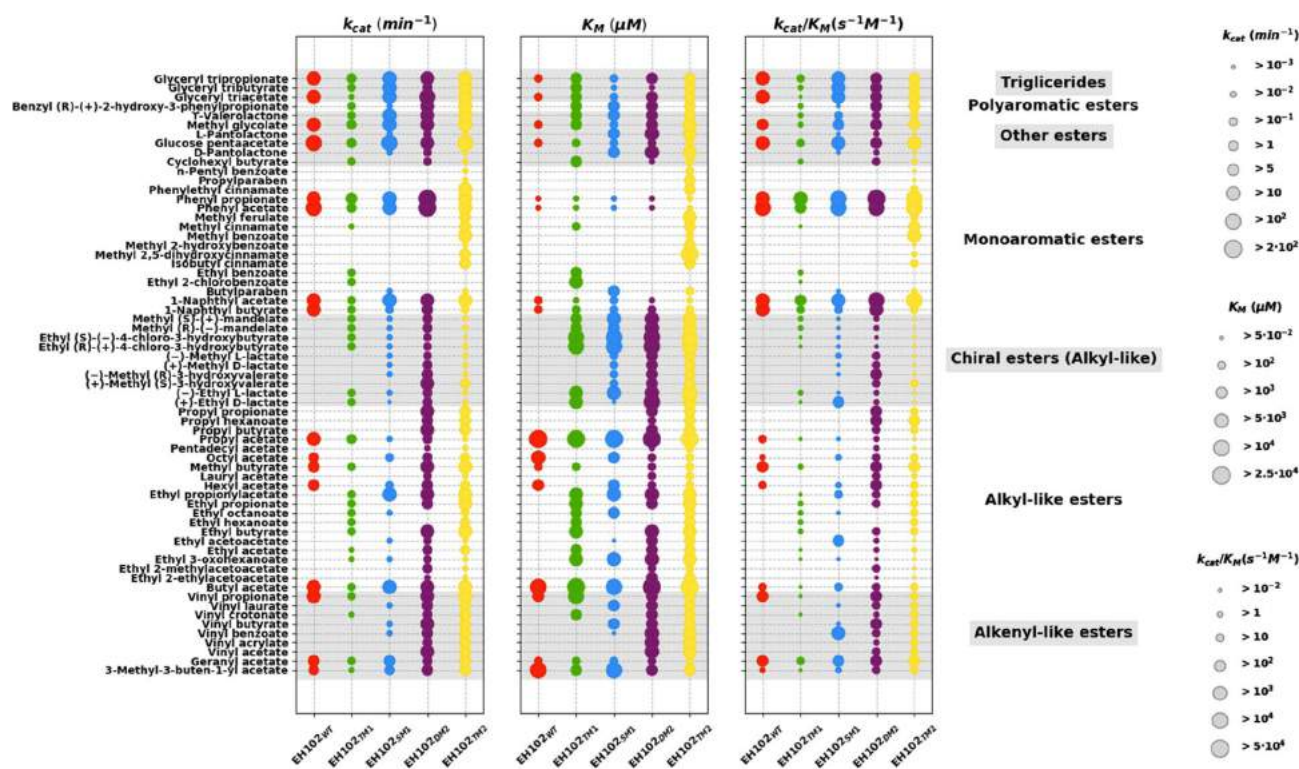
Figure 6. Substrate spectra of the EH102$_{WT}$, EH102$_{TM1}$, EH102$_{SM1}$, EH102$_{DM2}$, and EH102$_{TM2}$. The identification code of each variant can be found at the X axis, while the ester names are placed at the Y axis. The size of the points indicates a higher value for the represented property of each ester, which was determined as described in the Methodology. Substrates are grouped in the different ester families in which they can be classified to. The figure was created with the Matplotlib library.[51] For raw data, see Table S1.
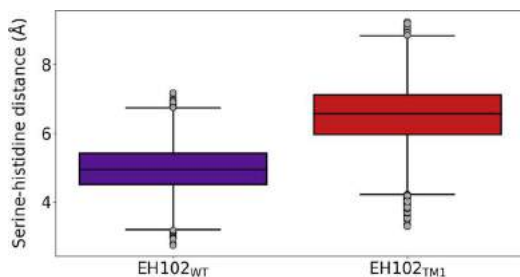

Figure 7. Box plot representing the serine−histidine distance ($d_{H_\gamma^{Ser}-N_\varepsilon^{His}}$) along the 500 ns of the 4 MD replicas performed for EH102$_{WT}$ and EH102$_{TM1}$. The figure was created with the Matplotlib library.[51]

expect an increase in substrate promiscuity. This time, however, we used MD simulations to probe the stability of the catalytic triad. EH102$_{TM2}$ (containing V16/A92/G96) and EH102$_{TM6}$ (G16/G92/G96) variants improved the catalytic distances and the overall geometry of the catalytic triad (Figures S9−S11). Regarding those variants that maintained similar values with the WT enzyme, we find the EH102$_{TM3}$ (V16/G92/G96), EH102$_{TM4}$ (A16/G92/G96), EH102$_{DM1}$ (G92/G96), and EH102$_{DM2}$ (A92/G96) variants. The remaining variants showed disturbed catalytic triads according to the studied metrics.

In addition, we estimated ΔΔGs of stability in the different variants (Figure S12). Although both EH102$_{TM2}$ and EH102$_{TM6}$ variants seemed the most promising ones, the large destabilization of the EH102$_{TM6}$ variant questions its

integrity. Thus, out of the triple mutants, EH102$_{TM2}$ seems the best candidate for properly folding and increasing the promiscuity of the WT enzyme without compromising its overall catalytic activity. Moreover, combining the results of single mutation stability analysis with the MD results (Figures S9−S11) suggests that residue 16 should be either substituted with a valine residue (EH102$_{SM1}$) or preserved because of its role in maintaining the catalytic integrity; to prove this, EH102$_{SM2}$ (A16) and EH102$_{SM3}$ (G16) were produced. Similarly, comparing the EH102$_{DM2}$ and EH102$_{DM1}$ variants seems to indicate that residue 92 must have some kind of side chain (larger than a Gly hydrogen).

We proceeded to express all the 10 additional mutants (for summary, see Table 1). Only three (EH102$_{SM1}$, EH102$_{DM2}$, and EH102$_{TM2}$) were produced in soluble active forms, and their substrate spectra and k_{cat} and K_M were determined, and compared with those of EH102$_{WT}$ and EH102$_{TM1}$, when tested against the 96 structurally diverse esters.

We first observed that, surprisingly, the mutant containing the single mutation I16V (EH102$_{SM1}$) was capable of hydrolyzing as many esters (36 in total) as EH102$_{TM1}$ (35 in total) (Figure 6, Table S1). However, a closer inspection of substrates being converted reveals that 24 esters were common substrates, but that compared to EH102$_{WT}$, EH102$_{SM1}$ gained the capacity to hydrolyze 12 substrates that EH102$_{TM1}$ could not; they mostly include esters with a volume lower than 200 Å^3, but one higher than 200 Å^3, which was vinyl laurate (k_{cat} 0.028 min^{-1}). In contrast, EH102$_{TM1}$ gained the capacity to convert 11 esters not hydrolyzed by the EH102$_{SM1}$, all of them being molecules with a volume below 200 Å^3. In addition, differences were also observed at the level of catalytic turnover

Figure 8. Comparison of the substrate promiscuity of EH102 variants against the published dataset of esterases.[13] The plot shows the number of esters hydrolyzed by EH102$_{WT}$, EH102$_{TM1}$, EH102$_{SM1}$, EH102$_{DM2}$, and EH102$_{TM2}$ in comparison with the other 146 ester hydrolases. The figure was created with the Matplotlib library.[51]
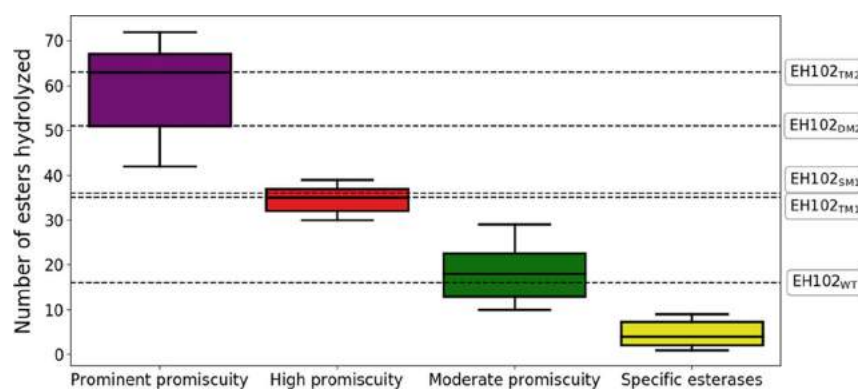
and substrate affinities, which range from *ca.* 9.966 to 0.032 min^{-1} and from *ca.* 0.043 to 51.612 mM, respectively for EH102$_{TM1}$ and from 122.732 to 0.002 min^{-1} from 8.2×10^{-5} to 30.002 mM, respectively for EH102$_{SM1}$. Interestingly, out of the 24 esters hydrolyzed by both mutants, we noticed that 14 were preferably converted by EH102$_{SM1}$ (from 2.5- to 400-fold in terms of k_{cat}), 9 by EH102$_{TM1}$ (from 7- to 95-fold), and only 1 was equally converted; interestingly, most esters above 200 Å^3 were better (2.5- to 32-fold) converted by EH102$_{SM1}$ and also had higher substrate affinities (average for these substrates: 0.934 mM for EH102$_{TM1}$ and 0.575 mM for EH102$_{SM1}$). We further observed that the overall catalytic turnover of EH102$_{SM1}$ (up to 122.7 min^{-1}) approaches that of the EH102$_{WT}$ (up to *ca.* 152.124 min^{-1}), although still the mutation reduced the rate of hydrolysis for most of the substrates (from 1.3- to 1258-fold in terms of k_{cat} for 9), while slightly increasing k_{cat} for a few substrates (from 1.3- to 2.5-fold for three esters), without altering substrate affinities. Thus, the added flexibility in the active site by the I16V mutation could enhance the catalytic binding of large but also relatively small substrates, but the active site environment still compromises the overall catalytic rate.

We further observed that the mutations I92A/W96G (in EH102$_{DM2}$) significantly promoted the substrate spectrum to 51 esters (Figure 6, Table S1). Its catalytic turnover (up to *ca.* 216.103 min^{-1}) was significantly higher than that observed for EH102$_{SM1}$ and EH102$_{TM1}$, and comparable or even higher (for 8 esters) to that of the EH102$_{WT}$. This confirms that the preservation of the hydrophobicity in the active site provided by Ile16 is important to maintain the geometry of the catalytic triad, and of the conversion rate. Compared to the EH102$_{WT}$ enzyme, EH102$_{DM2}$ gained activity toward 11 esters that were not hydrolyzed neither by EH102$_{SM1}$ nor EH102$_{TM1}$. They all represent molecules with a volume lower than 200 Å^3 but pentadecyl acetate (309 Å^3), and mostly include alkyl-(propyl propionate, propyl butyrate, propyl hexanoate, octyl acetate, pentadecyl acetate, ethyl 2-ethylacetoacetate, ethyl 2-methyl-acetoacetate, and (+)-methyl (S)-3-hydroxyvalerate) and alkenyl-(vinyl acetate and vinyl acrylate) like esters of different sizes, and they were converted at rates up to 31.268 min^{-1} (K_M values from *ca.* 0.084 to 8.538 mM).

Finally, the mutations I16V/I92A/W96G (in EH102$_{TM2}$) were found to promote a further step in the substrate spectrum. Indeed, the EH102$_{TM2}$ variant was capable of hydrolyzing 63 esters (Figure 6, Table S1), which represent 47,

28, 27 and 12 more esters than EH102$_{WT}$, EH102$_{TM1}$, EH102$_{SM1}$, and EH102$_{DM2}$, in the same order. In addition, the three mutations gained activity for eight esters, which could not be hydrolyzed by any of the previous variants nor the WT, namely, methyl benzoate, phenylethyl cinnamate, methyl ferulate, methyl 2,5-dihydroxycinnamate, isobutyl cinnamate, propylparaben, methyl 2-hydroxybenzoate, and *n*-pentyl benzoate; they all represent aromatic esters, which were converted at rates as high as 11.625 min^{-1} (K_M: from *ca.* 0.175 to 47.283 mM). The extended level of substrate docking provided by the three mutations, by meaning of the higher number of esters being accepted, does not entail, however, an overall increase in catalytic turnover (k_{cat} up to *ca.* 119.348 min^{-1}) compared to EH102$_{WT}$ and the mutants EH102$_{SM1}$ and EH102$_{DM2}$. The fact that this variant contains a mutation at I16, which was found to be the key for the optimal geometry of the active site and the conversion rate, may agree with the fact that the activity of EH102$_{TM2}$ was not further promoted. At the same time, it further demonstrates that this residue is important for favoring substrate docking freedom and thus expanding the substrate spectrum.

The above results first demonstrate that it is feasible to significantly expand the substrate spectrum of a hydrolase with an initial narrow substrate spectrum, by introducing mutations in a set of residues at the proximity of the active site. Second, although this can occur at the expense of activity, we further demonstrated that by applying rational design a balance between substrate promiscuity and the conversion rate can be achieved. Importantly, this was accomplished with a rather small library of mutants; all mutants assayed have been exposed here! Such success was possible, in our opinion, by efficiently combining molecular modeling and experimental validation, the latter being introduced very early in the design process. Using this approach and starting from a hydrolase capable of hydrolyzing only 16 esters out of 96 tested, we have engineered two hydrolases, which exhibit an impressive 51-(EH102$_{DM2}$) and 63-(EH102$_{TM2}$) substrate repertoire. Such a repertoire approximates that of the most substrate promiscuous esterases, such as EH1 and CalB, capable of hydrolyzing as much as 72 and 68 out of the 96 esters tested (Figure 8).[13] Likewise, no significant difference was observed in the size of the hydrolyzed esters between the WT enzyme and the successful variants (Figure S13), meaning that the increase in substrate promiscuity is independent of the substrate size. In addition, we also observed that our engineered hydrolase

gained the capacity to convert substrates that are rarely converted by esterases, such as pantolactone, pentadecyl acetate, and vinyl laurate, which under similar conditions used in the present study, were only hydrolyzed by four out of 147 such enzymes.[13] Also, our approach gained the capacity to hydrolyze a combination of esters such as coumaric and ferulic esters (e.g., isobutyl cinnamate, methyl 2,5-dihydroxycinnamate, methyl cinnamate, and methyl ferulate) rarely hydrolyzed together, for example, can only be observed in three other esterases.[13] Finally, it is noteworthy that the catalytic efficiency of mutant variants with the higher level of promiscuity (i.e., EH102$_{DM2}$ and EH102$_{TM2}$), measured under the same assay conditions, approached those of naturally promiscuous enzymes with a similar substrate range. This was confirmed by selecting six representative ester hydrolases reported to exhibit substrate repertoires ranging from 72 to 28 esters,[13] and 1-naphthyl acetate, and ester commonly hydrolyzed by ester hydrolases, as a model substrate (Table 2).

Table 2. Kinetic Parameters Determined for the Model Ester 1-Naphthyl Acetate

hydrolase	number of esters hydrolyzed	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} M^{-1})
EH102$_{WT}$	16	73.333	190	6418
EH102$_{TM1}$	35	0.730	84	145
EH102$_{SM1}$	36	50.174	254	3289
EH102$_{DM2}$	51	71.483	54	22,103
EH102$_{TM2}$	63	67.777	58	19,466
EH1	72	2247.0	272	40,126
EH3	69	173.260	231	33,706
EH5	67	416.670	56	32,477
EH7	64	1683.80	249	29,056
EH12	51	3360.0	657	25,240
EH37	28	59.840	35	23,827

CONCLUSIONS

Our study demonstrates that it is possible to infer computationally the properties that describe substrate promiscuity in serine ester hydrolases, and use them to increase the substrate range of a hydrolase with narrow substrate spectra. Moreover, we accomplished to enhance substrate promiscuity without compromising the turnover rate of the enzyme against its native substrates. In fact, it can be said that we transformed a low substrate promiscuous esterase into a prominently promiscuous one. Also, the created variants have gained the ability to hydrolyze esters rarely hydrolyzed by esterases, some of which, for example, lactones are of great interest for industrial purposes.

This rational design of substrate promiscuity would help surpass current problems in industrial settings, where multiple catalysts have to be used to convert different substrates with similar chemical groups.[53,54]

Thus, the substrate promiscuity of an enzyme can be enhanced by rationally optimizing several active site properties, involving not only its volume/size but also its exposure, enclosure, tightness, and so forth; such a balance is necessary for the preservation of its catalytic distances. Importantly, such an enhancement can be achieved with a very small mutant library when efficiently combining in silico and in vitro techniques.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.0c05015.

> Substrate spectra of the successful variants; auxiliary density plots and poses of the results from the PELE simulations; plots of the metrics obtained from the MD simulations; bar plot of the properties of the active site cavity for the remaining variants; representation of the predicted $\Delta\Delta G$ of the tested variants; violin plot of the volume of the hydrolyzed esters by the WT enzyme and the successful variants; and protein purity as determined by SDS-PAGE (PDF)

AUTHOR INFORMATION

Corresponding Authors

Manuel Ferrer − Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28049, Spain; orcid.org/0000-0003-4962-4714; Email: mferrer@icp.csic.es

Victor Guallar − Barcelona Supercomputing Center (BSC), Barcelona 08034, Spain; Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain; orcid.org/0000-0002-4580-1114; Email: victor.guallar@bsc.es

Authors

Sergi Roda − Barcelona Supercomputing Center (BSC), Barcelona 08034, Spain; orcid.org/0000-0002-0174-7435

Laura Fernandez-Lopez − Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28049, Spain; orcid.org/0000-0001-8861-3191

Rubén Cañadas − Barcelona Supercomputing Center (BSC), Barcelona 08034, Spain

Gerard Santiago − Barcelona Supercomputing Center (BSC), Barcelona 08034, Spain; Nostrum Biodiscovery S.L., Barcelona 08028, Spain; orcid.org/0000-0002-0506-3049

Complete contact information is available at:
https://pubs.acs.org/10.1021/acscatal.0c05015

Author Contributions

⊥S.R., L.-F.L. and R.C. contributed equally.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONSPARA

EH   ester hydrolase
CalB   *Candida antarctica* lipase B
rmsd root mean square deviation
PELE protein energy landscape exploration
SASA solvent-accessible surface area
MD   molecular dynamics
WT   wild type
TM   triple mutant
DM   double mutant
SM   single mutant

■ REFERENCES

(1) Li, S.; Yang, X.; Yang, S.; Zhu, M.; Wang, X. Technology Prospecting on Enzymes: Application, Marketing and Engineering. *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209017.

(2) Brannigan, J. A.; Wilkinson, A. J. Protein Engineering 20 Years on. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 964−970.

(3) Chen, R. A General Strategy for Enzyme Engineering. *Trends Biotechnol.* **1999**, *17*, 344−345.

(4) Chen, R. Enzyme Engineering: Rational Redesign versus Directed Evolution. *Trends Biotechnol.* **2001**, *19*, 13−14.

(5) Kuchner, O.; Arnold, F. H. Directed Evolution of Enzyme Catalysts. *Trends Biotechnol.* **1997**, *15*, 523−530.

(6) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew Chem. Int. Ed. Engl.* **2013**, *52*, 5700−5725.

(7) Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **2013**, *82*, 447−470.

(8) Nobeli, I.; Favia, A. D.; Thornton, J. M. Protein Promiscuity and Its Implications for Biotechnology. *Nat. Biotechnol.* **2009**, *27*, 157−167.

(9) Hult, K.; Berglund, P. Enzyme Promiscuity: Mechanism and Applications. *Trends Biotechnol.* **2007**, *25*, 231−238.

(10) Copley, S. D. An Evolutionary Biochemist's Perspective on Promiscuity. *Trends Biochem. Sci.* **2015**, *40*, 72−78.

(11) Amin, S. R.; Erdin, S.; Ward, R. M.; Lua, R. C.; Lichtarge, O. Prediction and Experimental Validation of Enzyme Substrate Specificity in Protein Structures. *Proc. Natl. Acad. Sci.* **2013**, *110*, E4195−E4202.

(12) Freund, G. S.; O'Brien, T. E.; Vinson, L.; Carlin, D. A.; Yao, A.; Mak, W. S.; Tagkopoulos, I.; Facciotti, M. T.; Tantillo, D. J.; Siegel, J. B. Elucidating Substrate Promiscuity within the FabI Enzyme Family. *ACS Chem. Biol.* **2017**, *12*, 2465−2473.

(13) Martínez-Martínez, M.; Coscolín, C.; Santiago, G.; Chow, J.; Stogios, P. J.; 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. P.; Alcaide, M.; Distaso, M.; Mesa, V.; Peláez, A. I.; Sánchez, J.; Buchholz, P. C. F.; Pleiss, J.; Fernández-Guerra, A.; 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.; The INMARE Consortium. The Inmare Consortium. Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chem. Biol.* **2018**, *13*, 225−234.

(14) Holmquist, M. Alpha/Beta-Hydrolase Fold Enzymes: Structures, Functions and Mechanisms. *Curr. Protein Pept. Sci.* **2000**, *1*, 209−235.

(15) Nardini, M.; Dijkstra, B. W. Alpha/beta Hydrolase Fold Enzymes: The Family Keeps Growing. *Curr. Opin. Struct. Biol.* **1999**, *9*, 732−737.

(16) Lenfant, N.; Hotelier, T.; Velluet, E.; Bourne, Y.; Marchot, P.; Chatonnet, A. ESTHER, the Database of the *α/β*-Hydrolase Fold Superfamily of Proteins: Tools to Explore Diversity of Functions. *Nucleic Acids Res.* **2013**, *41*, D423−D429.

(17) Raveendran, S.; Parameswaran, B.; Ummalyma, S. B.; Abraham, A.; Mathew, A. K.; Madhavan, A.; Rebello, S.; Pandey, A. Applications of Microbial Enzymes in Food Industry. *Food Technol. Biotechnol.* **2018**, *56*, 16−30.

(18) Kawai, F.; Kawabata, T.; Oda, M. Current State and Perspectives Related to the PET Hydrolases Available for Biorecycling. *ACS Sustainable Chem. Eng.* **2020**, *8*, 8894−8908.

(19) Jendrossek, D. Extracellular Polyhydroxyalkanoate (PHA) Depolymerases: The Key Enzymes of PHA Degradation. *Biopolymers* **2005**, DOI: 10.1002/3527600035.bpol3b03.

(20) Valero, F.; Ferreira-Dias, S.; Sandoval, G.; Plou, F. The Potential Use of Lipases in the Production of Fatty Acid Derivatives for the Food and Nutraceutical Industries. *Electron. J. Biotechnol.* **2013**, *16*, 12.

(21) Guerrand, D. Lipases Industrial Applications: Focus on Food and Agroindustries. *OCLC* **2017**, *24*, D403.

(22) Panda, T.; Gowrishankar, B. S. Production and Applications of Esterases. *Appl. Microbiol. Biotechnol.* **2005**, *67*, 160−169.

(23) Yang, J.; Fu, X.; Liao, J.; Liu, L.; Thorson, J. S. Structure-Based Engineering of E. Coli Galactokinase as a First Step toward in Vivo Glycorandomization. *Chem. Biol.* **2005**, *12*, 657−664.

(24) Engleder, M.; Strohmeier, G. A.; Weber, H.; Steinkellner, G.; Leitner, E.; Müller, M.; Mink, D.; Schürmann, M.; Gruber, K.; Pichler, H. Evolving the Promiscuity of Elizabethkingia Meningoseptica Oleate Hydratase for the Regio- and Stereoselective Hydration of Oleic Acid Derivatives. *Angew Chem. Int. Ed. Engl.* **2019**, *58*, 7480−7484.

(25) Spadiut, O.; Pisanelli, I.; Maischberger, T.; Peterbauer, C.; Gorton, L.; Chaiyen, P.; Haltrich, D. Engineering of Pyranose 2-Oxidase: Improvement for Biofuel Cell and Food Applications through Semi-Rational Protein Design. *J. Biotechnol.* **2009**, *139*, 250−257.

(26) Sundermann, U.; Bravo-Rodriguez, K.; Klopries, S.; Kushnir, S.; Gomez, H.; Sanchez-Garcia, E.; Schulz, F. Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase. *ACS Chem. Biol.* **2013**, *8*, 443−450.

(27) Cassidy, J.; Bruen, L.; Rosini, E.; Molla, G.; Pollegioni, L.; Paradisi, F. Engineering Substrate Promiscuity in Halophilic Alcohol Dehydrogenase (HvADH2) by in Silico Design. *PLoS One* **2017**, *12*, e0187482.

(28) Thorson, J. S.; Barton, W. A.; Hoffmeister, D.; Albermann, C.; Nikolov, D. B. Structure-Based Enzyme Engineering and Its Impact on in Vitro Glycorandomization. *Chembiochem* **2004**, *5*, 16−25.

(29) Li, R.; Wijma, H. J.; Song, L.; Cui, Y.; Otzen, M.; Tian, Y. e.; Du, J.; Li, T.; Niu, D.; Chen, Y.; Feng, J.; Han, J.; Chen, H.; Tao, Y.; Janssen, D. B.; Wu, B. Computational Redesign of Enzymes for Regio- and Enantioselective Hydroamination. *Nat. Chem. Biol.* **2018**, *14*, 664−670.

(30) Wang, Y.-S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum. *J. Am. Chem. Soc.* **2012**, *134*, 2950−2953.

(31) Su, B.-M.; Shao, Z.-H.; Li, A.-P.; Naeem, M.; Lin, J.; Ye, L.-D.; Yu, H.-W. Rational Design of Dehydrogenase/Reductases Based on Comparative Structural Analysis of Prereaction-State and Free-State Simulations for Efficient Asymmetric Reduction of Bulky Aryl Ketones. *ACS Catal.* **2020**, *10*, 864−876.

(32) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins* **2004**, *55*, 351−367.

(33) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221−234.

(34) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.

(35) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752−1780.

(36) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110−2142.

(37) 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*, 1304−1311.

(38) Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.

(39) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(41) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.

(42) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384−2393.

(43) Chow, K.-H.; Ferguson, D. M. Isothermal-Isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling. *Comput. Phys. Commun.* **1995**, *91*, 283−289.

(44) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm. *Chem. Phys. Lett.* **2004**, *384*, 288−294.

(45) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98−103.

(46) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(47) Halgren, T. New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem. Biol. Drug Des.* **2007**, *69*, 146−148.

(48) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377−389.

(49) Sumbalova, L.; Stourac, J.; Martinek, T.; Bednar, D.; Damborsky, J. HotSpot Wizard 3.0: Web Server for Automated Design of Mutations and Smart Libraries Based on Sequence Input Information. *Nucleic Acids Res.* **2018**, *46*, W356−W362.

(50) Alonso, S.; Santiago, G.; Cea-Rama, I.; Fernandez-Lopez, L.; Coscolín, C.; Modregger, J.; Ressmann, A. K.; Martínez-Martínez, M.; Marrero, H.; Bargiela, R.; Pita, M.; Gonzalez-Alfonso, J. L.; Briand, M. L.; Rojo, D.; Barbas, C.; Plou, F. J.; Golyshin, P. N.; Shahgaldian, P.; Sanz-Aparicio, J.; Guallar, V.; Ferrer, M. Genetically Engineered Proteins with Two Active Sites for Enhanced Biocatalysis and Synergistic Chemo- and Biocatalysis. *Nat. Catal.* **2020**, *3*, 319−328.

(51) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.

(52) Santiago, G.; Martínez-Martínez, M.; Alonso, S.; Bargiela, R.; Coscolín, C.; Golyshin, P. N.; Guallar, V.; Ferrer, M. Rational Engineering of Multiple Active Sites in an Ester Hydrolase. *Biochemistry* **2018**, *57*, 2245−2255.

(53) Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial Biocatalysis Today and Tomorrow. *Nature* **2001**, *409*, 258−268.

(54) Ferrer, M.; Bargiela, R.; Martínez-Martínez, M.; Mir, J.; Koch, R.; Golyshina, O. V.; Golyshin, P. N. Biodiversity for Biocatalysis: A Review of the *α/β*-Hydrolase Fold Superfamily of Esterases-Lipases Discovered in Metagenomes. *Biocatal. Biotransform.* **2015**, *33*, 235−249.

SUPPORTING INFORMATION

Computationally driven rational design of substrate promiscuity on serine ester hydrolases

Sergi Roda[1,⊥], Laura Fernandez-Lopez[2,⊥], Rubén Cañadas[1,⊥], Gerard Santiago[1,3], Manuel Ferrer[2,], Victor Guallar[1,4,*]*

[1]Barcelona Supercomputing Center (BSC), Barcelona, Spain

[2]Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain.

[3]Nostrum Biodiscovery S.L., Barcelona, Spain.

[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Corresponding Author

[*]E-mail M.F.: mferrer@icp.csic.es.

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

This supporting information contains the following information:

Substrate spectra of the successful variants; Auxiliary density plots and poses of the results from

the PELE simulations; Plots of the metrics obtained from the MD simulations; Bar plot of the

properties of the active site cavity for the remaining variants; Representation of the predicted $\Delta\Delta G$ of the tested variants; Violin plot of the volume of the hydrolyzed esters by the WT enzyme and the successful variants; Protein purity as determined by SDS-PAGE.

Table S1. Substrate spectra of the EH102$_{WT}$, EH102$_{TM1}$, EH102$_{SM1}$, EH102$_{DM2}$, and EH102$_{TM2}$. The chemically and structurally distinct esters out of the 96 tested for which activity was detected in one of the enzymes are listed on the left side of the table. The k_{cat} (in min^{-1}), K_M (in µM), and k_{cat}/K_M (in s^{-1}M^{-1}) for each substrate when tested at pH 8.0, 30ºC and 50 mM ester is shown for each variant.

Ester	Volume [Å^3]	k_{cat} (WT)	K_M (WT)	k_{cat} (TM1)	K_M (TM1)	k_{cat} (SM1)	K_M (SM1)	k_{cat} (DM2)	K_M (DM2)	k_{cat} (TM2)	K_M (TM2)	k_{cat}/K_M (WT)	k_{cat}/K_M (TM1)	k_{cat}/K_M (SM1)	k_{cat}/K_M (DM2)	k_{cat}/K_M (TM2)
1-Naphthyl acetate	172.56	73.333	190.431	0.73	84.201	50.174	254.254	71.483	53.903	67.777	58.029	6418.19	144.58	3288.95	22102.65	19466.32
1-Naphthyl butyrate	206.17	28.364	216.426	0.337	133.192	0.852	267.123	8.89	118.966	0.49	219.831	2184.3	42.18	53.15	1245.41	37.12
Glyceryl triacetate	196.65	31.157	140.649	0.07	2481.238	27.97	122.617	107.094	2051.888	33.921	2012.232	3692.1	0.47	3801.77	869.88	280.95
Glyceryl tripropionate	247.06	52.001	103.077	3.379	1032.629	41.337	117.546	87.769	1679.774	49.551	1740.74	8408.19	54.54	5861.03	870.84	474.43
Glyceryl tributyrate	297.46	0	0	0.608	2003.653	11.989	36.585	2.874	60.148	12.194	2019.009	0	5.06	5461.67	796.48	100.66
Hexyl acetate	157.73	5.523	2300	0	0	0.423	2019.227	6.3	712.398	1.64	933.16	40.02	0	3.49	147.39	29.3
Octyl acetate	191.34	2.518	5770.386	0	0	0.87	2286.156	2.882	779.375	1.305	783.017	7.27	0	6.34	61.63	27.77
Lauryl acetate	258.55	0	0	0	0	0	0	0.111	148.226	0.02	145.103	0	0	0	12.49	2.26
Pentadecyl acetate	308.95	0	0	0	0	0	0	0.034	84.337	0.016	291.367	0	0	0	6.68	0.9

Ethyl acetate	90.53	0	0	0.037	1583.028	0	0	1.421	3518.087	1.396	3007.892	0	0.39	0	6.73	7.73
Ethyl propionate	107.33	0	0	0.415	6581.62	0	0	6.698	1598.812	13.673	7032.178	0	1.05	0	69.82	32.41
Ethyl butyrate	124.13	0	0	0.496	7404.588	0	0	11.808	7960.656	10.423	7234.923	0	1.12	0	24.72	24.01
Ethyl hexanoate	157.73	0	0	0.658	2354.666	0	0	0	0	1.352	2050.21	0	4.66	0	0	10.99
Ethyl octanoate	191.34	0	0	0.331	1026.549	0.024	2008.336	0	0	0.818	1012.307	0	5.37	0.2	0	13.47
Ethyl benzoate	145.37	0	0	0.462	2251.863	0	0	0	0	0	0	0	3.42	0	0	0
Methyl (R)-(−)-mandelate	153.42	0	0	0.754	3914.657	0.032	4943.662	3.67	7408.73	0.035	6182.971	0	3.21	0.11	8.26	0.1
Methyl (S)-(+)-mandelate	153.42	0	0	0.806	3432.836	0.035	5807.692	3.908	7025.935	0.04	6298.71	0	3.91	0.1	9.27	0.1
Ethyl (R)-(+)-4-chloro-3-hydroxybutyrate	145.95	0	0	0.473	14287.858	0.043	10010.578	0.294	10029.986	0.028	10001.886	0	0.55	0.07	0.49	0.05
Ethyl (S)-(−)-4-chloro-3-hydroxybutyrate	145.95	0	0	0.592	14418.143	0.028	10006.348	0.26	10039.282	0.003	10368.904	0	0.68	0.05	0.43	0.01
(+)-Ethyl D-lactate	115.37	0	0	0.191	9147.034	0.002	0.082	3.968	10043.978	0.024	9000.167	0	0.35	493.25	6.58	0.04
(−)-Ethyl L-lactate	115.37	0	0	0.73	5273.524	0.055	7003.599	3.742	9596.182	0.052	10001.698	0	2.31	0.13	6.5	0.09
Propylparaben	170.19	0	0	0	0	0	0	0	0	0.028	3353.753	0	0	0	0	0.14
Butylparaben	187	0	0	0	0	0.032	1079.252	0	0	0.035	430.558	0	0	0.49	0	1.37
Methyl 2-hydroxybenzoate	136.59	0	0	0	0	0	0	0	0	0.024	4669.412	0	0	0	0	0.08
(−)-Methyl (R)-3-hydroxyvalerate	132.17	0	0	0	0	0.024	994.529	9.726	1613.536	0.024	1002.605	0	0	0.4	100.46	0.4

(+)-Methyl (S)-3-hydroxyvalerate	132.17	0	0	0	0	0	812.1	10.265	1970.987	2.997	1329.844	0	0	0	86.8	37.56
Benzyl (R)-(+)-2-hydroxy-3-phenylpropionate	241.87	0	0	0.325	1585.156	0.043	2404.225	25.923	1246.952	14.425	1063.935	0	3.42	0.3	346.48	225.97
Methyl benzoate	128.57	0	0	0	0	0	0	0	0	11.625	175.373	0	0	0	0	1104.8
Methyl butyrate	107.33	5.21	857.6	0.177	657.126	0	0	15.012	637.269	10.137	837.352	101.25	4.49	0	392.6	201.76
Propyl propionate	124.13	0	0	0	0	0	0	26.693	1766.309	9.903	1996.962	0	0	0	251.87	82.65
Propyl butyrate	140.93	0	0	0	0	0	0	16.185	3250.288	9.803	3169.595	0	0	0	82.99	51.55
Propyl hexanoate	174.54	0	0	0	0	0	0	8.209	935.21	7.972	1050.902	0	0	0	146.3	126.42
Phenylethyl cinnamate	244.44	0	0	0	0	0	0	0	0	10.051	4892.807	0	0	0	0	34.24
Isobutyl cinnamate	206.18	0	0	0	0	0	0	0	0	7.181	2801.922	0	0	0	0	42.71
Methyl 2,5-dihydroxycinnamate	172.03	0	0	0	0	0	0	0	0	5.386	47283.661	0	0	0	0	1.9
Methyl cinnamate	155.99	0	0	0.032	873.919	0	0	0	0	9.956	1039.315	0	0.6	0	0	159.66
Methyl ferulate	189.55	0	0	0	0	0	0	0	0	7.94	8482.143	0	0	0	0	15.6
Vinyl acetate	84.89	0	0	0	0	0	0	31.268	8538.378	7.513	8214.097	0	0	0	61.03	15.24
Vinyl propionate	101.7	28.886	3873	0.278	17461.903	0	0	44.158	4559.891	10.808	4539.135	124.31	0.27	0	161.4	39.68
Vinyl butyrate	118.5	0	0	0	0	0.02	3691.901	10.112	1274.796	6.382	1628.005	0	0	0.09	132.21	65.33
Vinyl laurate	252.91	0	0	0	0	0.028	1810.588	7.727	1380.855	6.35	1608.851	0	0	0.26	93.27	65.79
Vinyl benzoate	139.74	0	0	0	0	0.032	0.357	9.074	5634.586	9.64	5713.01	0	0	1474.65	26.84	28.12

Vinyl crotonate	112.31	0	0	0.037	3041.213	0	0	1.613	3194.303	6.817	3974.857	0	0.2	0	8.41	28.58
Vinyl acrylate	96.06	0	0	0	0	0	0	2.753	8124.545	7.046	7867.561	0	0	0	5.65	14.93
Geranyl acetate	212.09	5.454	348.68	0.361	127.961	5.879	346.301	2.908	226.124	7.374	275.352	260.69	47.06	282.97	214.32	446.36
3-Methyl-3-buten-1-yl acetate	135.06	1.93	14389.167	0.078	3988.244	4.734	14029.363	9.229	4700.354	5.165	4158.835	2.24	0.33	5.62	32.72	20.7
Ethyl 2-ethylacetoacetate	159.7	0	0	0	0	0	0	0.084	4574.094	0.02	4831.152	0	0	0	0.31	0.07
Ethyl 2-methylacetoacetate	142.9	0	0	0	0	0	0	1.067	816.971	0.032	720.065	0	0	0	21.76	0.73
Ethyl 3-oxohexanoate	159.92	0	0	0.034	6823.829	0.028	9258.226	0.212	7029.926	0.047	5433.857	0	0.08	0.05	0.5	0.15
Ethyl acetoacetate	126.31	0	0	0	0	0.02	0.472	0.808	4712.216	0.032	4171.959	0	0	695.48	2.86	0.13
Ethyl propionylacetate	143.11	0	0	0.496	9601.195	11.752	9166.816	29.798	9147.786	12.938	9041.194	0	0.86	21.37	54.29	23.85
ϒ-Valerolactone	96.75	0	0	0.939	1774.929	12.139	3025.181	24.979	4163.669	11.965	4092.198	0	8.82	66.88	99.99	48.73
Methyl glycolate	81.98	21.088	879.015	1.327	3348.122	27.786	582.189	96.858	4226.656	30.868	5057.499	399.84	6.6	795.46	381.93	101.72
D-Pantolactone	121.03	0	0	0	0	0.028	2574.535	0.773	7181.599	0.028	7453.532	0	0	0.18	1.79	0.06
L-Pantolactone	121.03	0	0	0	0	0.035	2050.751	0.746	6011.561	0.012	6882.659	0	0	0.29	2.07	0.03
Ethyl 2-chlorobenzoate	158.91	0	0	0.146	7253.831	0	0	0	0	0	0	0	0.34	0	0	0
Cyclohexyl butyrate	180.76	0	0	0.372	1346.775	0	0	0.264	77.684	0.035	318.4	0	4.6	0	56.59	1.86
n-Pentyl benzoate	195.78	0	0	0	0	0	0	0	0	0.02	437.012	0	0	0	0	0.75
(+)-Methyl D-lactate	98.57	0	0	0	0	0.016	312.168	1.011	1519.98	0.024	2969.279	0	0	0.84	11.09	0.13

|																
(−)-Methyl L-lactate	98.57	0	0	0	0	0.032	505.391	0.922	1161.796	0.095	2010.134	0	0	1.04	13.22	0.79
Propyl acetate	107.33	49.066	30295.44	1.613	51612.062	0.039	30002.36	3.185	50302.994	0.039	50007.071	26.99	0.52	0.02	1.06	0.01
Butyl acetate	124.13	37.904	13784.602	0.91	42471.288	28.637	14803.671	35.044	40107.187	32.186	40057.88	45.83	0.36	32.24	14.56	13.39
Phenyl acetate	128.57	114	49.72	4.504	98.035	90.501	51.782	208.072	92.457	85.723	90.754	38214	765.75	29128.89	37507.82	15742.7
Phenyl propionate	145.37	48.281	94.41	9.966	43.536	63.602	91.279	216.103	42.755	67.791	45.072	8523.31	3815.27	11613.23	84240.61	25067.46
Glucose pentaacetate	338.38	152.124	398.38	3.812	720.986	122.732	278.727	46.902	865.327	119.348	812.593	6364.28	88.13	7338.84	903.35	2447.88
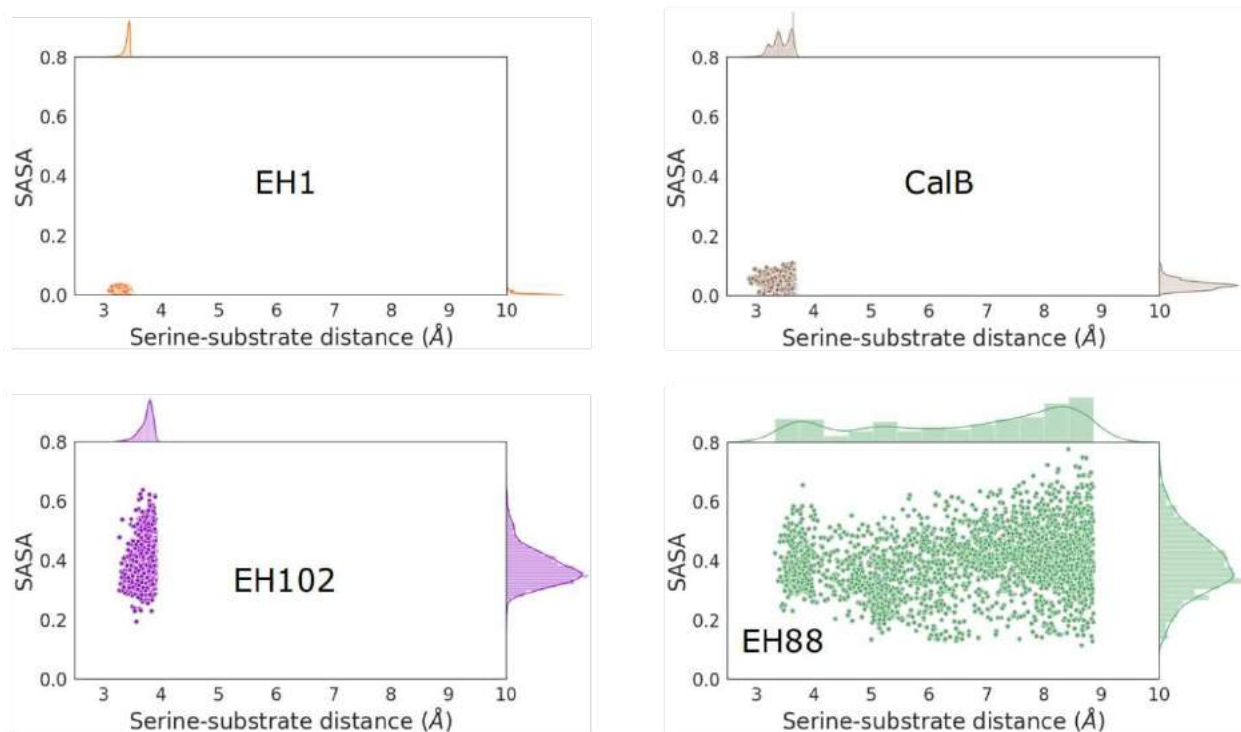
Figure S1. Density plots of the distribution of the catalytic serine-substrate distance against the SASA from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps

in the PELE simulations for the different studied esterases with phenyl acetate as substrate (3226 data points for CalB, 3095 for EH1, 2647 for EH102, and 2734 for EH88). The colors in the density plots represent each esterase: CalB (brown), EH1 (orange), EH102 (purple), and EH88 (green). The density plots were created with the Matplotlib library[1].
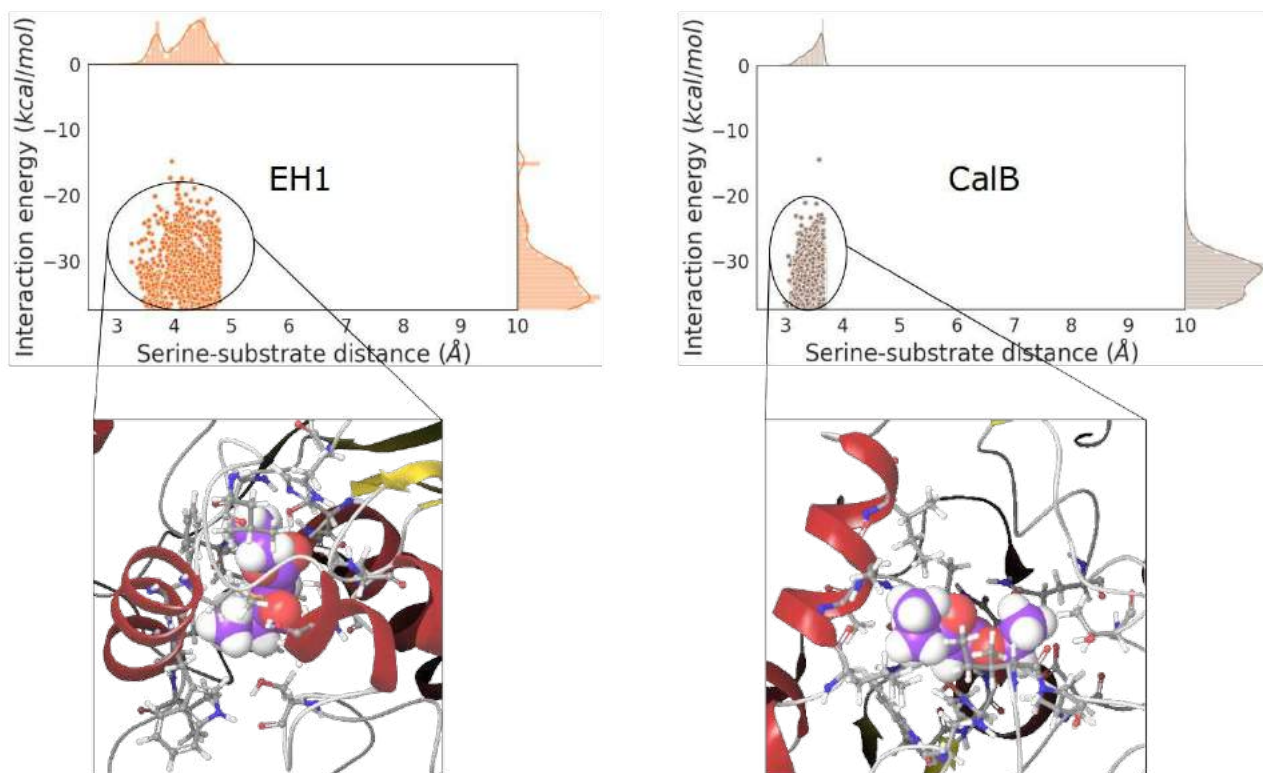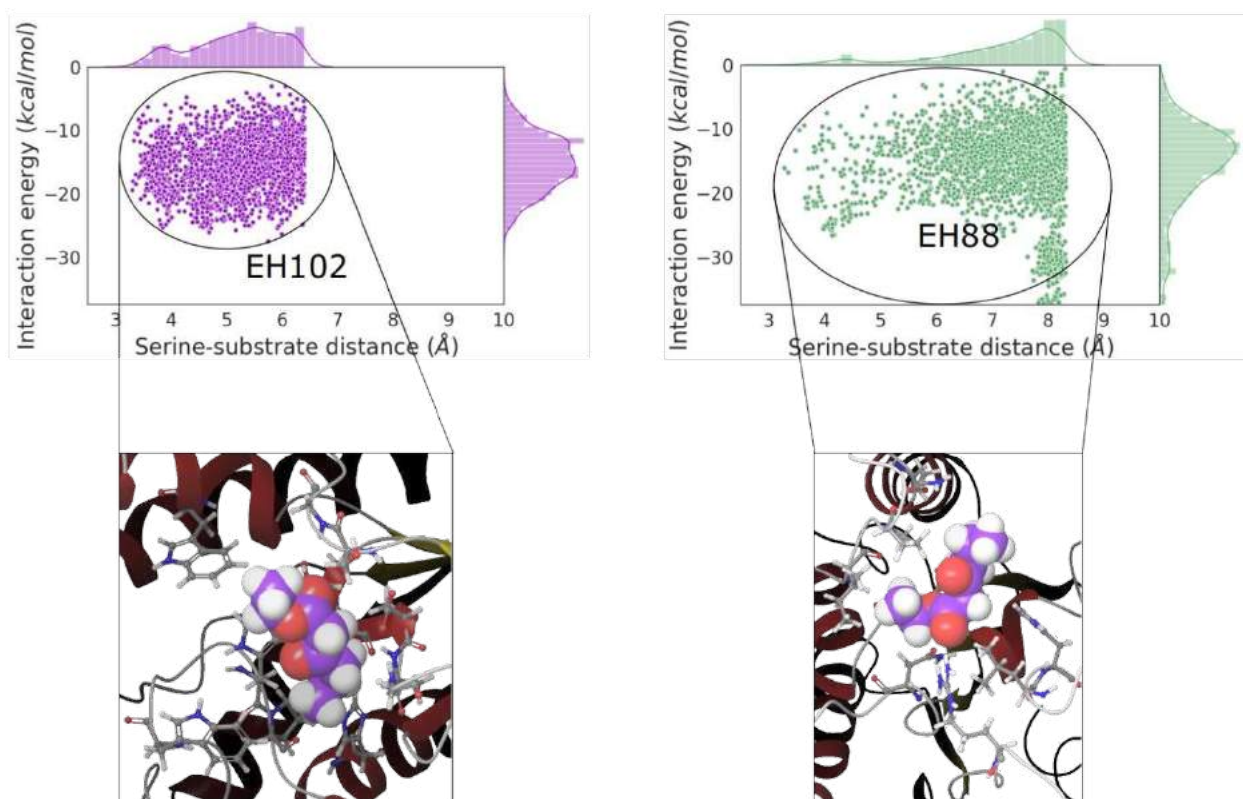
Figure S2. Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulations for the different studied esterases with ethyl-3-oxohexanoate as substrate (2732 data points for CalB, 2568 for EH1, 2332 for EH102, and 2557 for EH88), accompanied with one representing binding pose obtained during the simulation. The colors in the density plots represent each esterase: CalB (brown), EH1 (orange), EH102 (purple), and EH88 (green). The density plots were created with the Matplotlib library[1]. In the shown binding poses, the ligand has the C atoms stained in lilac and the overall structure represented in the CPK model, followed by the ball-and-stick representation of the residues 4 Å far from it. Protein ribbon is colored according to the secondary structure (ruby: α-helix, golden: β-sheet, and gray: loops).
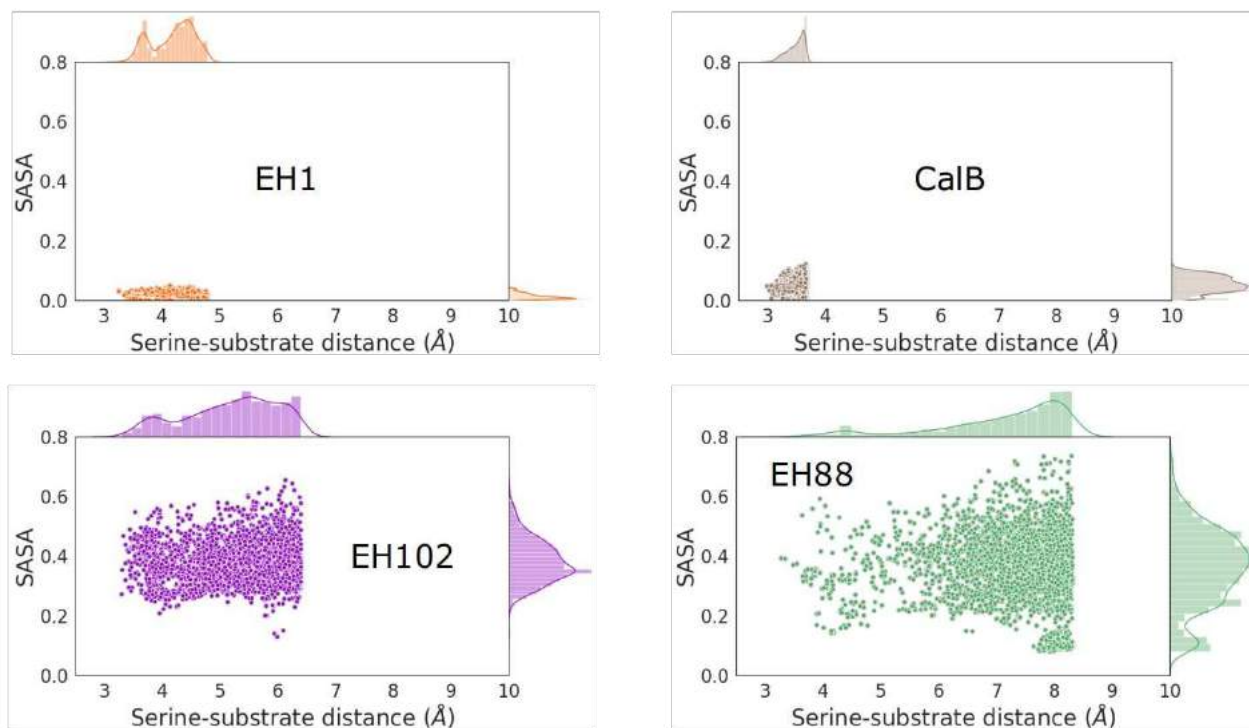
Figure S3. Density plots of the distribution of the catalytic serine-substrate distance against the SASA from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulation for the different studied esterases with ethyl-3-oxohexanoate as substrate (2732 data points for CalB, 2568 for EH1, 2332 for EH102, and 2557 for EH88). The colors in the density plots represent each esterase: CalB (brown), EH1 (orange), EH102 (purple), and EH88 (green). The density plots were created with the Matplotlib library[1].
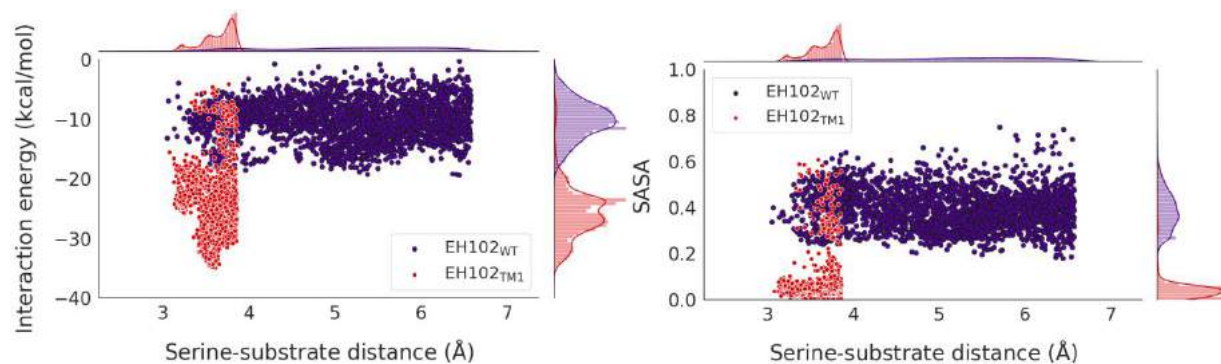
Figure S4. Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy (left) and SASA (right) from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulation for EH102$_{WT}$ and EH102$_{TM1}$ with vinyl crotonate as substrate (2910 data points for EH102$_{WT}$ and 3445 for EH102$_{TM1}$). The density plots were created with the Matplotlib library[1].
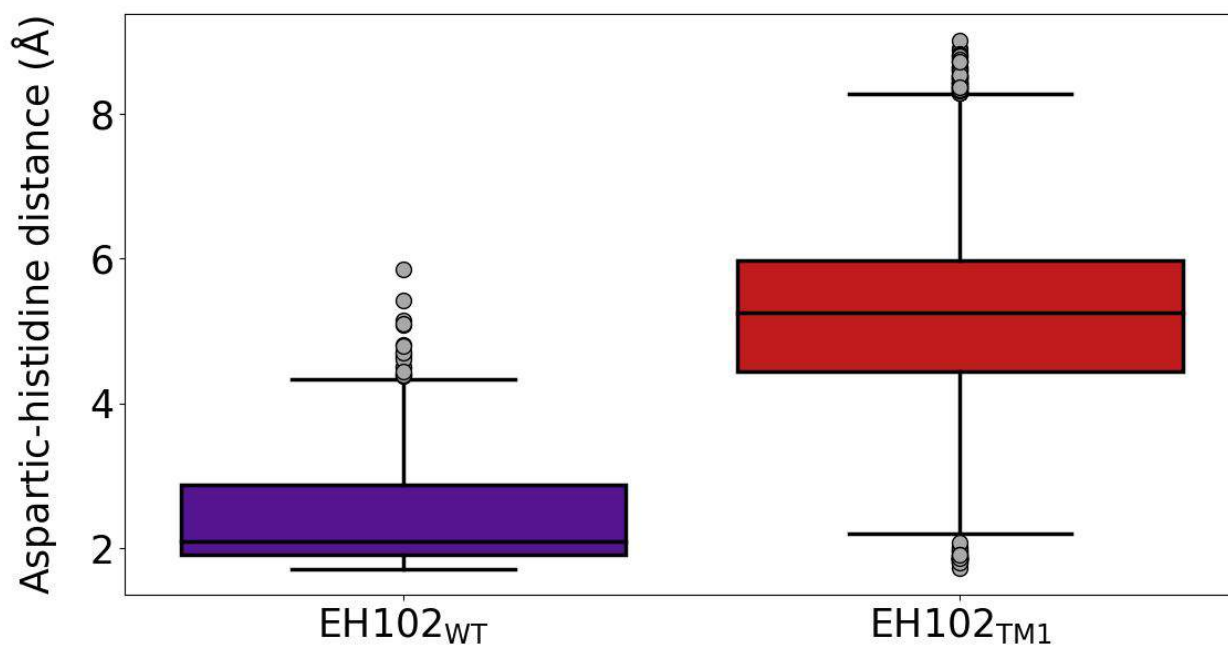

Figure S5. Box plot representing the aspartic-histidine distance ($d_{O_\delta^{Asp} - H_{\delta N}^{His}}$) along the 500 ns of the 4 MD replicas performed for EH102$_{WT}$ and EH102$_{TM1}$. The figure was created with the Matplotlib library[1].
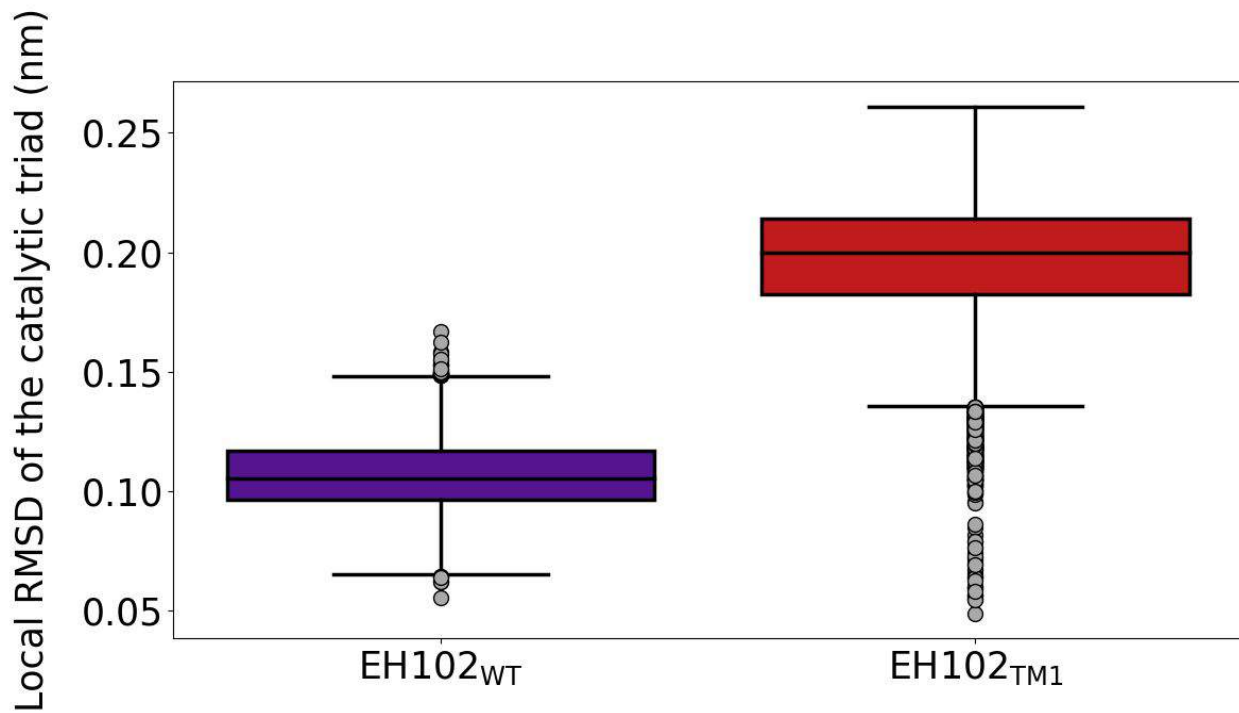
Figure S6. Box plot representing the local RMSD of the catalytic triad along the 500 ns of the 4 MD replicas (excluding the initial frame) performed for EH102_{WT} and EH102_{TM1}. The figure was created with the Matplotlib library[1].
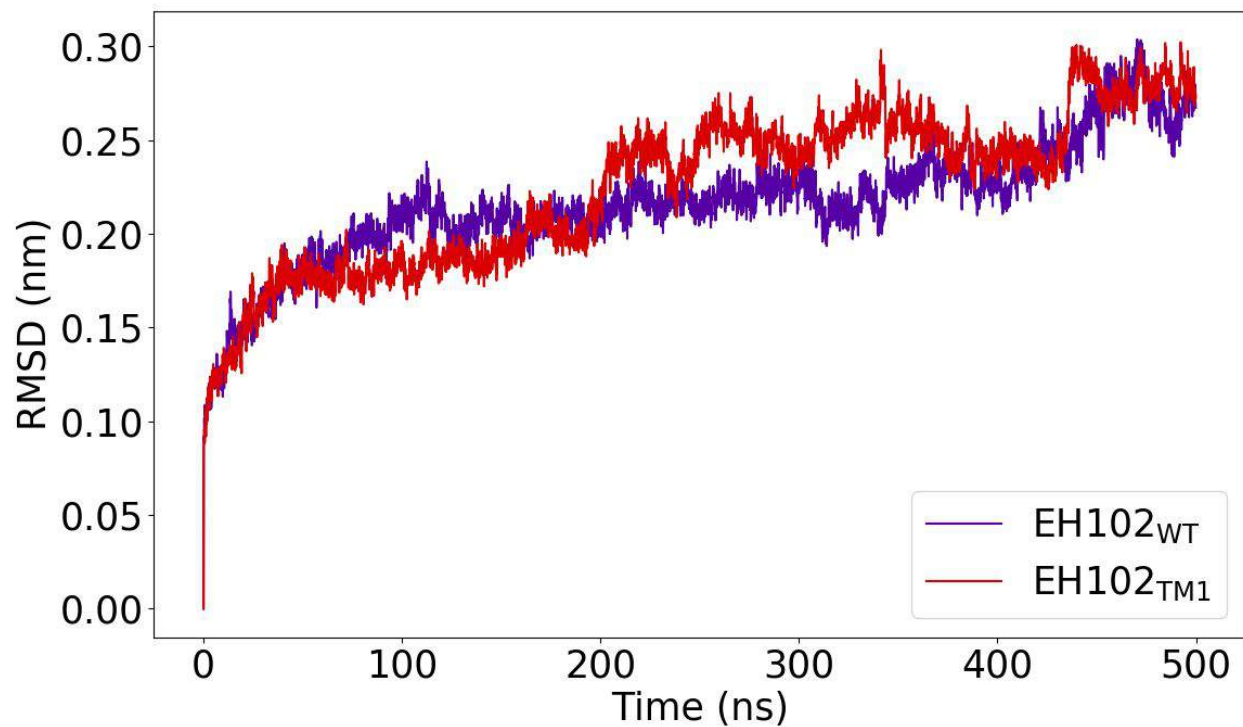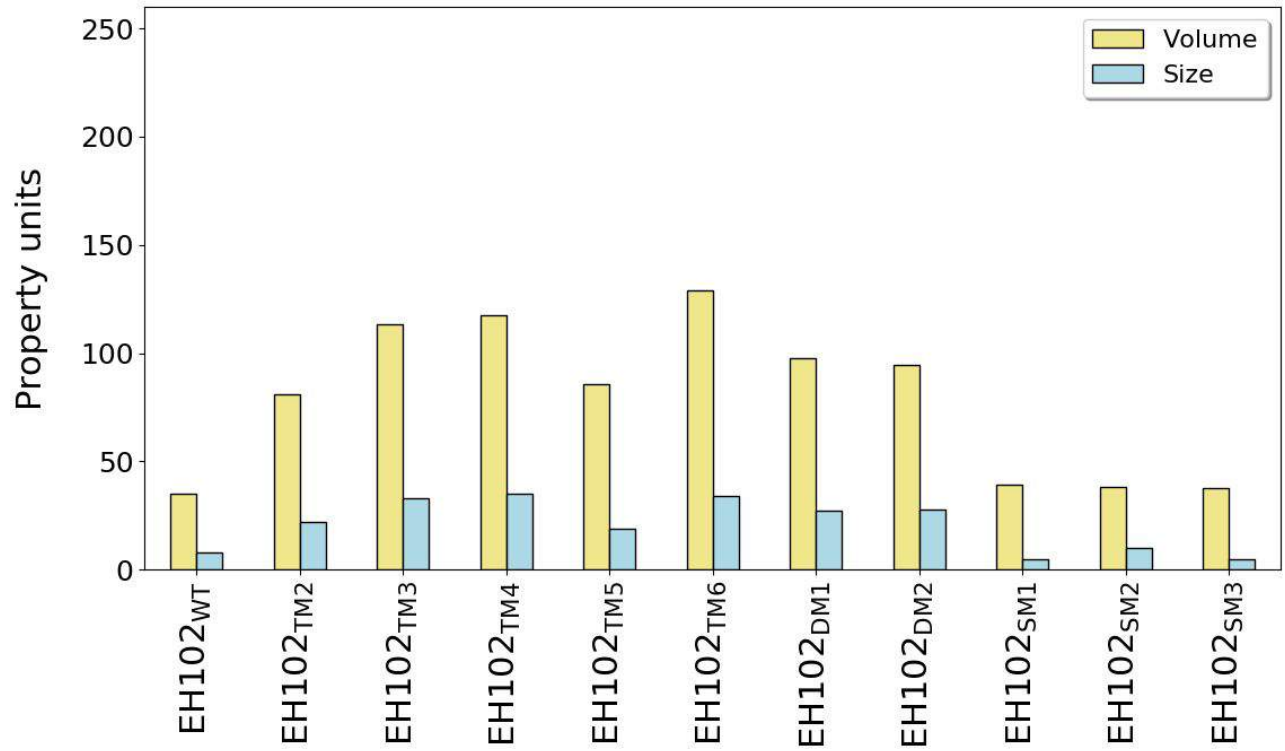
Figure S7. Plot representing the global RMSD of the protein backbone along the 500 ns of the 4 MD replicas performed for EH102$_{WT}$ and EH102$_{TM1}$. The plot was created with the Matplotlib library[1].
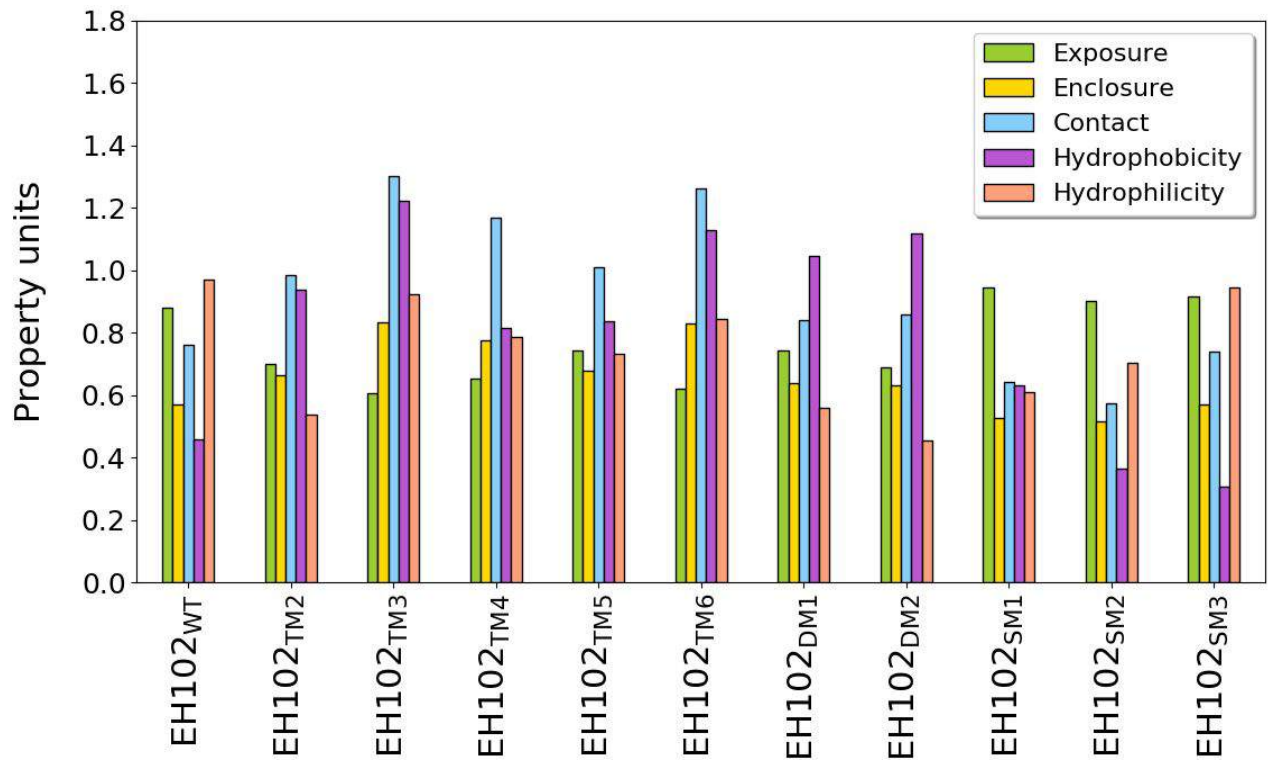
A

B

Figure S8. Bar plot representing the Volume (in Å^3) and Size of the active site cavity of the EH102$_{WT}$ in comparison with EH102$_{TM2}$, EH102$_{TM3}$, EH102$_{TM4}$, EH102$_{TM5}$, EH102$_{TM6}$, EH102$_{DM1}$, EH102$_{DM2}$, EH102$_{SM1}$, EH102$_{SM2}$, and EH102$_{SM3}$ obtained with SiteMap[2,3] (Figure S8A). Bar plot representing the Exposure, Enclosure, Contact, Hydrophobicity, and Hydrophilicity properties of the active site cavity of the EH102$_{WT}$ in comparison with EH102$_{TM2}$, EH102$_{TM3}$, EH102$_{TM4}$, EH102$_{TM5}$, EH102$_{TM6}$, EH102$_{DM1}$, EH102$_{DM2}$, EH102$_{SM1}$, EH102$_{SM2}$, and EH102$_{SM3}$ obtained with SiteMap[2,3] (Figure S8B). The figure was created with the Matplotlib library[1].

Figure S9. Box plot representing the serine-histidine distance ($d_{H_\gamma^{Ser} - N_\varepsilon^{His}}$) along the 500 ns of the 4 MD replicas performed for EH102$_{WT}$, EH102$_{TM2}$, EH102$_{TM3}$, EH102$_{TM4}$, EH102$_{TM5}$, EH102$_{TM6}$, EH102$_{DM1}$, EH102$_{DM2}$, EH102$_{SM1}$, EH102$_{SM2}$, and EH102$_{SM3}$. The figure was created with the Matplotlib library[1].
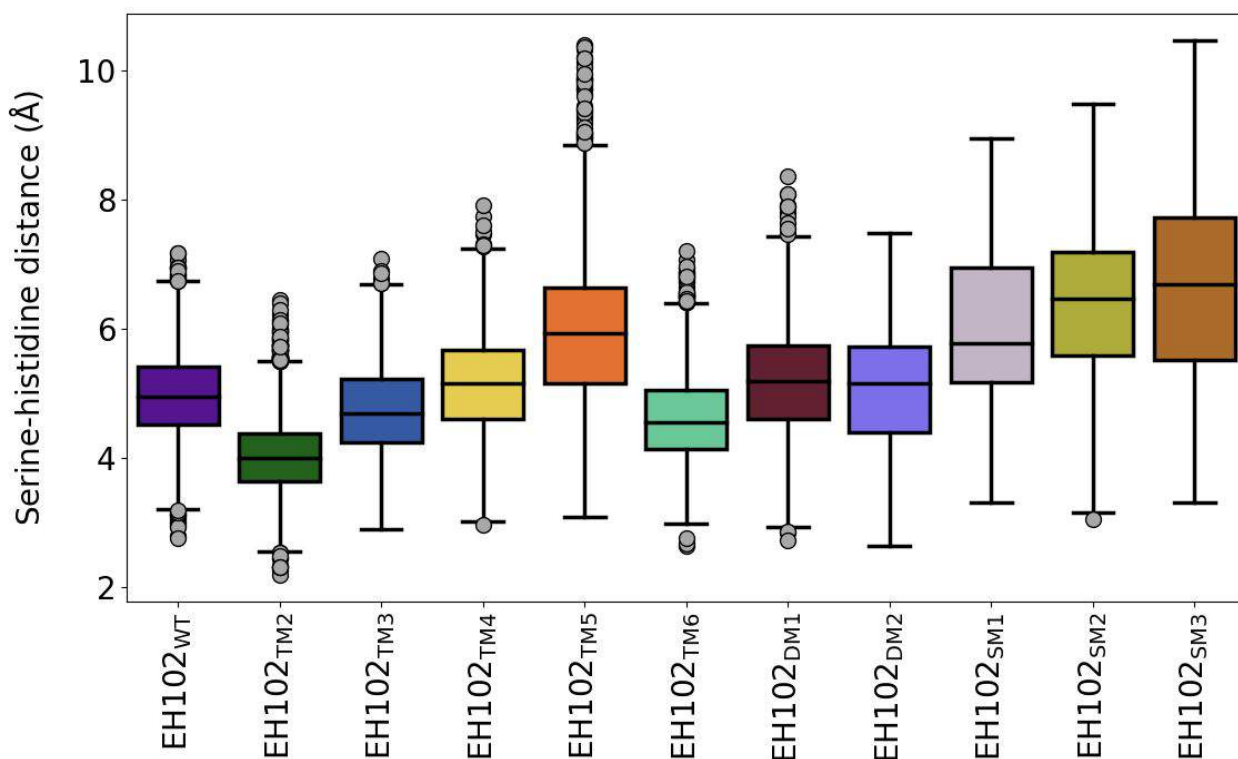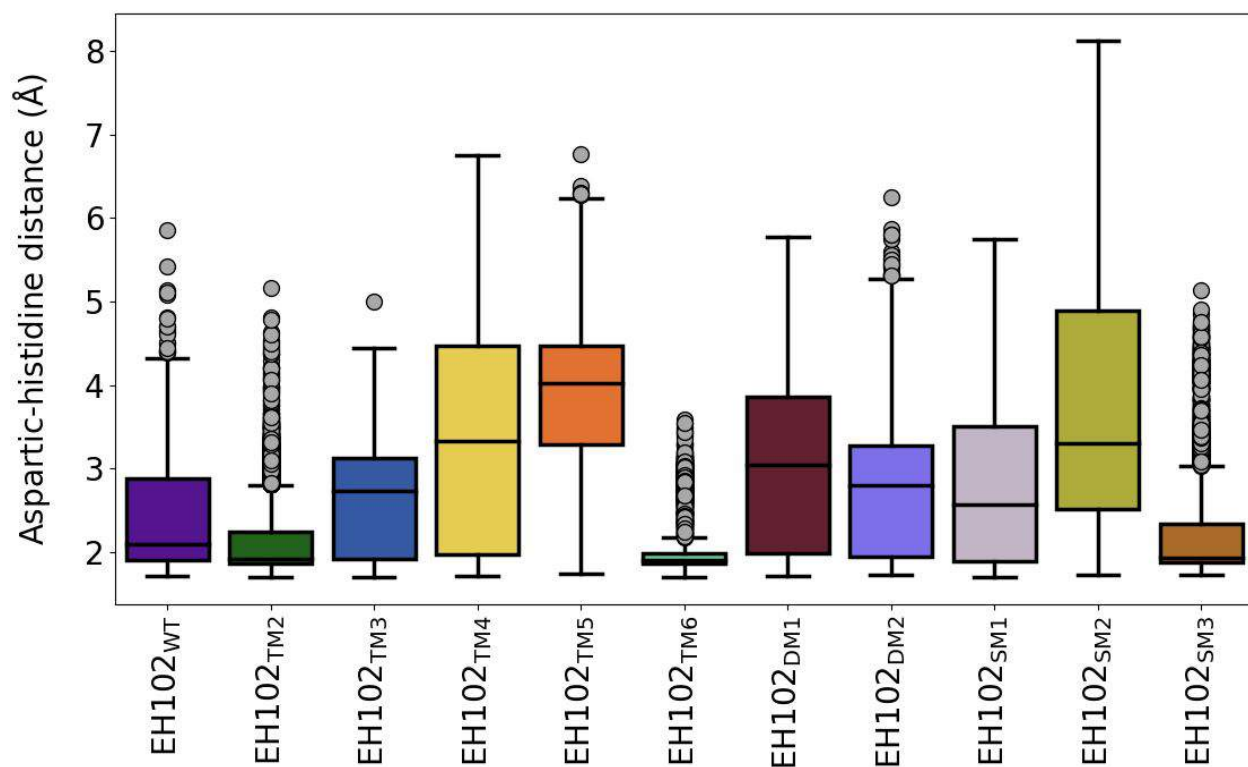
Figure S10. Box plot representing the aspartic-histidine distance ($d_{O_\delta^{Asp} - H_{\delta N}^{His}}$) along the 500 ns of the 4 MD replicas performed for EH102$_{WT}$, EH102$_{TM2}$, EH102$_{TM3}$, EH102$_{TM4}$, EH102$_{TM5}$, EH102$_{TM6}$, EH102$_{DM1}$, EH102$_{DM2}$, EH102$_{SM1}$, EH102$_{SM2}$, and EH102$_{SM3}$. The figure was created with the Matplotlib library[1].
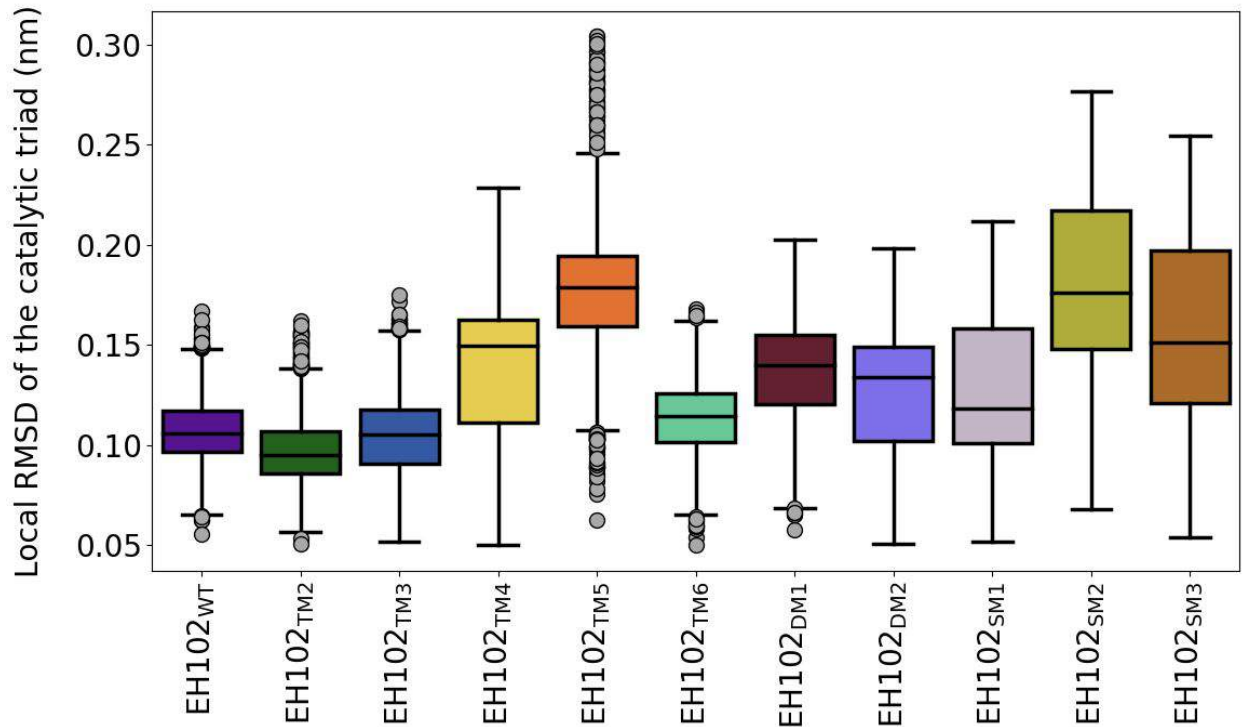
Figure S11. Box plot representing the local RMSD of the catalytic triad along the 500 ns of the 4 MD replicas (excluding the initial frame) performed for EH102$_{WT}$, EH102$_{TM2}$, EH102$_{TM3}$, EH102$_{TM4}$, EH102$_{TM5}$, EH102$_{TM6}$, EH102$_{DM1}$, EH102$_{DM2}$, EH102$_{SM1}$, EH102$_{SM2}$, and EH102$_{SM3}$. The figure was created with the Matplotlib library[1].
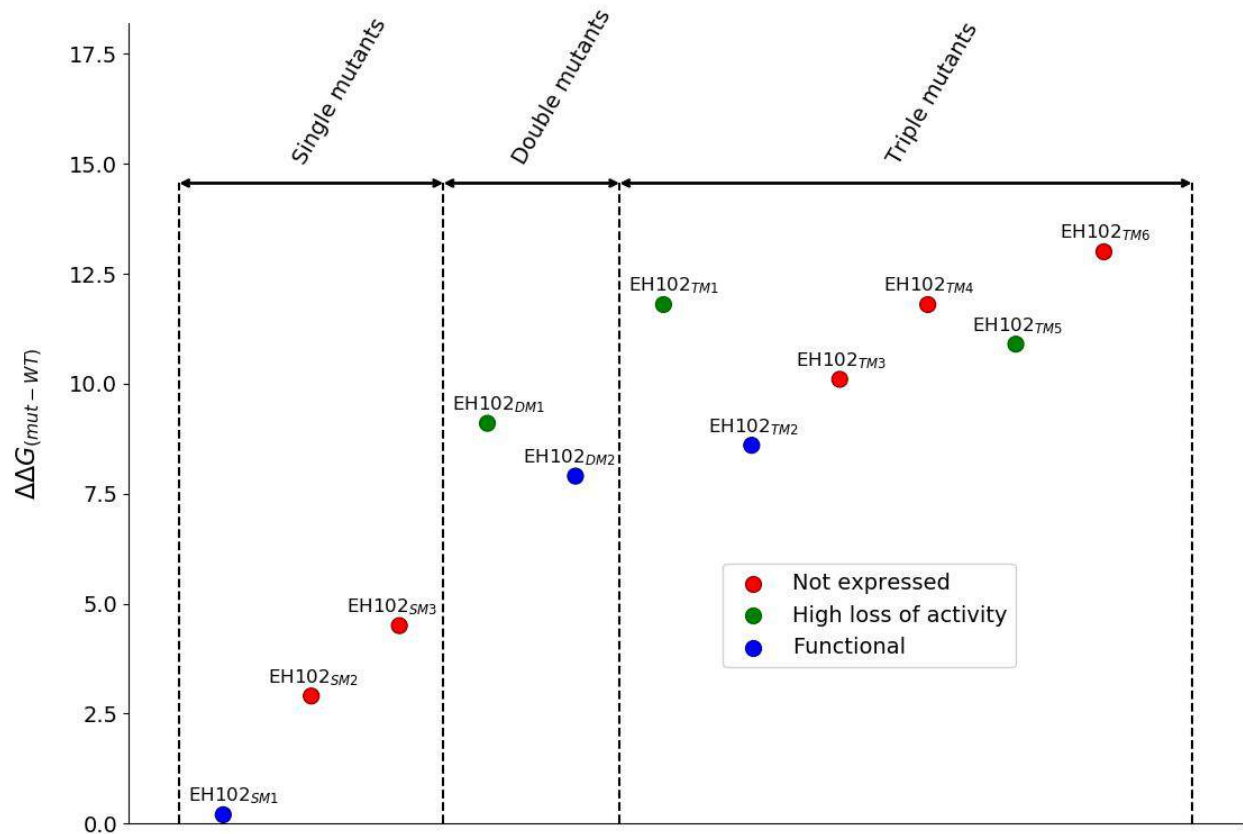
Figure S12. predicted $\Delta\Delta G_{(mut-WT)}$ of the experimentally tested variants calculated using the module of thermodynamic stability from HotSpot Wizard[4]. The variants that could not be expressed are highlighted in red, the ones that lost a big part of their activity in green, and the ones that were fully functional in blue. The figure was created with the Matplotlib library[1].

Figure S13. Violin plot of the volume of the hydrolyzed esters by $EH102_{WT}$, $EH102_{TM1}$, $EH102_{SM1}$, $EH102_{DM2}$, and $EH102_{TM2}$. The figure was created with the Matplotlib library[1].
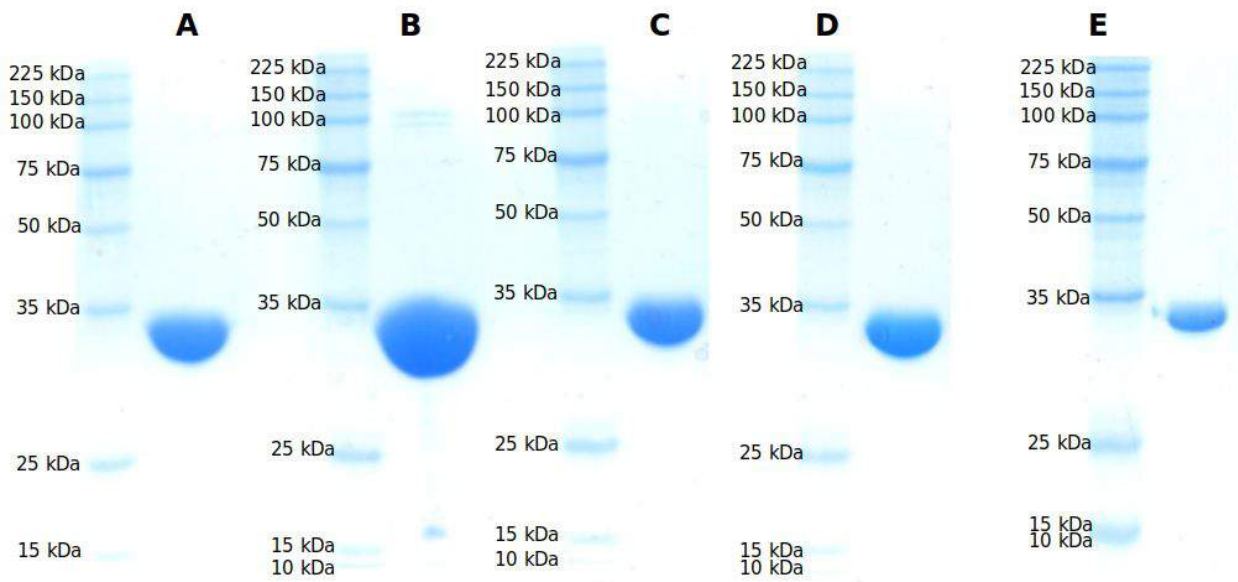
A: I16V (SM1)
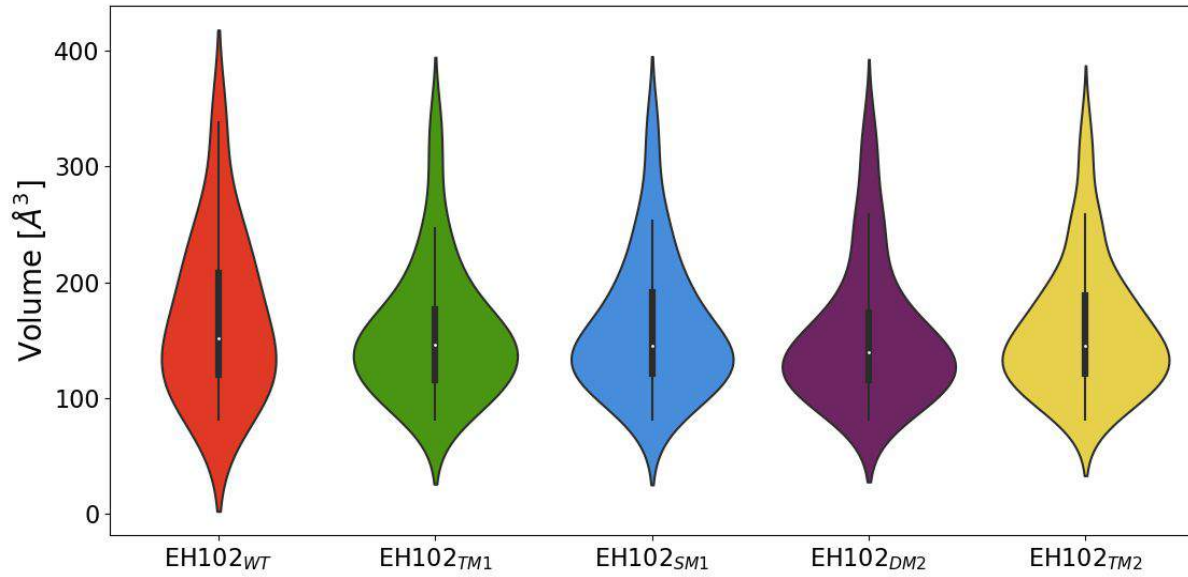**B:** I92A/W96G (DM2)
C: I16V/I92A/W96G (TM2)
D: I16G/I92A/W96G (TM1)
E: $EH102_{WT}$

Figure S14. Protein purity as determined by SDS-PAGE. A Coomassie brilliant blue (Protoblue Safe, National Diagnostics, GA, USA) stained 15% SDS-PAGE gel in which a total of 12-18 µg of proteins purified after the His6-tag purification step, are shown. Purity of the proteins was high (>99% by densitometry). The theoretical molecular mass of the native EH102 protein is 24328.8 Da, respectively. SDS-PAGE was performed in a Mini PROTEAN electrophoresis system (Bio-Rad).

REFERENCES

(1) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9* (3), 90–95.

(2) Halgren, T. New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem. Biol. Drug Des.* **2007**, *69* (2), 146–148.

(3) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49* (2), 377–389.

(4) Sumbalova, L.; Stourac, J.; Martinek, T.; Bednar, D.; Damborsky, J. HotSpot Wizard 3.0: Web Server for Automated Design of Mutations and Smart Libraries Based on Sequence Input Information. *Nucleic Acids Res.* **2018**, *46* (W1), W356–W362.

AsiteDesign: A semi-rational algorithm for automated enzyme design

Sergi Roda[1*], Henrik Terholsen[3*], Jule Ruth Heike Meyer[3], Victor Guallar[1,2§], Uwe Bornscheuer[3§], Masoud Kazemi[1,4§]

1. *Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain*
2. *Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain*
3. *Institute of Biochemistry, Department of Biotechnology & Enzyme Catalysis, University of Greifswald, Felix-Hausdorff-Str. 4, D-17487 Greifswald, Germany*
4. *Biomatter Designs, Žirmūnų g. 139A, Vilnius, Lithuania*

*Equal contribution.

§Corresponding authors.

ORCIDs of authors:
Sergi Roda: 0000-0002-0174-7435
Henrik Terholsen: 0000-0001-9625-8177
Victor Guallar: 0000-0003-3274-2482
Jule Ruth Heike Meyer:
0000-0003-0293-4686

Uwe Bornscheuer: 0000-0003-0685-2696
Masoud Kazemi: 0000-0002-0750-8865

Running title:

Keywords: AsiteDesign, rational design, active site, protein engineering, computational chemistry

Abstract

With advances in protein structure predictions, the number of available high quality structures has increased dramatically. In light of these advances, structure-based enzyme engineering is expected to become increasingly important for the optimization of biocatalysts for industrial processes. Here we present AsiteDesign, a Monte Carlo-based protocol for structure based engineering of active sites. AsiteDesign provides a framework for introducing new catalytic residues in a given binding pocket to either create a new catalytic activity or alter the existing one. AsiteDesign is implemented using pyRosetta and incorporates enhanced sampling techniques to efficiently explore the search space. The protocol was tested by designing an alternative catalytic triad in the active site of *Pseudomonas fluorescens* esterase (PFE). The designed variant was experimentally verified to be active, demonstrating that AsiteDesign can find alternative catalytic triads. Additionally, the AsiteDesign protocol was employed to enhance the hydrolysis of a bulky chiral substrate (1-phenyl-2-pentyl acetate) by PFE. The experimental verification of the designed variants demonstrated that F158L/F198A and F125A/F158L mutations increased the hydrolysis of 1-phenyl-2-pentyl acetate from 8.9% to 66.7% and 23.4%, respectively, and reversed the enantioselectivity of the enzyme from (*R*) to (*S*)-enantiopreference, with 32 and 55 %*ee*, respectively.

Introduction

Structure-based enzyme engineering is widely used in the development of biocatalysts for industrial purposes[1–5]. These approaches have been employed to engineer protein thermostability, enzyme activity, or substrate selectivity. FRESCO[6], FoldX[6,7], and Rosetta[8] are among the noteworthy methods for enhancing enzyme thermostability, but many more also have been developed [9–15]. Similarly, a wide range of computational methods are available for engineering enzyme activity or for changing the selectivity, for instance, Rosetta[8], nAPOLI[16], EnzymeMiner[17], HotSpot Wizard[18], Caver[19], FireProt[ASR20], LoopGrafter[21], and DaReUS-Loop [22]. However, Rosetta and its derived methods[23,24] have become one of the most widely used tools in this area.

Current advances in deep learning structural prediction methods increased the number of available protein structures dramatically[25,26]. Additionally, *de novo* protein design techniques are becoming increasingly accurate[24,27,28], which makes it possible to create tailor made protein scaffolds. Considering these developments, structure-based enzyme engineering is expected to play an even more critical role in optimizing enzymes for industrial applications. Structure-based enzyme engineering could be used to either optimize enzyme activity and selectivity for a given substrate or to introduce new functionality into a protein cavity[29–34]. The latter approach has the potential to create enzymes that are capable of catalyzing new chemical reactions[35]. For example, by mutating catalytic glutamate, a glycosidase enzyme was converted to a glycosynthase[33,34]. Alternatively, new catalytic residues can be introduced in a protein cavity. This approach, for instance, was used to create a second active site with hydrolysis activity in a transaminase. The engineered multifunctional enzyme combines transaminase and hydrolase activities in a single protein scaffold, allowing the conversion of β-keto esters into β-amino acids, which can be used for the synthesis of a key precursor of a family of antidiabetic drugs[32].

These applications require (re)designing catalytic residues in a given binding pocket. One way to approach such tasks is by grafting an active site into a protein scaffold, which has been performed, e.g. by Rosetta Match[36,37]. This method is based on identifying a suitable cavity by searching many protein scaffolds. Here we present an alternative method, AsiteDesign, for this task. AsiteDesign is capable of identifying the best positions for a set of predefined catalytic residues in a given active site without the need for searching different protein scaffolds. The method is based on a Monte Carlo (MC) simulation and has been implemented using pyRosetta. Additionally, the protocol employs enhanced sampling techniques to improve the simulation convergence. It also includes the sampling of rotatable

substrate bonds, which can potentially improve the identification of design solutions that may not be found otherwise. Furthermore, the substrate sampling can easily be restricted to predefined intervals, allowing to perform partial optimizations for large molecules. Here we demonstrate the application of the method by designing an alternative catalytic triad in the active site of the *Pseudomonas fluorescens* esterase (PFE)[38]. The protocol was further employed to alter the enzyme selectivity for a bulky chiral substrate.

Computational methods

Design of catalytic residues

For a given enzymatic chemical reaction, the amino acid identity of the catalytic residues is known *a priori*. Considering PFE as an example, the catalytic triad consists of a nucleophile residue (Ser) that attacks the substrate, a general base (His) that accepts the proton, and an acidic residue (Asp/Glu) that activates the base. The principle of this catalytic triad is common to all serine hydrolases. As such, the main task here was to identify a set of positions in the binding pocket that can accommodate the catalytic residues at the correct distances relative to each other. This goal was achieved by first assigning random positions to each catalytic residue and performing an MC simulation to identify the optimal positions for these residues. During the simulation, in each iteration, one catalytic residue was then assigned to a new random position by mutating it, and the previously assigned position is recovered to the native amino acid. The correct distances of the catalytic residues are enforced by imposing distance restraints during simulations.

Adequate sampling could be challenging as the number of design elements increases. To overcome this, the adaptive reinforcement learning protocol[39] was incorporated into the simulation. In this scheme, the simulation is performed in epochs in which the MC sampling of the catalytic residues is performed by multiple explorers in parallel using a distributed memory parallelization scheme. At the end of each epoch, the results from explorers are collected and ranked based on a given objective function (total energy, ligand energy, restraint energy, etc.), and the next epoch is spawned by the top ranking results. To further improve the sampling, a simulated annealing scheme was also used during the simulation.

Ligand sampling is performed by dividing its atoms into core and side chains. The core atoms are only subjected to the rigid body rotation translation (RGB) perturbation. The ligand side chain atoms, in addition to RGB transformation, are perturbed by first choosing a random interval for rotatable bonds and then performing a partial optimization in the selected intervals. Compared to the conventional method of choosing a random angle, this approach can identify the most suitable conformation for chosen dihedral intervals. The minimization is performed using a two step process. The first step is an iterative grid search to identify the minimum energy conformation in internal coordinates. In this step, a multidimensional grid is constructed for the side chain rotatable bonds in the corresponding angle intervals, and the conformation with the lowest energy is identified by evaluating the ligand energy on each grid point. In the next iteration, a new higher resolution grid is constructed on a smaller

interval centered on the dihedral angles of the identified conformation from the previous step. This process is repeated three times to identify the ligand conformation with the lowest energy for given intervals. The final ligand conformation is then minimized with a gradient descent approach in cartesian coordinates.

Molecular Dynamics

Four replicas of 100 ns of molecular dynamics (MD) simulations with OPENMM[40] were performed to analyze the stability of the newly designed catalytic triads. A water cubic box (distance of 8 Å between the closest protein atom and the edge of the box) was created around the system using the TIP3P water model, and the charge of the system was stabilized using monovalent ions (Na$^+$ and Cl$^-$). The protein system was parameterized with the AMBER99SB force field. Andersen thermostat and MC barostat were applied for the NPT ensemble (constant pressure and temperature, being 1 bar and 300 K, respectively). An NVT equilibration phase lasted 400 ps using a constraint of 10 kcal/(mol·Å^2) to the whole solute system, followed by a 1 ns NPT equilibration with a milder constraint of 5 kcal/(mol·Å^2); the production run only included constraints between H and heavy atoms. The Verlet integrator with a 2 fs time step was used with an 8 Å non-bonded long-range interactions cutoff.

Protein Energy Landscape Exploration simulations

Protein Energy Landscape Exploration (PELE) was used to analyze the substrate binding of the evolved variants using AsiteDesign. PELE is a heuristic MC-based algorithm coupled with protein structure prediction methods[41,42]. The software begins by sampling the different microstates of the ligand through small rotations and translations. Applying normal modes through the anisotropic network model (ANM) approach[43], the protein's flexibility is also considered. Once the whole system has been perturbed, side chains of the residues close to the ligand are sampled to avoid steric clashes. Lastly, a truncated Newton minimization with the OPLS2005 force field is performed[44], and the new microstate is accepted or rejected based on the Metropolis criterion. The Variable Dielectric Generalized Born Non-Polar (VDGBNP) implicit solvent model[45] was used to mimic the effect of water molecules around the protein.

The exploration of the substrate was enhanced with Adaptive-PELE[39] to improve the exploration of the search space[46].

Experimental Description

Material

The chemicals *rac*-1-phenyl-2-pentanol (≥99%) and *rac*-1-phenylethyl acetate (≥98%) were ordered from Sigma-Aldrich. All other chemicals and solvents were purchased from Sigma-Aldrich, VWR or Carl Roth and were used without further treatment.

Synthesis of 1-phenyl-2-pentyl acetate

500 µL acetic anhydride and 100 µL *rac*-1-phenyl-2-pentanol were added in a 1.5 mL tube. The reaction was started by adding 10 µL pyridine to the mixture. The reaction was shaken at 25°C and 500 rpm until complete conversion was achieved. Samples of 2 µL were withdrawn and diluted in 198 µL ethyl acetate for gas chromatographic (GC) analysis. The reaction was quenched by adding the mixture to a 15 mL tube containing 2 mL ddH$_2$O. The product *rac*-1-phenyl-2-pentyl acetate (**1**) formed a second phase, was separated and dried over anhydrous sodium sulfate. The oily *rac*-1-phenyl-2-pentyl acetate was obtained in 51% yield.

Plasmid construction and site-directed mutagenesis

Synthetic genes of the PFE variants 1, 4, 8, 11, 12 in pET28a were ordered from BioCat (Heidelberg, Germany) using seamless cloning with the flanking regions 5' aaggagatatacc 3' (5' flanking) and 5' CACCACCACCACCACCACTGAGATCCGG 3' (3' flanking). The mutants are based on the sequence of the PFE wild-type (GeneBank: WP_120448209.1). The sequences were extended by a C-terminal linker (GS) and a His$_6$-tag, as is the case of the sequence used for the 1VA4 crystal structure.

PFE variants 2 and 3 were constructed based on PFE_1, PFE_5-7 were based on PFE_4, and PFE_9-10 were based on PFE_8 using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs GmbH, Ipswich, UK). Non-overlapping DNA-oligonucleotides were designed using the online NEBaseChanger tool for the mutations. The list of primers used for mutagenesis is given in Table S7. The annealing temperatures suggested by NEBaseChanger online tool (https://nebasechanger.neb.com/) were used for the polymerase chain reaction (PCR), which was performed according to the manufacturer's protocol. The obtained constructs were amplified in *E. coli* Top10 and used for heat-shock transformation of *E. coli* BL21 (DE3).

Protein preparation

Pre-cultures (4 mL LB containing kanamycin) of *E. coli* BL21 (DE3) colonies harboring the constructs for the expression of the PFE variants were incubated overnight (37°C, 180 rpm). LB medium (50 mL containing kanamycin) was inoculated with 1% (v/v) of the preculture and incubated (37°C, 180 rpm) until it reached an OD_{600} of 0.6. Protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM followed by incubation for ~20 h at 20°C at 180 rpm. Cells were harvested by centrifugation at 10,000 × g and 4°C for 3 min, and the cell pellets were resuspended with 4 mL equilibration buffer (50 mM potassium phosphate, 300 mM sodium chloride, 10 mM imidazole, pH 8.0). Cells were disrupted by sonication on ice (five cycles of 1 min sonication at 30% intensity, and 50% pulsed cycle) using a SONOPULS HD 2070 (BANDELIN Electronic GmbH & Co. KG, Berlin, Germany), and the lysates were clarified by centrifugation at 10,000 × g and 4°C for 30 min. For purification, the crude lysates were applied to 1.5 mL Roti® Garose-His/Ni Beads (Carl Roth, Karlsruhe, Germany). The resins were washed with 15 mL washing buffer (50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, pH 8.0) before target proteins were eluted with elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 250 mM imidazole, pH 8.0). Protein-containing fractions were pooled and re-buffered in 50 mM KP_i pH 7.5 using PD10 columns (GE Healthcare, Buckinghamshire, UK). PFE wild-type was expressed as previously reported[47].

Activity assays

The activity of the PFE variants towards the hydrolysis of *para*-nitrophenyl acetate (*p*NPA) was analyzed. For this purpose, 20 μL of a 100 mM *p*NPA solution in DMSO was added to a 96-well plate and 180 μL of a PFE solution of known concentration in 50 mM KPi pH 7.5 was added to start the reaction. The absorbance was followed at 405 nm and the initial slope was calculated. The reaction was carried out at 25°C. Autohydrolysis was determined by adding 180 μL of the 50 mM KPi pH 7.5 buffer and subtracting the value from the hydrolysis rate of the PFE mutants. Specific activity was calculated using a standard curve of *para*-nitrophenol (0-1 mM). The hydrolysis of *rac*-1-phenyl-2-pentyl acetate (**1**) and *rac*-1-phenylethyl acetate (**2**) was analyzed for all PFE mutants. For this purpose, 975 μL of a 90 μg/mL PFE solution was added to a 1.5 mL tube. The reaction was started by adding 25 μL of a 200 mM substrate solution in acetonitrile (final concentration 5 mM), and was run for 24 h at 37°C and 1000 rpm. Time samples of 200 μL were taken after 1, 2, 4, and 24 h and extracted with 200 μL of ethyl acetate twice. The organic phases were pooled, dried over anhydrous sodium sulfate, and analyzed by GC.
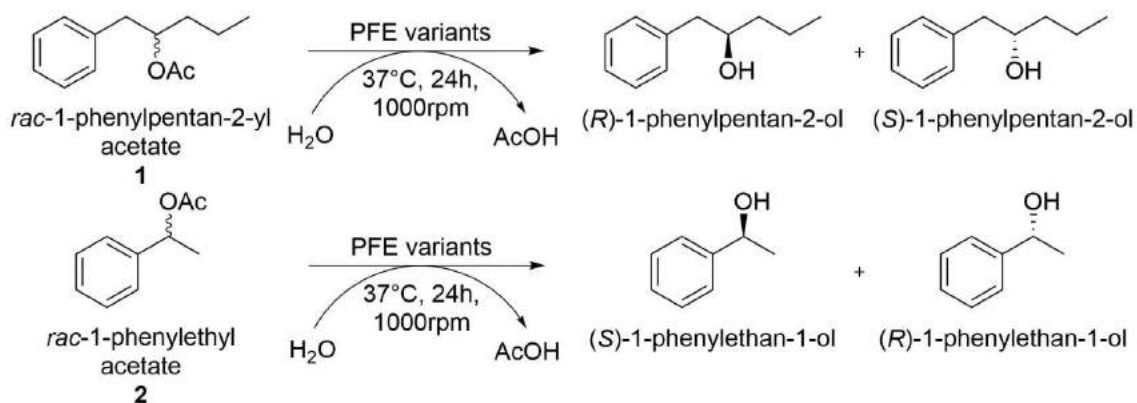
Scheme 1. Kinetic resolution of substrates **1** and **2** studied using PFE and its variants.

Gas chromatography (GC) analysis

Analysis was performed by gas chromatography with a flame ionization detector (GC-2010, Shimadzu, Kyoto, Japan) equipped with a Hydrodex β3P column (25.0 m × 0.25 mm, 0.25 µm film thickness, Macherey–Nagel, Düren, Germany). For the detection of the synthesis or hydrolysis of **1** column temperature was held at 95°C for 30 min, increased to 110°C with 5°C/min, and held for 45 min. Retention times: (*S*)-1-phenyl-2-pentyl acetate 50.5 min, (*R*)-1-phenyl-2-pentyl acetate 51.5 min, (*S*)-1-phenyl-2-pentanol 66.7 min, (*R*)-1-phenyl-2-pentanol 69.2 min. For the detection of the hydrolysis of **2** column temperature was held at 110°C for 30 min. Retention times: (*S*)-1-phenyl-ethyl acetate: 4.3 min, (*R*)-1-phenyl-ethyl acetate: 5.9 min, (*S*)-1-phenylethanol: 8.0 min, (*R*)-1-phenylethanol: 7.1 min.

Results and Discussion

Catalytic residues redesign

The esterase I from *Pseudomonas fluorescens* (PFE), with wide experimental validation [47–49], was chosen as the test scaffold protein. This enzyme hydrolyzes small aliphatic esters and its active site contains the typical Ser-His-Asp catalytic triad (Figure S1).

To test the performance of AsiteDesign in identifying optimum positions for placing the catalytic residues, the amino acids forming the esterase catalytic triad in the wild-type (WT) enzyme (Ser94, His251, and Asp222) were mutated to Ala. Using this mutated structure, an

MC search was performed, employing ethyl acetate as the probing substrate. In the simulation, all residues listed in Table S1 were allowed to be mutated to one of the residues of the catalytic triad. As mentioned in the method section, during the simulation once a new position is accepted for a given catalytic residue, the previous position of the catalytic residue is mutated back to the WT amino acid. Encouragingly, the catalytic residues of the WT enzyme were recovered as the best solution (Table S2). This result demonstrates that the protocol can identify the optimal positions for the catalytic residues.

In addition, the simulation resulted in three enzyme variants (PFE_1, PFE_1*+I155D, PFE_1*+A183D) in which both positions 28 and 29 are assigned to the Ser and His residues, respectively (Figure S1). The only difference resides in the location of the acid residue. It is interesting to highlight that, in these enzyme variants, the catalytic residues are the mirror image of the WT enzyme (Figure 1). Therefore, these variants are expected to exhibit opposite enantioselectivity relative to the WT enzyme.
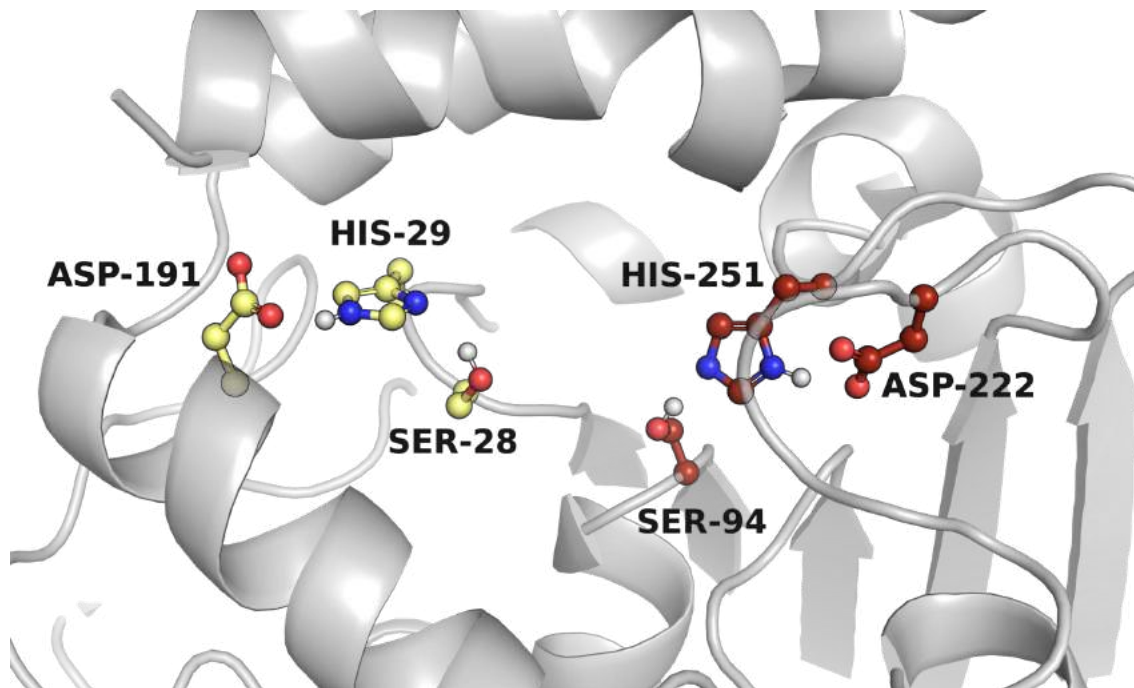

Figure 1. PFE WT and the newly designed active site. The catalytic triad residues of the WT are colored in red and labeled, while the ones from the PFE_1 design are also shown in yellow and labeled. The figure displays that the designed variants are the mirror image of the WT active site in the same protein cavity.

To test the stability of the predicted variants, 100 ns MD simulations were performed for each enzyme. These simulations indicated that the variants are less stable compared to the WT (Figure S2). The PFE_1 variant, however, exhibited acceptable catalytic distances (Figure 2 and S3). The analysis of the MD simulations suggests that the W28S mutation in PFE_1 results in a cavity that is occupied with water, thereby destabilizing the structure. To address this, compensatory mutations (C194T, V195M, and C194T/V195M) were introduced. These mutants, however, exhibited similar values for the metrics of the MD simulations.
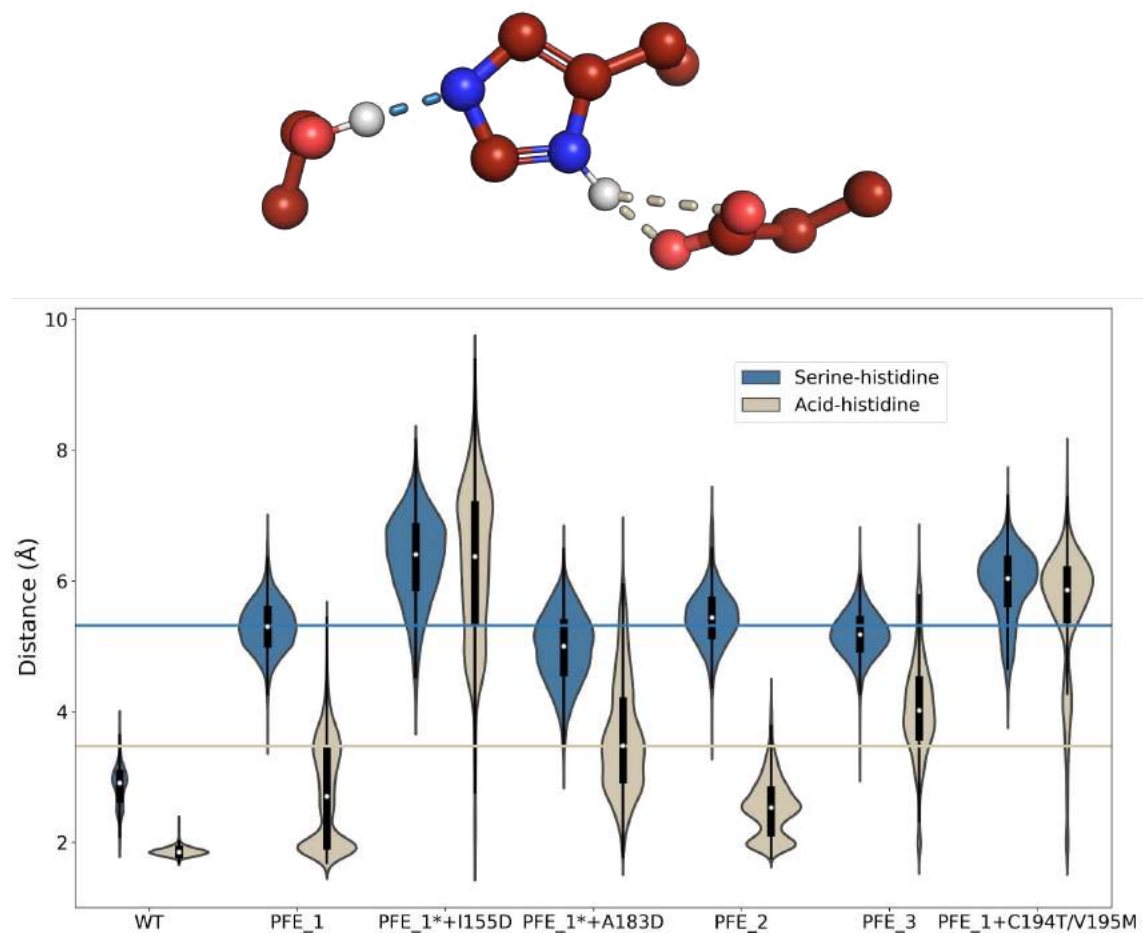
Figure 2. Distribution of the catalytic distances (serine–histidine distance, dHγSer–Nδ/εHis; acid–histidine distance, dOδ/εAsp/Glu–Hδ/εNHis) along the MD simulations. Four MD replicas of 100 ns were performed for both the WT and all selected mutants from the AsiteDesign pipeline. PFE_1* stands for W28S/L29H/S94A.

The computationally designed variants, recombinantly expressed in *E. coli* and purified, were then verified experimentally to characterize the enzymes' activities (Tables 1 and S2). In these variants, the native catalytic machinery was disabled by mutating the nucleophilic Ser94 to Ala. The experimental results showed that the PFE_1 variant is indeed active in the

hydrolysis of *p*NPA and the racemic compounds **1** and **2**. Its activity, however, was lower than the WT enzyme. This could be due to the destabilization effect of mutations, as can be seen from the decreased melting temperature of the designed variants. Alternatively, the lower observed activity could be because of less optimum catalytic distances (Figure 2) and less stability of the active site as it can be seen from the catalytic triad's local RMSD (Figure S2). These observations imply that this variant has a less organized catalytic geometry. This phenomenon has been observed in other designed (or natural) hydrolase active sites, where improving these distances gave better overall activities[30,50]. This was predicted by the MD simulations of the mutants. Interestingly, the PFE_1 variant exhibited an inverse selectivity for the bulky compound **1**. Variants PFE_2 and PFE_3 were both inactive toward **1** but showed activity toward *p*NPA (Table S4).

It was confirmed that the *p*NPA hydrolysis was not caused by potential imidazole impurities from purification. This was demonstrated by the hydrolysis of **2**, which is not hydrolyzable by imidazole and is a less challenging PFE substrate than the bulky substrate **1** (Table S4). Since all the purified PFE variants showed activity and selectivity in the hydrolysis of **2**, and autohydrolysis was not observed, the experimental data clearly confirm enzymatic hydrolysis.

Table 1. Experimental measured activities for the catalytic designs in the hydrolysis of substrate **1**. The activity is reported as conversion. The residue numbering corresponds to the 1VA4 structure. [a]: after 24 hours, [b]: after 1 hour.

PFE variants	Mutations	Substrate **1**	Predicted selectivity
WT	-	8.9 % (8 %*ee* (*R*), E 1)[a]	-
PFE_1	W28S/L29H/T191D/ S94A	2.3 % (13 %*ee* (*S*), E 1)[a]	(*S*)
PFE_2	W28S/L29H/T191D/ S94A/C194T	Not detectable	(*S*)
PFE_3	W28S/L29H/T191D/ S94A/V195M	Not detectable	(*S*)

E values were calculated according to Chen et al[51].

These results suggest that, for a given binding pocket, the protocol is able to identify multiple viable solutions for designing catalytic residues, which can be used as a starting point for further optimization.

Binding pocket redesign

To test the performance of AsiteDesign for non-catalytic residues, **1** was chosen as the substrate (Scheme 1). The WT enzyme exhibits low activity and enantioselectivity for **1**, which makes it a good candidate for improvement. Additionally, the previous site-directed mutagenesis of this enzyme did not yield any variants with high activity for **1**[47] . The binding pocket design simulations were performed by including 31 residues of the active site (Table S3; design domain, notice that catalytic residues were excluded). In these simulations, no assumptions were made for the positions of mutations and all residues are present in the first shell of the active site, 11 residues (Figure 3; highlighted in yellow), were allowed to mutate while the rest were only repacked. Since the enzyme was expected to hydrolyze a hydrophobic substrate, the allowed mutations were limited to hydrophobic amino acids (A, I, L, F, P, W, V, Y).
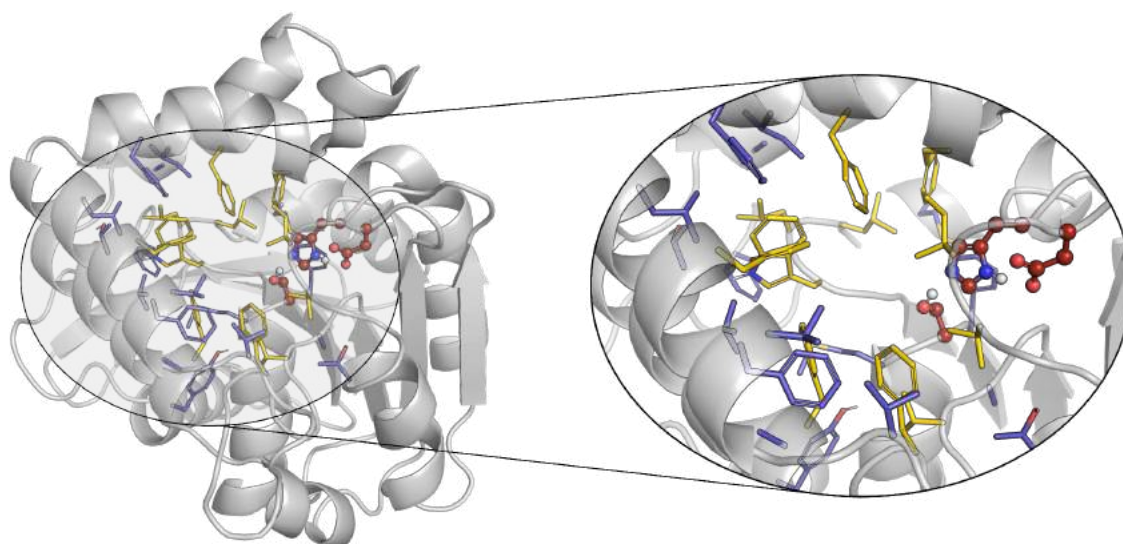
Figure 3. Active site of PFE and the used design domain. The catalytic triad residues are colored in red, the mutable residues in yellow, and the only repackable residues in violet.

In addition, sequence restraints were imposed on the active site residues (which correspond to the WT active site) to avoid large divergence from the WT enzyme, thereby favoring sequence conservation with an energy penalty. The substrate was placed in the active site manually and MC simulations were performed while imposing distance restraint between the carbonyl carbon of substrate and Ser94.

Two separate MC simulations were performed for the (*R*)- and (*S*)-enantiomers of **1**, thereby evolving the active site for each enantiomer independently. For each enantiomer, the 50 variants with the overall best energies (protein and substrate binding) were selected. These

structures were then clustered based on the binding mode of the substrate and from each cluster one variant was chosen (Table S5). Encouragingly, the simulations targeted many of the active site positions that were previously suggested to be important for enantioselectivity (F125, F158, and I224)[47] in addition to some new positions (W28, V121, and F198). However, the predicted mutations for these positions may differ from the previous study.

To test the predicted variants activities, substrate binding was simulated by PELE software[42]. In these simulations, the substrate was placed outside the active site and the binding was monitored by counting catalytic events and computing the average energy of the ligand in the active site (Figure 4). A catalytic event is defined as a conformation where the distance between the carbonyl C of the substrate and the alcoholic O of the catalytic Ser residue is within 4 $\text{\AA}$, while the H-bonds of the catalytic triad are within reasonable distances (<= 3.5 $\text{\AA}$)[31]. Also, the distance between the catalytic His residue and the ether O of the substrate is less than 6.5 $\text{\AA}$ (as the protonated His residue will give a proton to this atom during the reaction, to release the alcohol product[52]). Overall, the predicted variants exhibited a higher number of catalytic events compared to the WT (Table S5). Moreover, the distribution of the average interaction energy (Figure S4) is better in many variants compared to the WT, and key catalytic distances have good values as well (Figure S5-8).
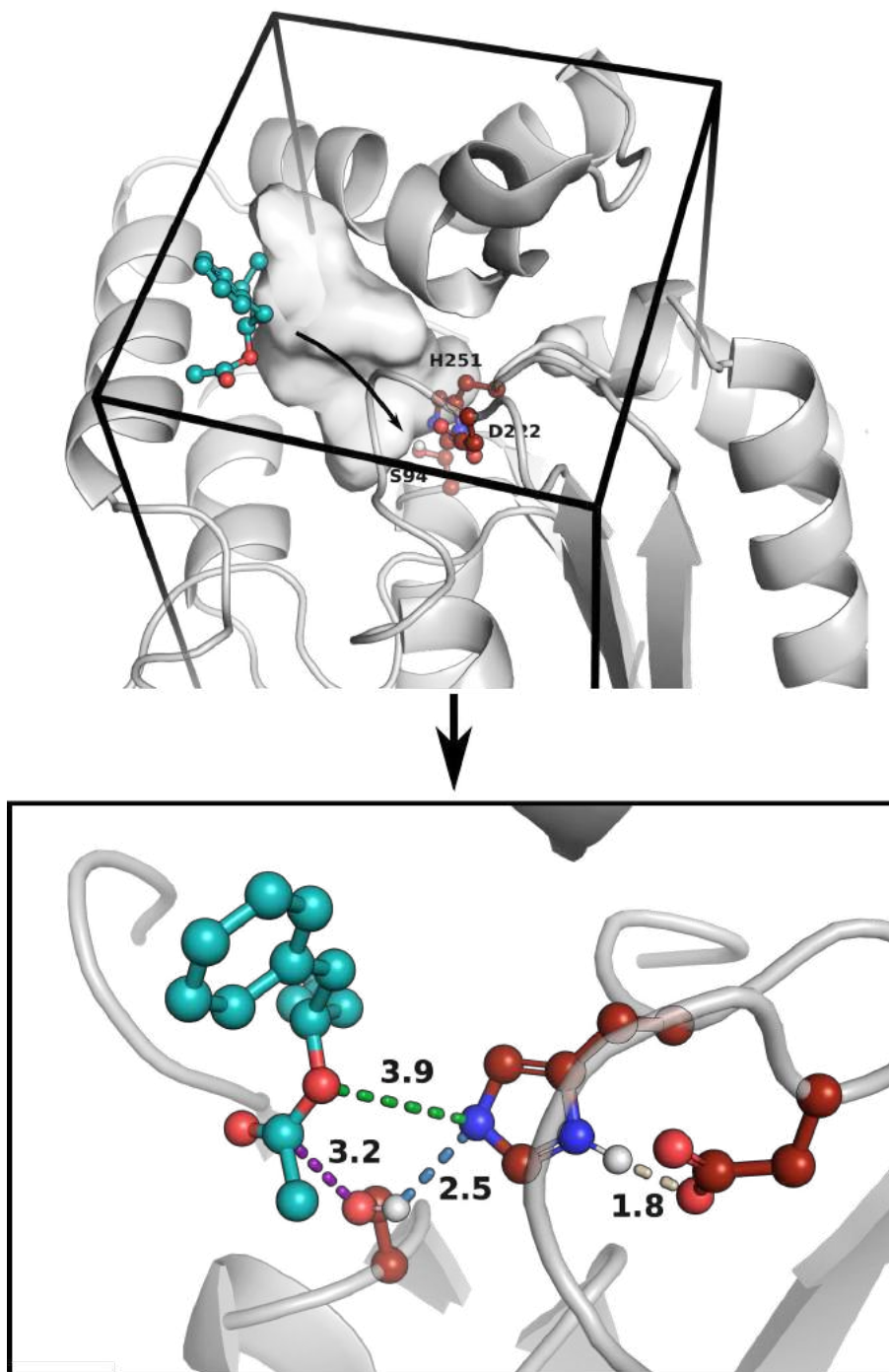
Figure 4. Initial setup for PELE simulations and the representation of a catalytic event. The substrate is placed outside the active site and allowed to explore around the drawn box (top). The catalytic event is represented with every key distance highlighted in a different color (blue for serine-histidine, beige for acid-histidine, violet for serine-substrate, and green for histidine-substrate) (bottom). The catalytic triad residues are colored in red, and the substrate in cyan.

Table 2. Experimental measured activities for the binding pocket redesigns in the hydrolysis substrate **1**. The activity is reported as conversion. The residue numbering corresponds to 1VA4 structure. [a]: after 24 hours, [b]: after 1 hour, [c]: after 2 hours.

PFE variants	Mutations	Substrate **1**	Predicted selectivity
WT	-	8.9 % (8 %*ee* (*R*), E 1)[a]	-
PFE_4	W28A/F158L/F198A	6.0 % (3 %*ee* (*S*), E 1)[a]	-
PFE_5	F158L/F198A	66.7 % (32 %*ee* (*S*), E 4)[a]	(*S*)
PFE_6	W28A/F125A/F158L/F198A	3.3 % (9 %*ee* (*S*))[a]	(*S*)
PFE_7	W28A/F158L/F198A/I224L	5.2 % (8 %*ee* (*R*), E 1)[a]	(*S*)
PFE_8	F125A/F158L	23.4 % (55 %*ee* (*S*), E 4)[a]	(*R*)
PFE_9	F125A/F158L/I224L	1.4 % (29 %*ee* (*R*))[a]	(*R*)
PFE_10	F125A/F158L/F198A	16.3 % (60 %*ee* (*S*), E 4)[a]	(*R*)
PFE_11	V121A/F125A/I224L	2.9 % (38 %*ee* (*S*))[a]	(*R*)
PFE_12	V121A/F158A/F198V	1.6 % (100 %*ee* (*S*))[a]	(*R*)

Based on the *in silico* analysis, 8 variants with the highest number of catalytic events relative to the WT enzyme, were selected for experimental verification. The experimental results show that three predicted variants (PFE_5, PFE_8, and PFE_10) exhibited significant improvement over the WT enzyme in the hydrolysis of **1** and, in contrast to the WT enzyme, they are selective for the (*S*) substrate (Table 2). These variants contain F158L, F125A, and F198A substitutions. Mutation of these bulky residues to smaller hydrophobic ones opens up the active site to accommodate the bulky substrates (Figure 5).

These results demonstrate that the design protocol is able to identify the hot spots in the active site and propose mutations that can potentially improve the enzyme activity for a

given substrate. However, the simulations were not able to predict the variants enantioselectivity accurately. The main reason for this could be that the design/predictions were performed based on ligand binding energies, which does not necessarily correlate with enantioselectivity. This can potentially be improved by incorporating a transition state analog as the probing substrate, which is the main deciding factor for enantioselectivity. Nevertheless, an accurate prediction of the enantioselectivity is notoriously challenging as the energy difference between the activation energies of enantiomers are often very small.

The binding pocket designed variants also exhibited lower melting temperatures. The main reason for this is that the simulations are driven by improving the ligand binding energies, at the expense of protein stability. This issue can be alleviated by downstream enzyme stability optimization of the designed variants either computationally or experimentally[6–10].
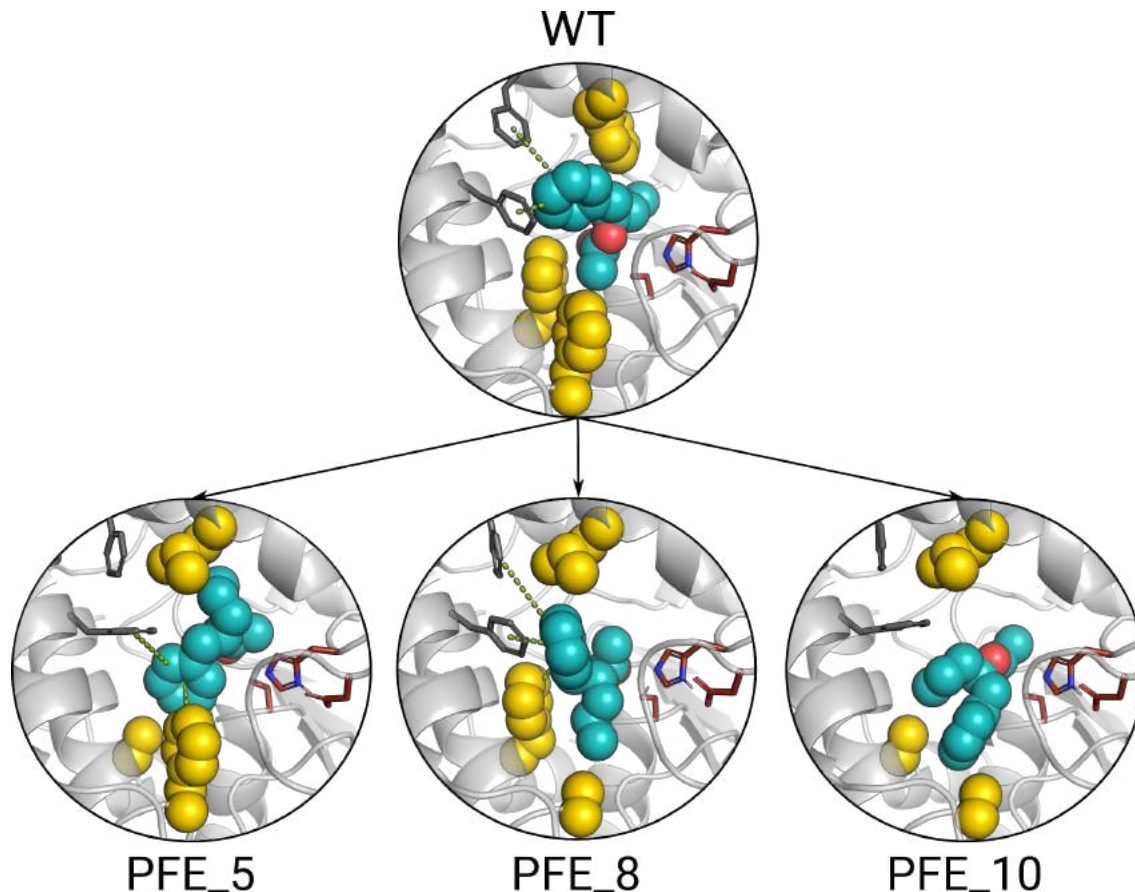

Figure 5. Representative catalytic pose of the WT enzyme and the successful *in silico* evolved variants. The catalytic triad residues are colored in red, the substrate in cyan, and the mutated residues in yellow. Possible pi-pi interactions between the substrate and Phe residues are shown with a dashed green line.

Conclusions

This work presents the AsiteDesign protocol, which aims at engineering active sites of enzymes to either introduce new catalytic residues or to modify an existing active site *in silico*. The protocol is implemented using the pyRosetta library, and it combines MC sampling of the active site residues with enhanced sampling techniques to identify the most suitable positions for a given active site. The ligand sampling is also included in the simulation, which is necessary to identify the optimal solutions.

To demonstrate the performance of the protocol, a new catalytic triad was designed in the active site of the esterase I from *Pseudomonas fluorescens* (PFE). The experimental characterization demonstrated that the designed variants are not only active biocatalysts, but they also exhibited inverse enantioselectivity for the bulky chiral substrate **1**. Thus, the binding pocket of the enzyme was also successfully engineered to improve the activity for **1**.

Overall, these examples demonstrate that the AsiteDesign protocol is able to identify multiple viable solutions for designing active site residues for a given active site. This approach, thus, can be used in the engineering of multifunctional catalysts or designing new catalytic residues in a given putative binding pocket.

Code availability

The code used in the study is available at https://github.com/masoudk/AsiteDesign

References

1. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).

2. Yi, D. *et al.* Recent trends in biocatalysis. *Chem. Soc. Rev.* **50**, 8003–8049 (2021).

3. Lovelock, S. L. *et al.* The road to fully programmable protein catalysis. *Nature* **606**, 49–58 (2022).

4. Bell, E. L. *et al.* Biocatalysis. *Nat Rev Methods Primers* **1**, (2021).

5. Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: enzymatic synthesis for industrial applications. *Angew. Chem. Int. Ed Engl.* **60**, 88–119 (2021).

6. Wijma, H. J. *et al.* Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27**, 49–58 (2014).

7. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–8 (2005).

8. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).

9. Marques, S. M., Planas-Iglesias, J. & Damborsky, J. Web-based tools for computational enzyme design. *Curr. Opin. Struct. Biol.* **69**, 19–34 (2021).

10. Kulshreshtha, S., Chaudhary, V., Goswami, G. K. & Mathur, N. Computational approaches for predicting mutant protein stability. *J. Comput. Aided Mol. Des.* **30**, 401–412 (2016).

11. Musil, M., Konegger, H., Hon, J., Bednar, D. & Damborsky, J. Computational design of stable and soluble biocatalysts. *ACS Catal.* **9**, 1033–1054 (2019).

12. Goldenzweig, A. & Fleishman, S. J. Principles of protein stability and their application in computational design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).

13. Shirke, A. N. *et al.* Toward rational thermostabilization of *Aspergillus oryzae* cutinase: Insights into catalytic and structural stability. *Proteins* **84**, 60–72 (2016).

14. Nguyen, V. *et al.* Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289–294 (2017).

15. Bednar, D. *et al.* FireProt: energy- and evolution-based computational design of thermostable multiple-point mutants. *PLoS Comput. Biol.* **11**, e1004556 (2015).

16. Fassio, A. V., Santos, L. H., Silveira, S. A., Ferreira, R. S. & de Melo-Minardi, R. C. nAPOLI: a graph-based strategy to detect and visualize conserved protein-ligand interactions in large-scale. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17**, 1317–1328 (2020).

17. Hon, J. *et al.* EnzymeMiner: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities. *Nucleic Acids Res.* **48**, W104–W109 (2020).

18. Sumbalova, L., Stourac, J., Martinek, T., Bednar, D. & Damborsky, J. HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res.* **46**, W356–W362 (2018).

19. Pavelka, A. *et al.* CAVER: algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **13**, 505–517 (2016).

20. Musil, M. *et al.* FireProtASR: a web server for fully automated ancestral sequence reconstruction. *Brief. Bioinform.* **22**, (2021).

21. Planas-Iglesias, J. *et al.* LoopGrafter: a web tool for transplanting dynamical loops for protein engineering. *Nucleic Acids Res.* (2022) doi:10.1093/nar/gkac249.

22. Karami, Y. *et al.* DaReUS-Loop: a web server to model multiple loops in homology models. *Nucleic Acids Res.* **47**, W423–W428 (2019).

23. Khersonsky, O. *et al.* Automated design of efficient and functionally diverse enzyme repertoires. *Mol. Cell* **72**, 178–186.e5 (2018).

24. Anishchenko, I. *et al.* De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).

25. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

26. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

27. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).

28. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* eadd2187 (2022).

29. Roda, S., Santiago, G. & Guallar, V. Mapping enzyme-substrate interactions: its potential to study the mechanism of enzymes. *Adv. Protein Chem. Struct. Biol.* **122**, 1–31 (2020).

30. Alonso, S. *et al.* Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. *Nat. Catal.* **3**, 319–328 (2020).

31. Roda, S., Robles-Martín, A., Xiang, R., Kazemi, M. & Guallar, V. Structural-based modeling in protein engineering. A must do. *J. Phys. Chem. B* **125**, 6491–6500 (2021).

32. Roda, S. *et al.* A plurizyme with transaminase and hydrolase activity catalyzes cascade reactions. *Angew. Chem. Int. Ed Engl.* (2022) doi:10.1002/anie.202207344.

33. Mayer, C., Zechel, D. L., Reid, S. P., Warren, R. A. & Withers, S. G. The E358S mutant of Agrobacterium sp. beta-glucosidase is a greatly improved glycosynthase. *FEBS Lett.* **466**, 40–44 (2000).

34. Jahn, M., Marles, J., Warren, R. A. J. & Withers, S. G. Thioglycoligases: mutant glycosidases for thioglycoside synthesis. *Angew. Chem. Int. Ed Engl.* **42**, 352–354 (2003).

35. Chen, K. & Arnold, F. H. Engineering new catalytic activities in enzymes. *Nat. Catal.* **3**, 203–213 (2020).

36. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).

37. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De novo enzyme design using Rosetta3. *PLoS One* **6**, e19230 (2011).

38. Khalameyzer, V., Fischer, I., Bornscheuer, U. T. & Altenbuchner, J. Screening,

nucleotide sequence, and biochemical characterization of an esterase from *Pseudomonas fluorescens* with high activity towards lactones. *Appl. Environ. Microbiol.* **65**, 477–482 (1999).

39. Lecina, D., Gilabert, J. F. & Guallar, V. Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Rep.* **7**, 8466 (2017).

40. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).

41. Borrelli, K. W., Vitalis, A., Alcantara, R. & Guallar, V. PELE: protein energy landscape exploration. A novel Monte Carlo based technique. *J. Chem. Theory Comput.* **1**, 1304–1311 (2005).

42. Municoy, M., Roda, S., Soler, D., Soutullo, A. & Guallar, V. aquaPELE: a Monte Carlo-based algorithm to sample the effects of buried water molecules in proteins. *J. Chem. Theory Comput.* **16**, 7655–7670 (2020).

43. Atilgan, A. R. *et al.* Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **80**, 505–515 (2001).

44. Banks, J. L. *et al.* Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.* **26**, 1752–1780 (2005).

45. Bashford, D. & Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129–152 (2000).

46. Gilabert, J. F. *et al.* PELE-MSM: A Monte Carlo based protocol for the estimation of absolute binding free energies. *J. Chem. Theory Comput.* **15**, 6243–6253 (2019).

47. Schliessmann, A., Hidalgo, A., Berenguer, J. & Bornscheuer, U. T. Increased enantioselectivity by engineering bottleneck mutants in an esterase from *Pseudomonas fluorescens*. *ChemBioChem* **10**, 2920–2923 (2009).

48. Schmidt, M. *et al.* 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* **7**, 805–809 (2006).

49. Ding, Q. & Kazlauskas, R. J. Improving *Pseudomonas fluorescens* esterase for

hydrolysis of lactones. *Catal. Sci. Technol.* **7**, 4756–4765 (2017).

50. Roda, S. *et al.* Computationally driven rational design of substrate promiscuity on Serine ester hydrolases. *ACS Catal.* **11**, 3590–3601 (2021).

51. Chen, C. S., Fujimoto, Y., Girdaukas, G. & Sih, C. J. Quantitative analyses of biochemical kinetic resolutions of enantiomers. *J. Am. Chem. Soc.* **104**, 7294–7299 (1982).

52. Rauwerdink, A. & Kazlauskas, R. J. How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of α/β-hydrolase fold enzymes. *ACS Catal.* **5**, 6153–6176 (2015).

Supporting Information

Table S1. Recompilation of all the residues in the design domain of the catalytic residues redesign's experiment. The residue numbering corresponds to 1VA4 structure.

Residue number
27
28
29
30
34
57
69
93
94
95
120
121
125
135

139
140
143
147
154
155
158
162
183
191
195
198
199
222
224
225
230
251

Table S2. Top 10 catalytic designs given at the end of the ASiteDesign simulation with PFE's structure (where the original catalytic residues have been mutated to ALA). The variants are ranked according to the total energy.

Total Energy	Mutations
-3170.1	A94S/A251H/A222D (WT)
-3160.1	W28S/L29H/T191D (PFE_1)
-3159.6	A94S/V225H/A222D
-3157.8	A94S/A251H/F162D
-3156.8	A94S/A251H/I224D
-3156.1	W28S/L29H/I155D (PFE_1*+I155D)
-3150.8	A94S/V225H/F125D

-3150.1	W28S/L29H/A183D (PFE_1*+A183D)
-3148.5	W28S/V195H/T191D
-3147.7	W28S/M95H/V121D

Table S3. Recompilation of all the residues in the design domain of the binding redesign pocket's experiment. The residue numbering corresponds to 1VA4 structure.

Residue number	Type of alteration along simulation
28	Mutable
29	Mutable
30	Mutable
34	Only repackable
57	Only repackable
69	Only repackable
93	Only repackable
94	Only repackable
95	Only repackable
120	Only repackable
121	Mutable
125	Mutable
135	Only repackable
139	Only repackable
140	Only repackable
143	Mutable
147	Only repackable
154	Only repackable
155	Only repackable
158	Mutable

162	Mutable
183	Only repackable
191	Only repackable
195	Only repackable
198	Mutable
199	Only repackable
222	Only repackable
224	Mutable
225	Mutable
230	Only repackable
251	Only repackable

Table S4. Experimental measured activities for the catalytic designs with the other tested substrates; pNPA and **2**. pNPA activity is recorded according to specific activity, while for substrate **2** the conversion is reported. The melting points (T_m) of the PFE variants were determined by NanoDSF. The residue numbering corresponds to the 1VA4 structure. [a]: after 24 hours, [b]: after 1 hour.

PFE variants	Mutations	T_m [°C]	pNPA activity [U/mg]	Substrate **2**
WT	-	71.8	162.2	48.7 % (91 %ee (R), E 59)[b]
PFE_1	W28S/L29H/T191D/S94A	44.9	0.9	17.7 % (80 %ee (R), E 11)[a]
PFE_2	W28S/L29H/T191D/S94A/C194T	50.9	0.2	2.1 % (43 %ee (R), E 3)[a]
PFE_3	W28S/L29H/T191D/S94A/V195M	44.8	1.3	0.9 % (11 %ee (R), E 1)[a]

Table S5. PELE simulation results of selected mutants with **1**. Some variants were selected based on two structures by combining the mutations. The residue numbering corresponds to the 1VA4 structure.

| Probing substrate	Mutations	Structure	Catalytic events (*S*-substrate)	Average Interaction Energy (*S*-substrate) [kcal/mol]	Catalytic events (*R*-substrate)	Average Interaction Energy (*R*-substrate) [kcal/mol]
(*R*,*S*)-1	-	1VA4	0	-29.8	24	-34.4
(*S*)-1	F158L/F198A	PFE_5	1849	-33.8	1291	-33.0
(*S*)-1	W28A/F125A/ F158L/F198A	PFE_6	2731	-32.1	1721	-32.5
(*S*)-1	W28A/F158L/F 198A/I224L	PFE_7	3497	-32.3	2916	-32.9
(*R*)-1	F125A/F158L	PFE_8	0	-	3155	-33.1
(*R*)-1	F125A/F158L/I 224L	PFE_9	15	-28.1	810	-31.1
(*R*)-1	F125A/F158L/ F198A	PFE_10	1079	-31.6	1775	-33.2
(*R*)-1	V121A/F125A/I 224L	PFE_11	820	-33.0	989	-34.0
(*R*)-1	V121A/F158A/ F198V	PFE_12	156	-32.2	1153	-34.7
(*S*)-1	F143A/F198A	PFE_13	491	-32.9	2003	-33.6
(*S*)-1	V121A/F125A/ F198I	PFE_14	2525	-31.8	2835	-32.1
(*S*)-1	W28A/V121A/ F158L	PFE_15	17	-31.3	746	-34.0
(*R*)-1	W28A/F158L/I 224L	PFE_16	3684	-31.5	2484	-33.0
(*R*)-1	V121A/F198A/ V225A	PFE_17	274	-31.9	164	-28.9

Table S6. Experimental measured activities for the binding pocket redesigns with the other tested substrates; *p*NPA and **2**. *p*NPA activity is recorded according to specific activity, while for substrate **2** the conversion is reported. The melting points (T_m) of the PFE variants were determined by NanoDSF. The residue numbering corresponds to the 1VA4 structure. [a]: after 24 hours, [b]: after 1 hour, [c]: after 2 hours.

PFE variants	Mutations	T_m [ºC]	pNPA activity [U/mg]	Substrate **2**
WT	-	71.8	162.2	48.7 % (91 %ee (R), E 59)[b]
PFE_4	W28A/F158L/F198A	57.0	19.1	47.4 % (95 %ee (R), E 115)[a]
PFE_5	F158L/F198A	69.5	79.0	43.8 % (89 %ee (R), E 36)[b]
PFE_6	W28A/F125A/F158L/F198A	58.2	1.3	2.0 % (50 %ee (R), E 3)[a]
PFE_7	W28A/F158L/F198A/I224L	57.3	1.7	10.6 % (74 %ee (R), E 7)[a]
PFE_8	F125A/F158L	63.6	162.6	48.4 % (95 %ee (R), E 118)[c]
PFE_9	F125A/F158L/I224L	62.0	83.4	47.6 % (98 %ee (R), E >200)[c]
PFE_10	F125A/F158L/F198A	63.6	2.9	39.6 % (96 %ee (R), E 94)[c]
PFE_11	V121A/F125A/I224L	59.7	13.4	31.6 % (71 %ee (R), E 8)[a]
PFE_12	V121A/F158A/F198V	60.4	26.9	13.1 % (88 %ee (R), E 18)[a]

Table S7. Primers used for the construction of the PFE_2 (C194T_fw/rv; from PFE_1), PFE_3 (V195M_fw/rv; from PFE_1), PFE_5 (A28W_fw/rv; from PFE_4), PFE_6 (F125A_fw/rv; from PFE_4), PFE_7 (I224L_fw/rv; from PFE_4), PFE_9 (I224L_fw/rv; from PFE_8), PFE_10 (F198A_fw/rv; from PFE_8).

Primer	Sequence (5'→3')	T_a [ºC]
C194T_fw	CGATGTTGATaccGTGACCGCCTTCG	63
C194T_rv	GCTTTCAGGCTTGCCAGC	63
V195M_fw	TGTTGATTGTatgACCGCCTTCG	66
V195M_rv	TCGGCTTTCAGGCTTGCC	66
A28W_fw	CAGCCATGGTtggCTGCTGGATG	61

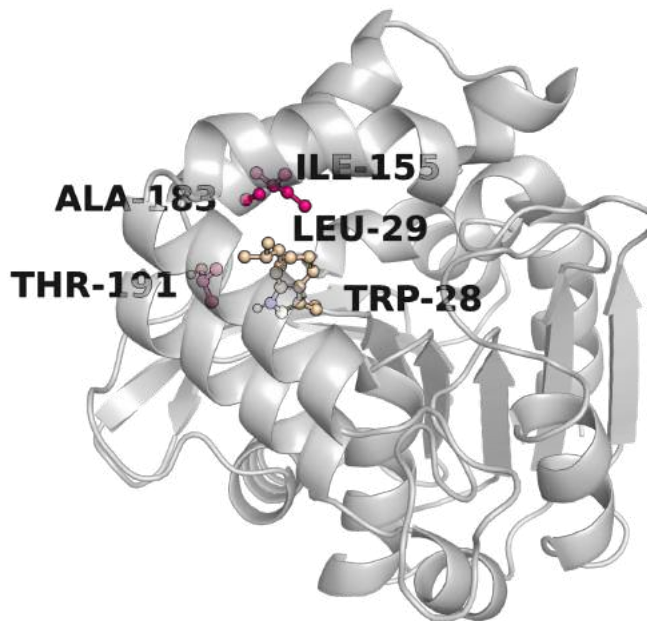|		
A28W_rv	AACAGCACCGGTTTACCA	61
F125A_fw	GACCCCGCTGgcaGGCCAGAAACCG	68
F125A_rv	ACGGCACCCAGCAGAACC	68
I224L_fw	TGGCGATCAGctgGTGCCGTTCG	65
I224L_rv	TCGCCATGAATCACCAGG	65
F198A_fw	CGTTACCGCCgcaGCCGAAACCG	62
F198A_rv	CAATCCACGGTTGCTTTC	62
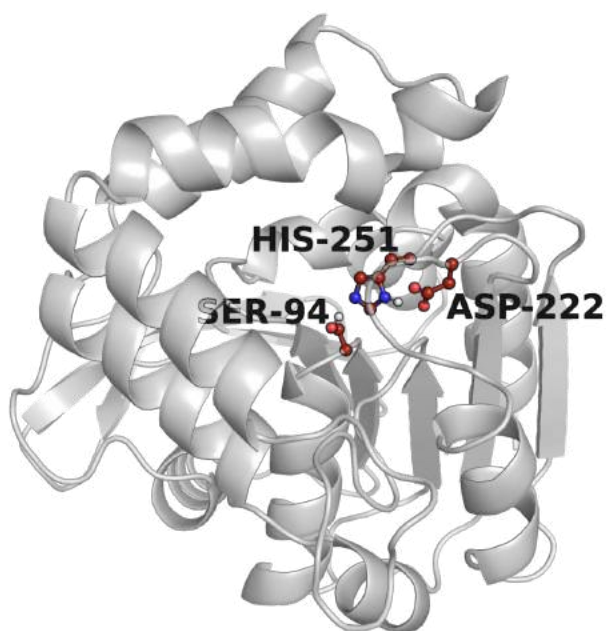
Figure S1. PFE and its catalytic residues (top). The catalytic triad residues are colored in red and labeled (PDB code: 1VA4). Potential positions for the design of an alternative catalytic triad (bottom). The amino acid positions Trp28 and Leu29, which are mutated to Ser and His, respectively, in the three variants PFE_1, PFE_1*+I155D, and PFE_1*+A183D, are colored beige. The introduced Asp, whose location differs in PFE_1, PFE_1*+I155D, and PFE_1*+A183D, is shown in pink.
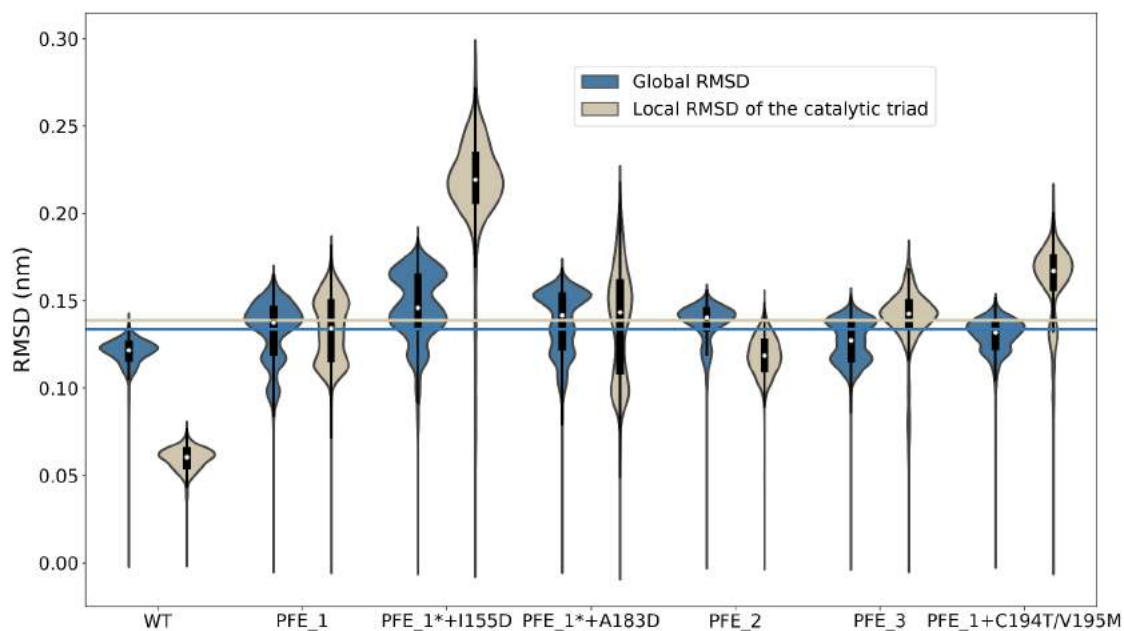
Figure S2. Distribution of the RMSD of the whole protein and its catalytic triad along the MD simulations. Four MD replicas of 100 ns were performed for both the WT and all selected mutants from the AsiteDesign pipeline. PFE_1* stands for W28S/L29H/S94A.

Figure S3. Distribution of the key distances between the substrate and the active site along the MD simulations. Four MD replicas of 100 ns were performed for both the WT and all selected mutants from the AsiteDesign pipeline. PFE_1* stands for W28S/L29H/S94A.
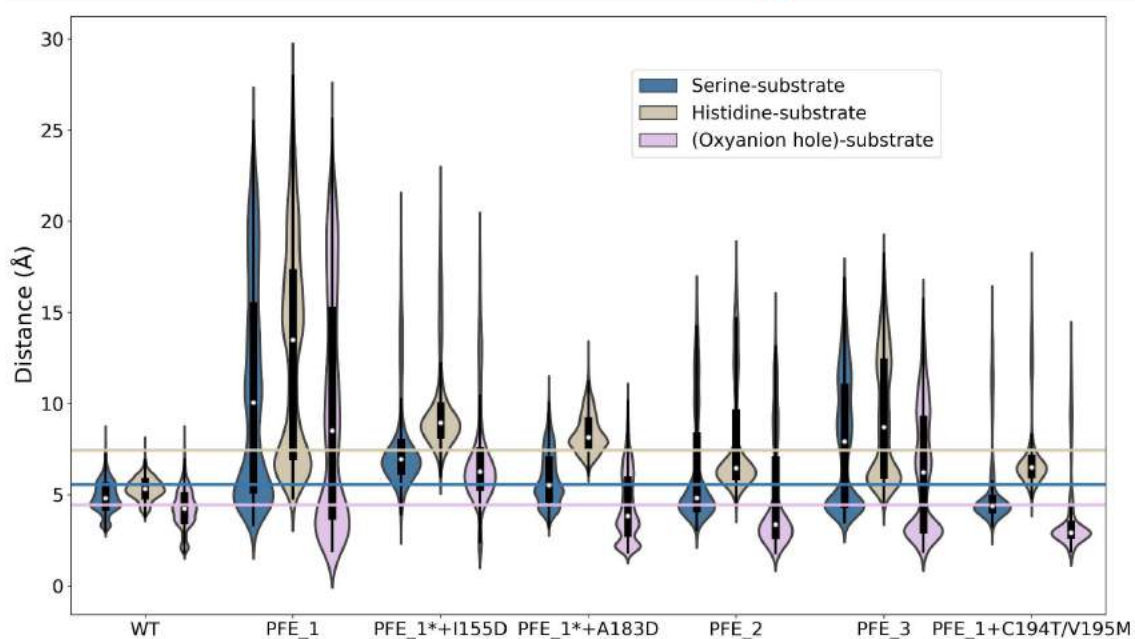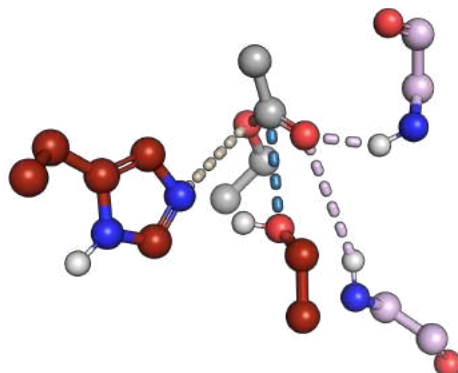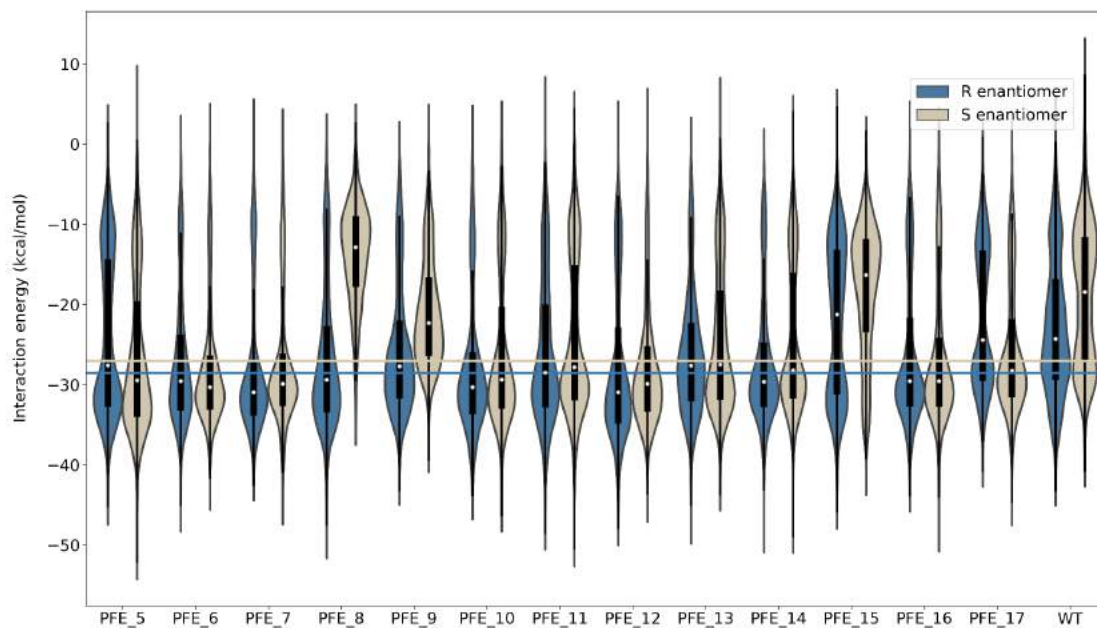
Figure S4. Distribution of the interaction energy between the substrate and the enzyme along the PELE simulations of the WT and all assayed variants.
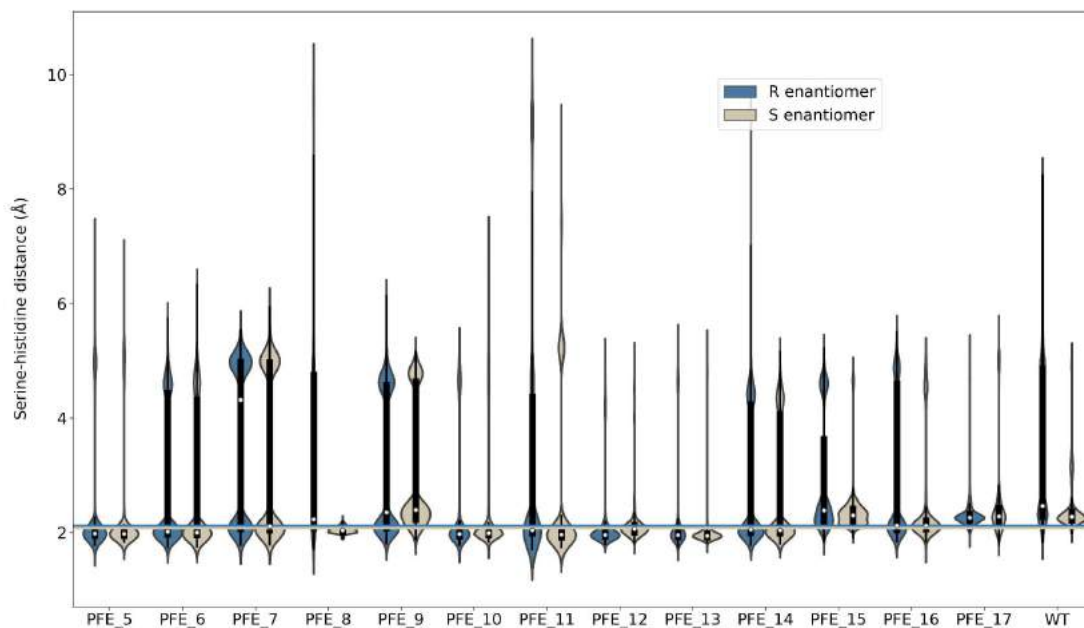

Figure S5. Distribution of the serine-histidine distance along the PELE simulations of the WT and all assayed variants.
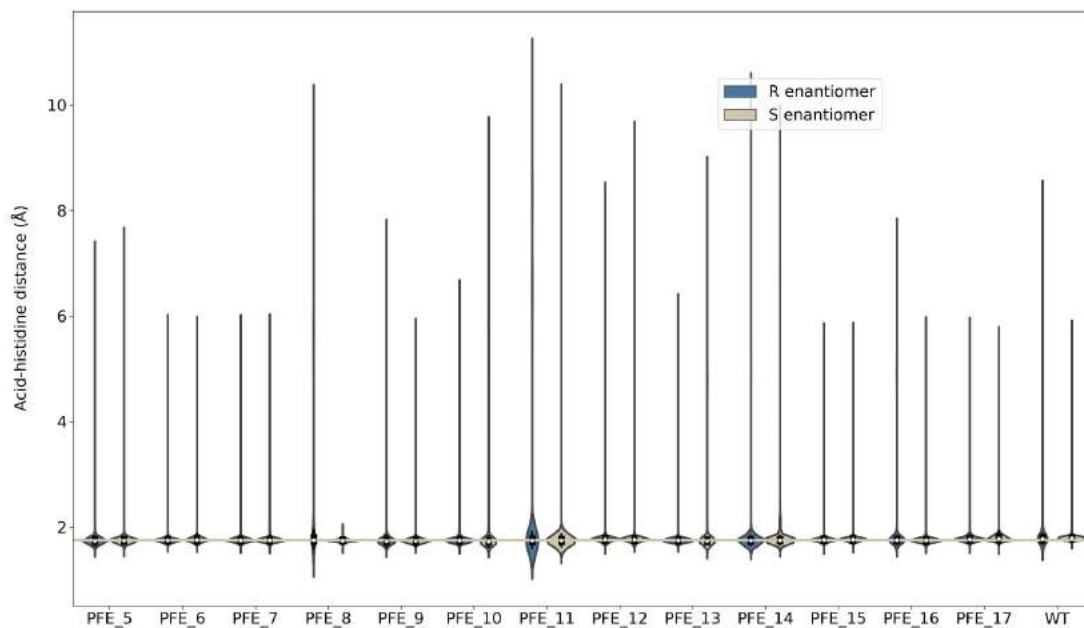
Figure S6. Distribution of the acid-histidine distance along the PELE simulations of the WT and all assayed variants.
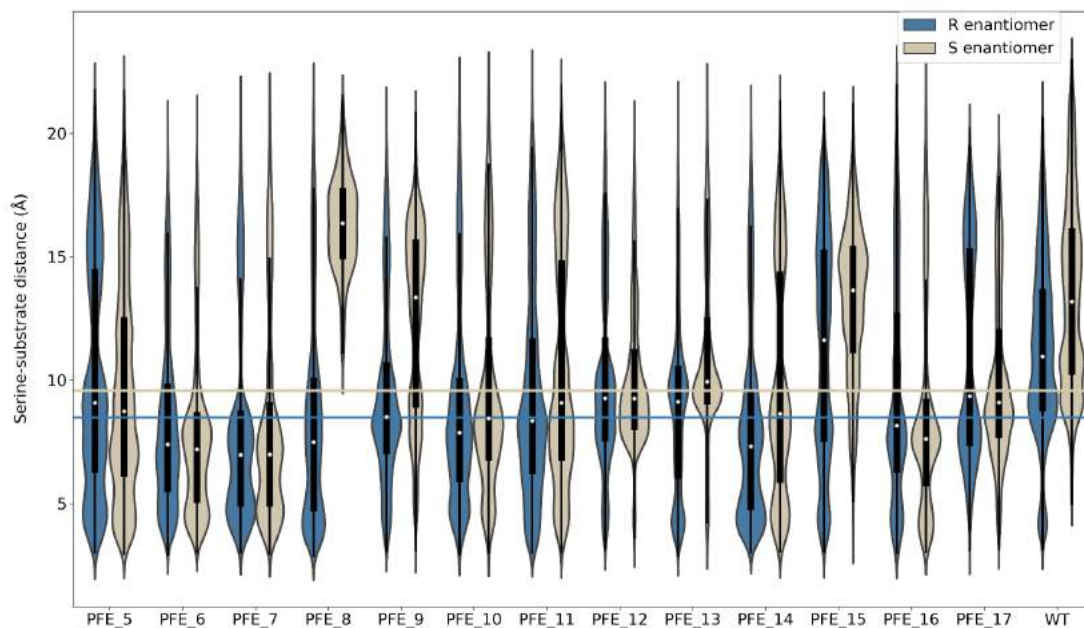

Figure S7. Distribution of the serine-substrate distance along the PELE simulations of the WT and all assayed variants.
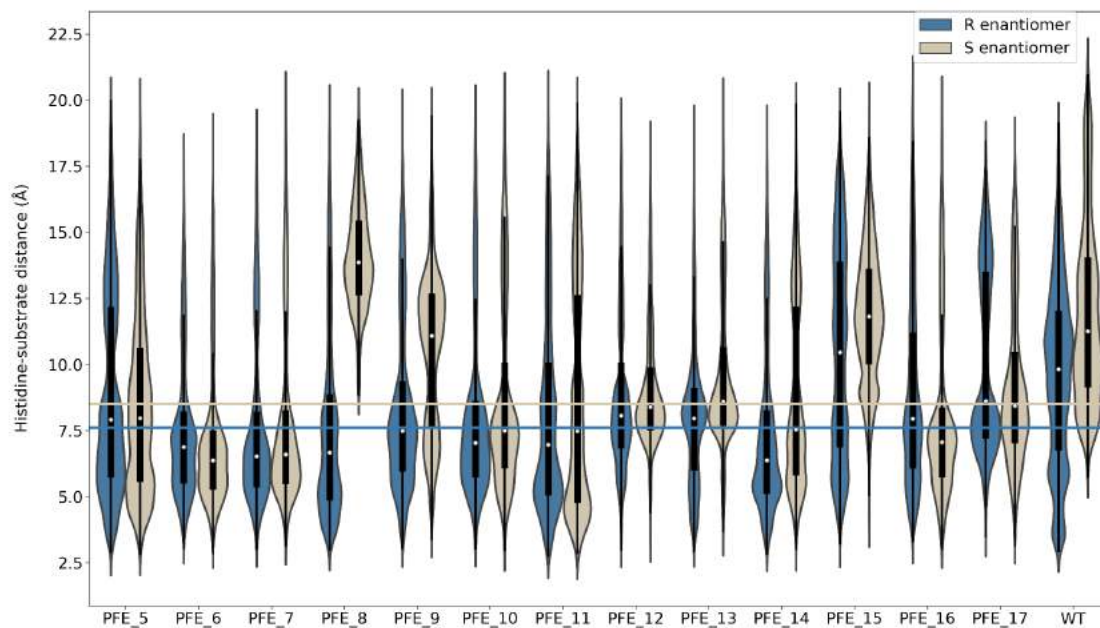
Figure S8. Distribution of the histidine-substrate distance along the PELE simulations of the WT and all assayed variants.

4 | Discussion

This chapter tries to easily describe my computational work of the articles presented in *Chapter 3*. The experimental work of the shown research has been performed by others who deserve credit for it, meaning I will only shortly report the outcome. Thus, I will go one by one, explaining my contributions to each paper.

4.1 Design of a pluriZyme with more than two active sites

As mentioned in the first goal of *Chapter 2*, we wanted to create an enzyme with 3 biological active sites or more. As a starting point, the pluriZyme created from a serine esterase extracted from a metagenomic sample of Lake Arreo was used (PDB ID: 6RB0) (Martínez-Martínez et al., 2013; Alonso et al., 2019). First, we performed a global exploration of the protein surface to find alternative binding sites using PELE (Kenneth W. Borrelli et al., 2005; Diáz et al., 2020). The used substrate was glyceryl tripropionate, which contains three ester bonds and is a widely standard ester compound. These alternative binding sites are local minima where the catalytic residues or the functional site are not present (*Figure 4.1*).

Interestingly, an alternative binding site was found next to the cavity of the artificial active site (named site II). To ensure that the active sites would not compromise each other, the simulations were run with the crystal structure that has the suicide inhibitor molecules bound to the catalytic Ser residues of both existing active sites. These results showcased that the cavity of site II was quite big (it could accommodate both the suicide inhibitor and a glyceryl tripropionate molecule), and thus, we imagined this pocket could fit peptides.
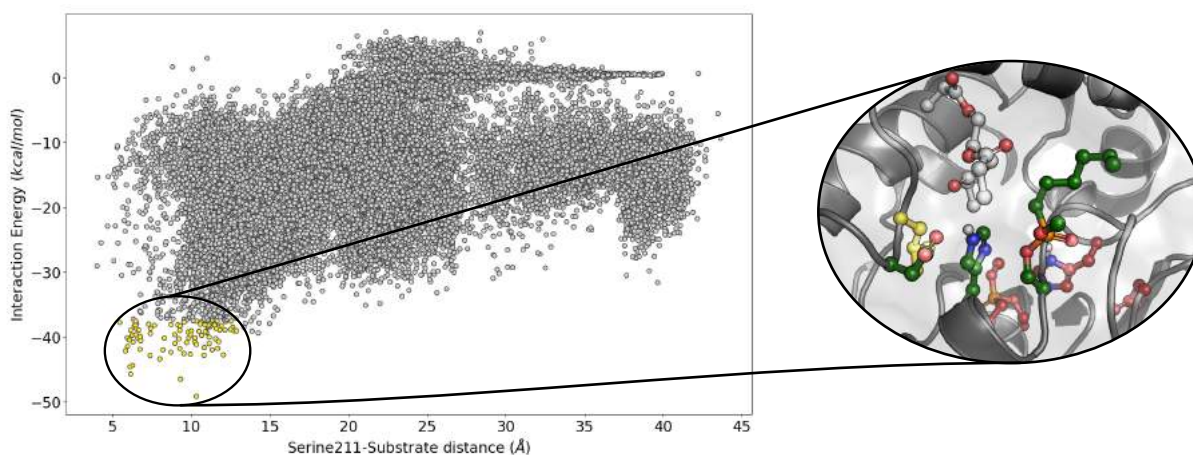
FIGURE 4.1: Representative plot of a global exploration of alternative
ester binding sites in the pluriZyme (PDB ID: 6RB0) structure with the
bound suicide inhibitors. Accepted PELE steps are depicted as a dot in the
scatter plot. The X axis refers to the distance to site II (distance between
the ester C atoms of the substrate and the nucleophilic O of the catalytic
Ser residue) and the Y axis to the interaction energy. The steps exploring
the new found site are highlighted in yellow color. A representative pose
of the alternative binding site is depicted on the right (main active site C
 atoms are stained in maroon and the artificial active site in dark green).

For this reason, we decided to add a cysteine-histidine dyad, which facilitates hydrolysis
of peptide bonds. Also, this approach allowed us to both create a new type of active site,
as well as reuse the histidine residue of site II to create the catalytic dyad. The candidate
residue to change for a cysteine was Leu24 due to its proximity to the catalytic histidine
residue of site II and its location in the cavity (*Figure 4.2*).

Then, several dipeptides (AH, AQ, DI, EA, FF, KA, LA, LL, NV, PF, QQ, RG, SW, TM,
YN, and YY) of different nature were prepared, and local PELE explorations of the newly
functionalized site (site III) were performed. The intention was to explore the likelihood of
the new active site binding peptides in catalytic positions (Cys residue close to the peptide
bond) to see whether the new site could have protease activity or not. All dipeptides
reached good catalytic poses ($d_{C_{amide}-S_\gamma^{Cys}} \leq threshold$ Å & $d_{H_\gamma^{Cys}-N_\epsilon^{His}} \leq 3.5$ Å) around
site III, except for YY (*Table 4.1*). One example of the results with the simulation of two
dipeptides is shown in *Figure 4.3*.

These results encouraged us to send the single mutant to our collaborators from Ferrer's
laboratory (CSIC) for experimental verification of protease activity. Once the protease
and esterase activities of the developed pluriZyme were characterized, a cascade reaction
using both sites was elaborated. The chosen reaction was the synthesis of L-histidine
methyl ester, an intermediate for the design of Schiff base ligands (Ogawa et al., 2018),
from the dipeptide L-carnosine. To computationally test whether the artificial protease
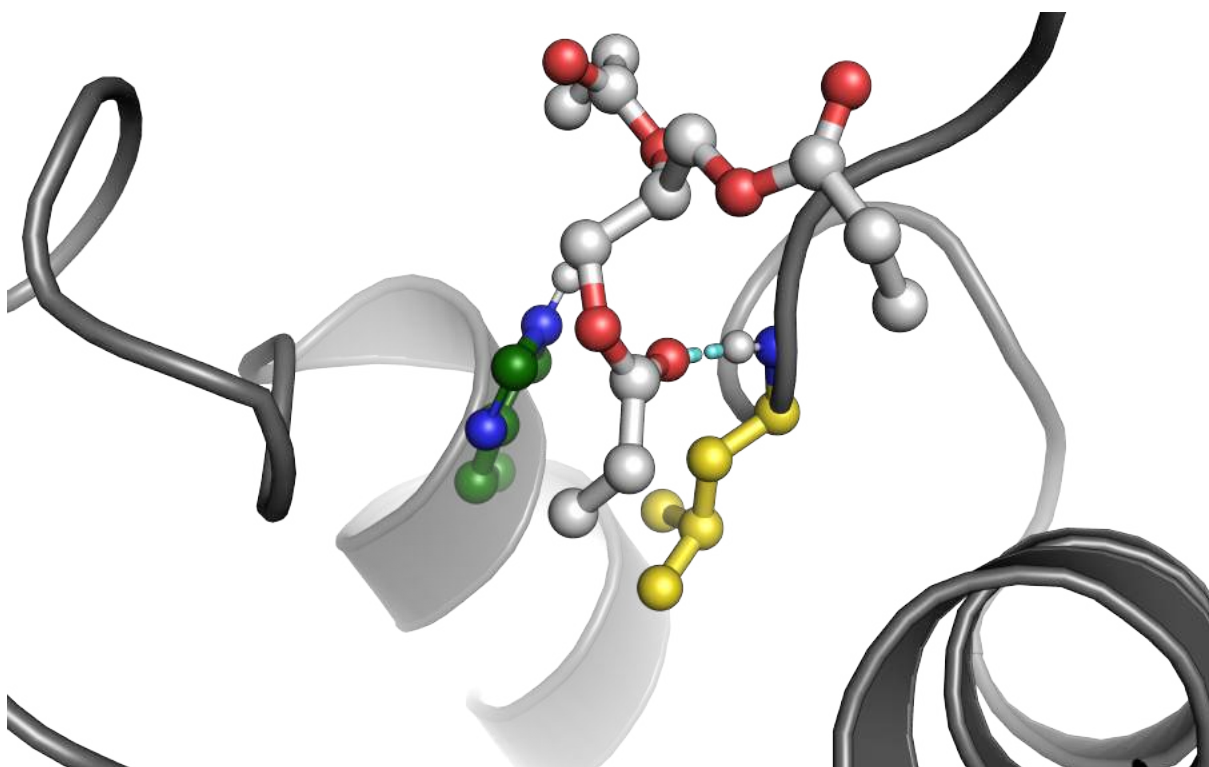site could break the peptide bond in L-carnosine, a local PELE exploration was per-

FIGURE 4.2: Illustrative binding pose of glyceryl tripropionate with the Leu24 and catalytic His from site II are highlighted. Leu24 makes a H-bond with the carbonyl O atom in the ester bond by the NH group in the backbone (indicated by a cyan dash line).

TABLE 4.1: Absolute and relative (divided by all the other poses. Value in %) number of catalytic poses with $threshold$ of $d_{C_{amide}-S_\gamma^{Cys}} \leq threshold$ Å being 3.75, 4,25, 5 Å.

| Dipeptide substrate	Number of accepted catalytic poses [3.75 Å]	Number of accepted catalytic poses [4.25 Å]	Number of accepted catalytic poses [5 Å]
AH	129 (0.265 %)	443 (0.91 %)	779 (1.6 %)
AQ	17 (0.04 %)	105 (0.248 %)	421 (0.995 %)
DI	3 (0.012 %)	68 (0.279 %)	144 (0.59 %)
EA	14 (0.034 %)	584 (1.416 %)	1200 (2.91 %)
FF	0	7 (0.025 %)	41 (0.145 %)
KA	5 (0.013 %)	412 (1.093 %)	1197 (3.177 %)
LA	14 (0.033 %)	777 (1.858 %)	1163 (2.781 %)
LL	0	17 (0.044 %)	67 (0.172 %)
NV	0	7 (0.018 %)	371 (0.935 %)
PF	58 (0.171 %)	495 (1.456 %)	657 (1.933 %)
QQ	20 (0.065 %)	128 (0.415 %)	295 (0.955 %)
RG	9 (0.029 %)	171 (0.553 %)	1016 (3.286 %)
SW	47 (0.147 %)	444 (1.393 %)	513 (1.609 %)
TM	38 (0.095 %)	404 (1.014 %)	635 (1.594 %)
YN	20 (0.081 %)	56 (0.227 %)	72 (0.292 %)
YY	0	0	0
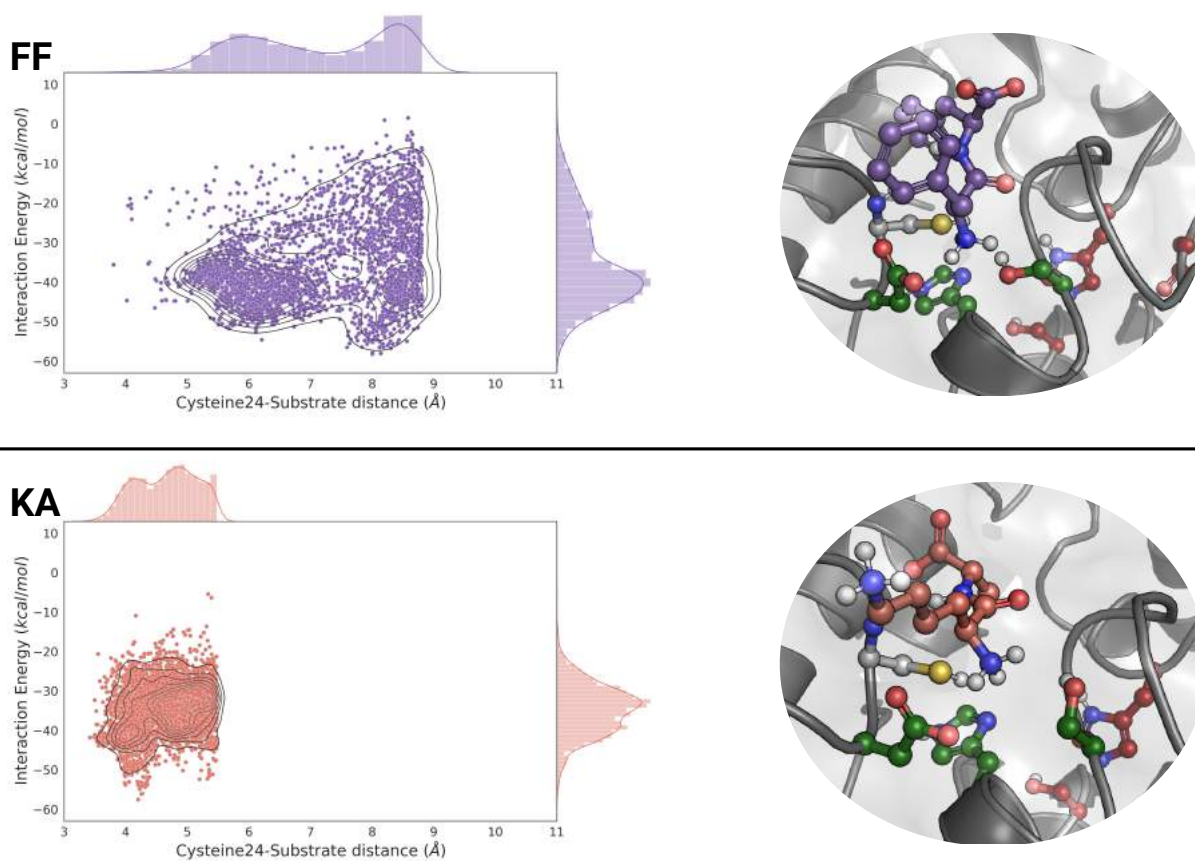
FIGURE 4.3: Density plots of the distribution of the catalytic cysteine-substrate distances against the interaction energy of FF and KA peptides. Only the poses of the 10% lowest percentile regarding the distance are shown. A representative pose of the bound dipeptide in site III is depicted on the right.
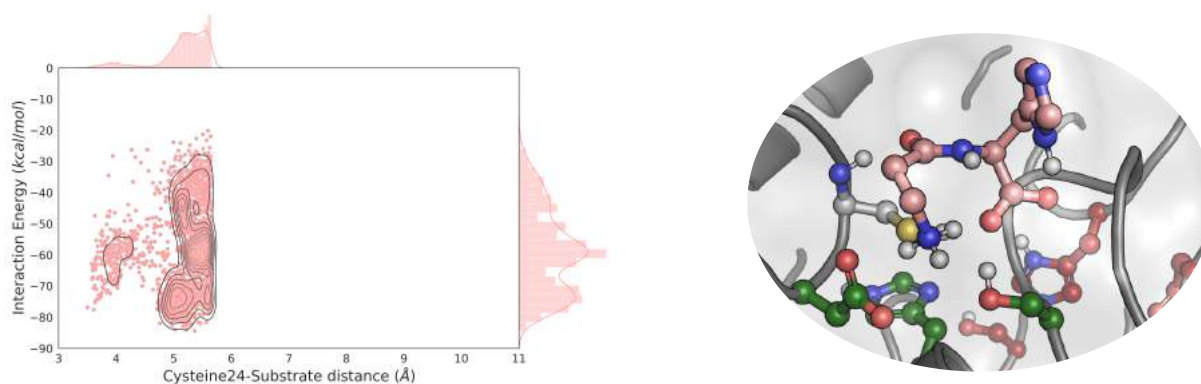
FIGURE 4.4: Density plot of the distribution of the catalytic cysteine-substrate distances against the interaction energy of L-carnosine peptide. Only the poses of the 10% lowest percentile regarding the distance are shown. A representative pose of the bound dipeptide in site III is depicted on the right.

formed. The results showed good catalytic poses (26 for $d_{C_{amide}-S_\gamma^{Cys}} \leq 3.75$Å, 166 for $d_{C_{amide}-S_\gamma^{Cys}} \leq 4.25$Å, 425 for $d_{C_{amide}-S_\gamma^{Cys}} \leq 5$Å) and overall distribution of the catalytic cysteine-substrate distance versus the interaction energy (*Figure 4.4*).

The resulting designed catalyst could hydrolyze azo(casein) (a substrate model for protease activity). Likewise, it could synthesize L-histidine methyl ester from L-carnosine in the presence of methanol, showing its potential to perform one-pot cascade reactions.

4.2 A Plurizyme for the pharmaceutical industry

Until recently, the designed pluriZymes in our laboratory had no specific application for the industry. Thus, we wanted to show whether we could develop a compelling cascade reaction with a single protein scaffold using another type of enzyme family. The pharmaceutical industry must catalyze several chemical reactions to transform inexpensive starting fragments into valuable drugs/chemicals (requiring several inorganic or biological catalysts). As explained in *Chapter 1*, around 40 % of pharmaceutical drugs contain a chiral amino group within their structure (Ghislieri and Turner, 2014; Kelly et al., 2018). Hence, the design of a hydrolase site in an ω-TAs could be profitable for the pharmaceutical industry.

The used ω-TA (named TR_2) for that purpose is a class III ω-TA extracted from an acidic beach pool on Vulcano Island (Coscolín et al., 2019). The design protocol starts with the global exploration of the protein surface to find noncatalytic ester binding sites with PELE and glyceryl tripropionate as the probe. Since ω-TAs are functional dimeric proteins and each monomer had 457 residues, several noncatalytic ester binding sites were

found (*Figure 4.5*).

Several variants were designed to introduce catalytic triads in the different found ester binding sites, assaying more options (*Table 4.2*). The variants were created manually, taking into account a few rules/tips; the fewer mutations, the better (meaning trying to reuse present residues in the found cavity is always better), and the less drastic the changes, the better.

The variants were computationally tested with local PELE explorations, MD simulations, and $\Delta\Delta G$ estimations using the HotSpot Wizard server (Sumbalova et al., 2018). As in the previously explained paper, local PELE explorations were used to examine whether catalytic poses can be reached or not. MD simulations describe how well the catalytic architecture of the active site is maintained over time and the overall geometry of the protein. The $\Delta\Delta G$ estimations were used to see if some of the mutations were highly destabilizing. All the plots and data regarding these simulations can be seen in *Chapter 3*.

All variants were tried experimentally at Ferrer's laboratory (CSIC) to gather more information about designing pluriZyme variants and see if the best ones, according to the computational analysis, would be good in real life.

Once we had the experimental feedback of the variants, we studied whether we could tailor a cascade reaction with the newly created pluriZyme. The considered example was one precursor block of all gliptins (a family of antidiabetic drugs), (R)-3-amino-4-(2,4,5-trifluorophenyl)-butanoic acid (3-ATfBA) (Savile et al., 2010; Kim et al., 2019; Khobragade et al., 2021). 3-ATfBA is synthesized from an ester precursor (methyl 3-oxo-4-(2,4,5-trifluorophenyl)butanoate, 3-OTfBE) by using both a serine esterase and an ω-TA (Kim et al., 2019; Khobragade et al., 2021). To see if 3-ATfBA could be synthesized from the pluriZyme alone, we computationally studied whether 3-OTfBE and the amino ester, methyl 3-amino-4-(2,4,5-trifluorophenyl)butanoate (3-ATfBE), could bind to the added hydrolase site in a catalytic manner. Both substrates showed catalytic poses (23 for 3-ATfBE and 42 for 3-OTfBE), but with an order of magnitude lower than glyceryl tripropionate, expecting some catalytic activity, but slower than tripropionin. Representative found catalytic poses with PELE are shown in *Figure 4.6*.

The experimental validation showed that the developed enzyme could synthesize 3-ATfBA from 3-OTfBE in a single protein scaffold with a $\frac{k_{cat}}{K_M} = 202.5 M^{-1} \cdot s^{-1}$ for the amination of the β-keto acid and a $\frac{k_{cat}}{K_M} = 224.9 M^{-1} \cdot s^{-1}$ for the hydrolysis of the ester bond in the β-keto ester. Moreover, the added hydrolase site could hydrolyze 52 esters from a set of 96 diverse molecules, meaning the artificial active site was highly promiscuous. Thus, we decided to experimentally try other transformations of β-keto esters into enantiopure β-amino acids, and 3 out of 7 substrates were successfully converted.
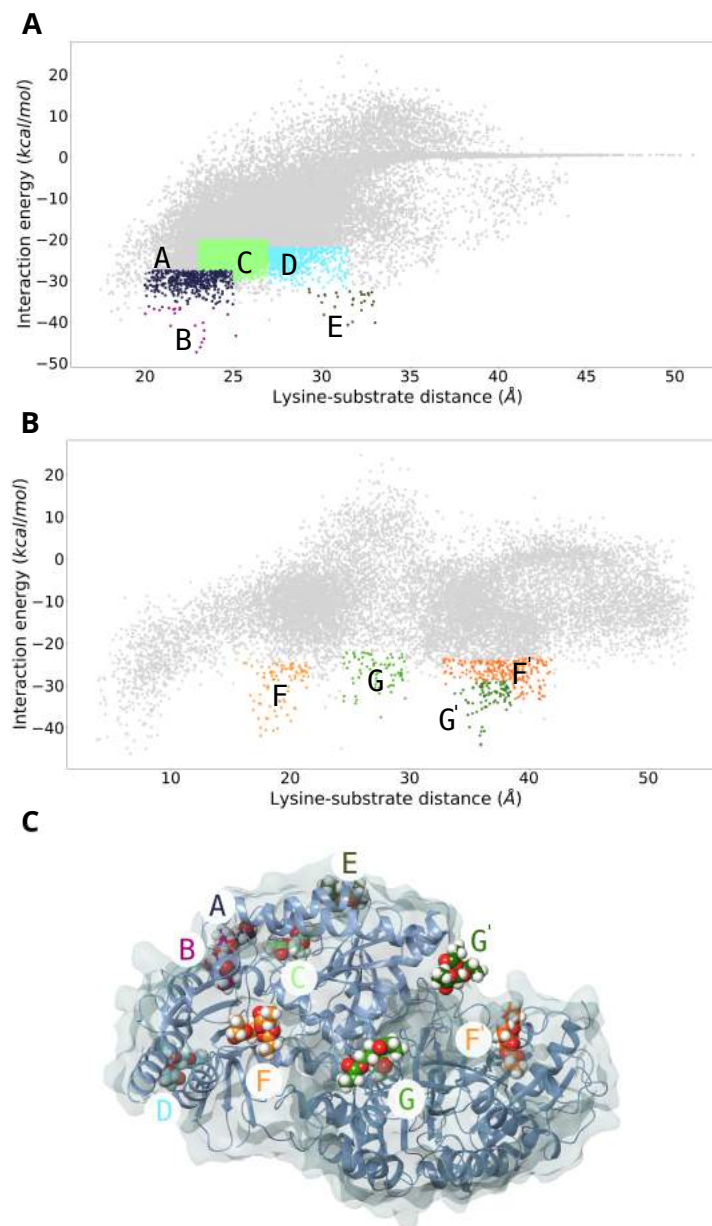
FIGURE 4.5: Global exploration of noncatalytic ester binding sites on the TR_2 surface with PELE. Scatter plots represent the global exploration of the monomeric (A) and dimeric (B) structures. In both plots, the X axis refers to the distance of the substrate to the transaminase site (the catalytic Lys residue bound to the PLP cofactor) and the Y axis to the interaction energy. The steps exploring around found ester binding sites are highlighted in a particular color (site A in dark blue, site B in purple, site C in lime, site D in turquoise, site E in dark green, sites F/F' in orange for chains A and B and sites G/G' in green for chains A and B). Representative poses of glyceryl tripropionate in the found ester binding sites are depicted (C atoms of the ester molecules are stained according to in which ester binding site they are placed).

TABLE 4.2: TR_2 designed variants, their mutations, and which site is being functionalized.

ID	Mutations	Number of mutations	Ester binding site
TR_2	WT	-	-
TR_2E_1	K110D	1	E
TR_2E_2	A172S Q173H	2	G
TR_2E_3	W272S V107H	2	C
TR_2E_4	M112H M302D	2	E
TR_2E_5	G420S G399H	2	F
TR_2E_6	M112H T108D	2	E
TR_2E_7	V377S Y277H T271D	3	A
TR_2E_8	V377S Y277H Y278D	3	A
TR_2E_9	Q52S Y44H Q435E	3	D
TR_2E_{10}	F190S Y188H I236D	3	B
TR_2E_{11}	F190S Y188H G191D	3	B
TR_2E_{12}	F190S Y188H R192E	3	B
TR_2E_{13}	A169S M167H M170D	3	G
TR_2E_{14}	W272S I347H E348D	3	C
TR_2E_{15}	V107S W272H G351K	3	C
TR_2E_{16}	V377S Y277H T271D Q276E	4	A
TR_2E_{17}	V107S W272H I347D G351K	4	C
TR_2E_{18}	A251S D283H R306D	3	E
TR_2E_{19}	F190S E384H T422D R375A	4	B
TR_2E_{20}	F190S E373H E384D R375G R394G	5	B

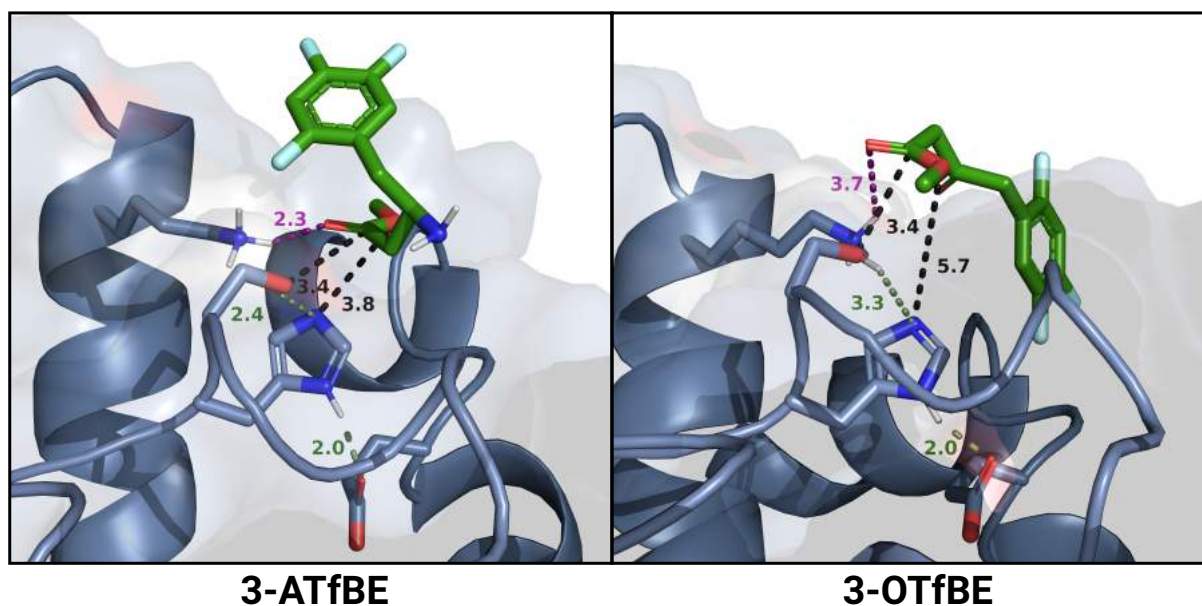**3-ATfBE**   **3-OTfBE**

FIGURE 4.6: Representative catalytic poses of 3-ATfBE and 3-OTfBE in
the hydrolase site of the successful variant. Catalytic triad distances (in
Å) are shown in green, the substrate-serine/histidine distances are shown
in black, and the oxyanion-hole distances are displayed in purple.

To computationally demonstrate the option of substrate channeling between both active sites, we simulated the migration of 3-oxo-4-(2,4,5-trifluorophenyl)-butanoic acid (3-OTfBA), the product of the hydrolysis of 3-OTfBE, from the added esterase site to the WT transaminase site with PELE. Likewise, we designed two mutants (A232F/L60F and A232F/F89W) to close the access channel, hindering the access of the reaction intermediate to the transaminase site. These simulations showed that the substrate could easily reach the transaminase site from the designed hydrolase site but not in the two enclosing mutants (*Figure 4.7*). Experimentally, we saw that the hydrolase activity of the assayed mutant (A232F/L60F) was similar to the original pluriZyme, but the transaminase activity was lower. Moreover, the conversion of 3-OTfBE to 3-ATfBA was 3 times slower in the enclosing mutant at short incubation times. Hence, both computational and experimental results proved the possible substrate channeling between both functional sites.

To conclude, the pluriZyme is capable of using both active sites synergistically to transform β-keto esters into enantiopure β-amino acids. Moreover, the mutations of the successful pluriZyme were introduced in a missing loop at the original crystal structure of TR_2, which was modeled with Prime (Madhavi Sastry et al., 2013) and refined with MD simulations (*Figure 4.8*). Hence, inserting a catalytic triad stabilized the missing loop by adding two internal H-bonds, improving the thermostability of the enzyme, meaning a second biological active site does not always come at the expense of losing stability or activity of the original active site.
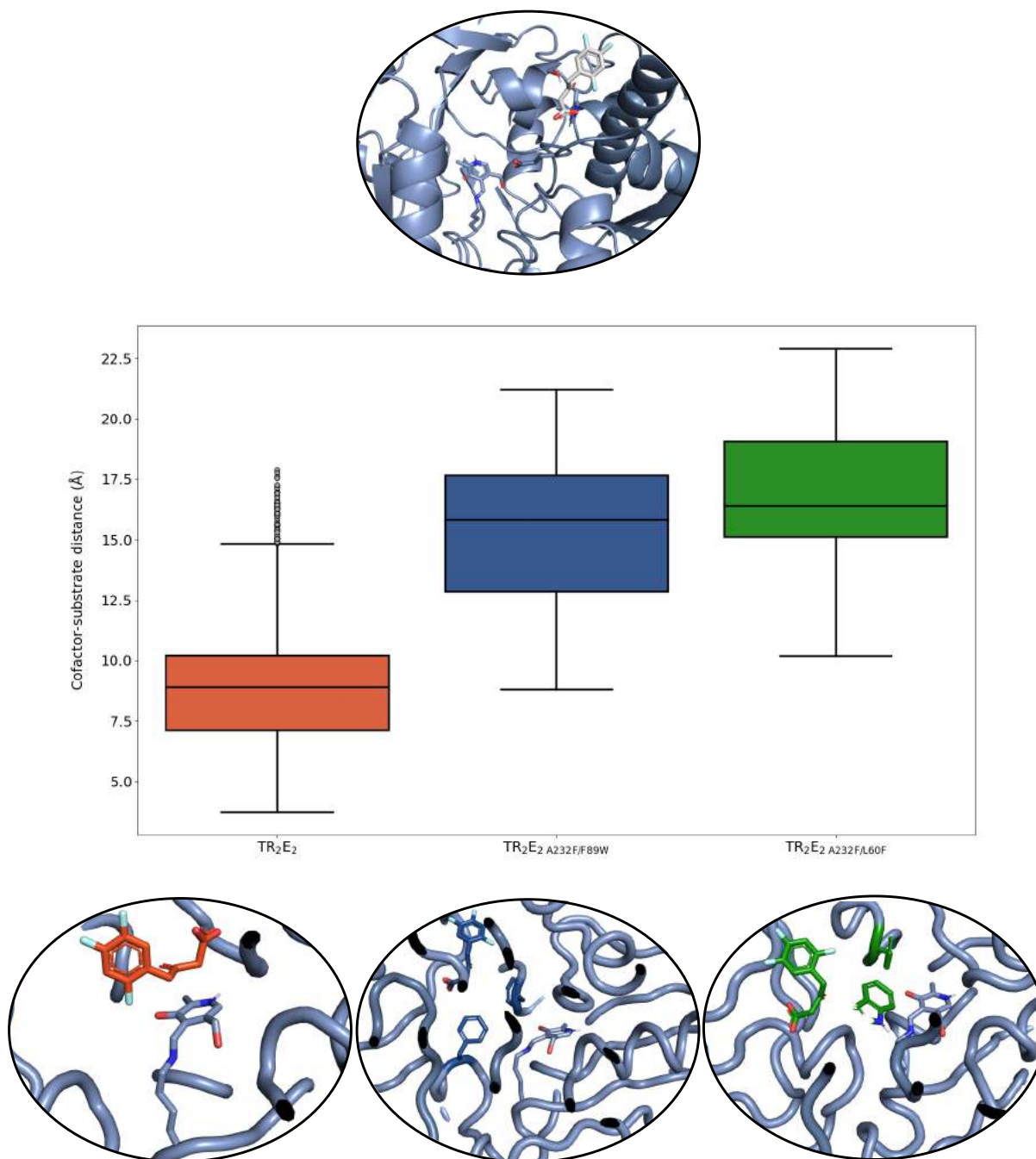
FIGURE 4.7: Initial pose near the esterase site from where 3-OTfBA started in PELE's migration simulations (top). Box plot representing the cofactor-substrate distance (distance to PLP in the transaminase site) of all the accepted PELE steps with a relative SASA of 0.1 or lower from these simulations (middle). Representative 3-OTfBA poses of the PELE's migration simulations in TR_2E_2, $TR_2E_{2\ A232F/L60F}$, and $TR_2E_{2\ A232F/F89W}$ (bottom).
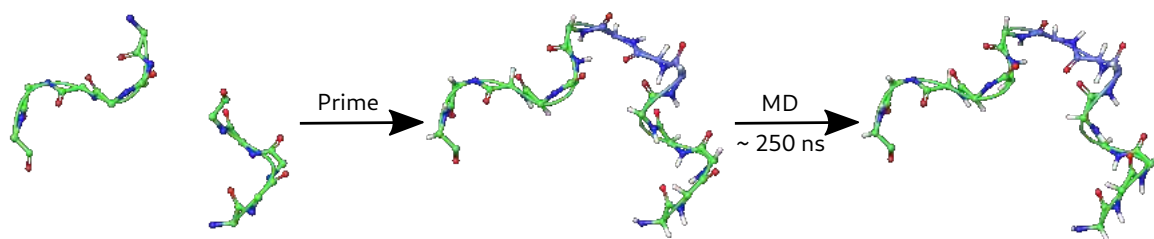
Figure 4.8: Modeling of the missing loop where the catalytic triad of the TR_2E_2 variant resides. The initial crystal structure had the missing loop modeled using Prime (Madhavi Sastry et al., 2013). Then, the modeled loops were further refined with 250 ns (4 replicas) of apo MD of the dimer structure to obtain the final structure used in the analysis. Both the secondary structure and the C atoms are shown in light green for the residues present in the crystal structure and in blue for the missing residues.

4.3 Rational design of substrate promiscuity on serine esterases

Once Ser-His-Asp/Glu catalytic triads are inserted into a functional enzyme, creating a pluriZyme, it would be appealing to be able to engineer the new active site to enhance the property of interest. Improving the overall activity of a serine hydrolase has been performed widely (Ma et al., 2018; Liu et al., 2018; Moore and Arnold, 1996; Bornscheuer et al., 1998). Still, the engineering of substrate promiscuity (the ability of an enzyme to catalyze several substrates of different chemical nature) is less explored.

An in-depth analysis of the active site's cavity and the study of the enzyme-substrate migration pathways in two-high promiscuous and two low-promiscuous serine hydrolases were performed before starting the design of substrate promiscuity on a serine hydrolase active site. The four studied enzymes were previously published (Martínez-Martínez et al., 2018, 2013), where three (EH1, EH88, and EH102) were extracted from a metagenomic sample of Lake Arreo (Martínez-Martínez et al., 2018, 2013), and the other one was CalB. EH1 and CalB were used as the high-promiscuous esterases (capable of hydrolyzing 72 and 68 esters of a library of 96 different esters), and EH88 and EH102 as the low-promiscuous ones (capable of hydrolyzing 13 and 16 esters of a library of 96 distinct esters).

First, the active site's cavity was analyzed with SiteMap (Halgren, 2007, 2009) and seven features were extracted; volume, size (the number of inside points found in the cavity according to SiteMap), exposure, enclosure, contact (tightness, the higher, the more compact), hydrophobicity, and hydrophilicity. As seen in *Figure 4.9*, high-promiscuous esterases have bigger, more hydrophobic, and more compact (higher contact value) cavities with more shelter from the solvent (higher enclosure/exposure ratios).
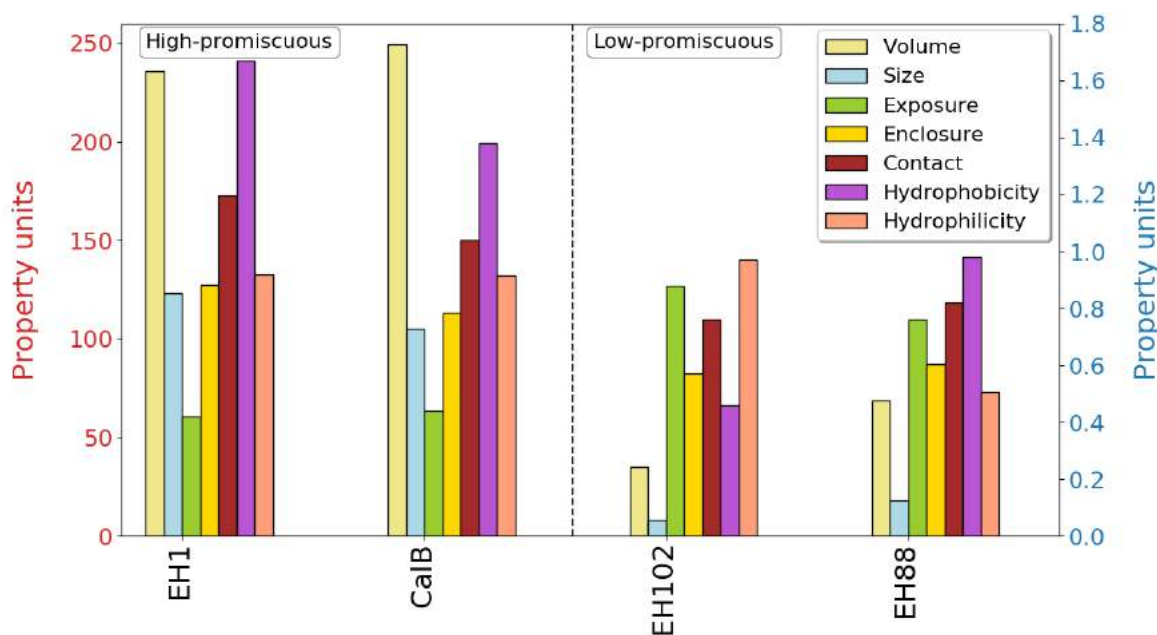
FIGURE 4.9: Bar plot showing the SiteMap features of the active site's cavities of the four studied esterases. The Y red axis comprises the volume (in Å^3) and the size. The blue Y axis represents the remaining properties.

Then, the substrate-enzyme interactions were simulated with PELE to see how they correlate with their substrate range. Two substrates were studied for all systems; phenyl acetate, which could be hydrolyzed by all of them, and ethyl-3-oxohexanoate, which could only be hydrolyzed by the high-promiscuous biocatalysts. All the plots regarding these simulations can be found in *Chapter 3*. Basically, they show that high-promiscuous esterases tend to have better interaction energies with the substrates and the reactive region of the substrate tends to stay closer to the active site than the low-promiscuous ones. Moreover, the SASA values are lower, meaning that the substrate is more buried in the cavity in the high-promiscuous serine hydrolases. Plots representing the simulations of EH1 and EH102 against phenyl acetate are shown in *Figure 4.10* to exemplify these results.

This rationalization of substrate promiscuity in serine hydrolases led us to transform EH102 into a high-promiscuous esterase, using the different features and simulations that let us know what a high-promiscuous esterase is. EH102 was selected over EH88 because of the availability of its crystal structure (PDB ID: 5JD3), having a more reliable starting point. By taking a look at the vicinity of the catalytic triad, several bulky residues can be found, hiding a potentially bigger and more enclosed cavity that would allow the binding of many more substrates (*Figure 4.11*). These residues are Ile16, Ile92, and Trp96. The first assayed mutant aimed at drastically enlarging the active site's cavity with three mutations; I16G/I92A/W96G ($EH102_{TM1}$). The active site's cavity properties were computed with SiteMap for the variant to see if the applied changes made the enzyme's cavity resemble one from a high-promiscuous esterase. Encouragingly, the volume, size,

FIGURE 4.10: Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy (top) and SASA (bottom) from the 10 % lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulations for EH1 and EH102 with phenyl acetate as substrate. A representative binding pose is displayed (where the ligand C atoms are stained in lilac and the protein ribbon colored according to the secondary structure; ruby: α-helix, golden: β-sheet, and gray: loops).
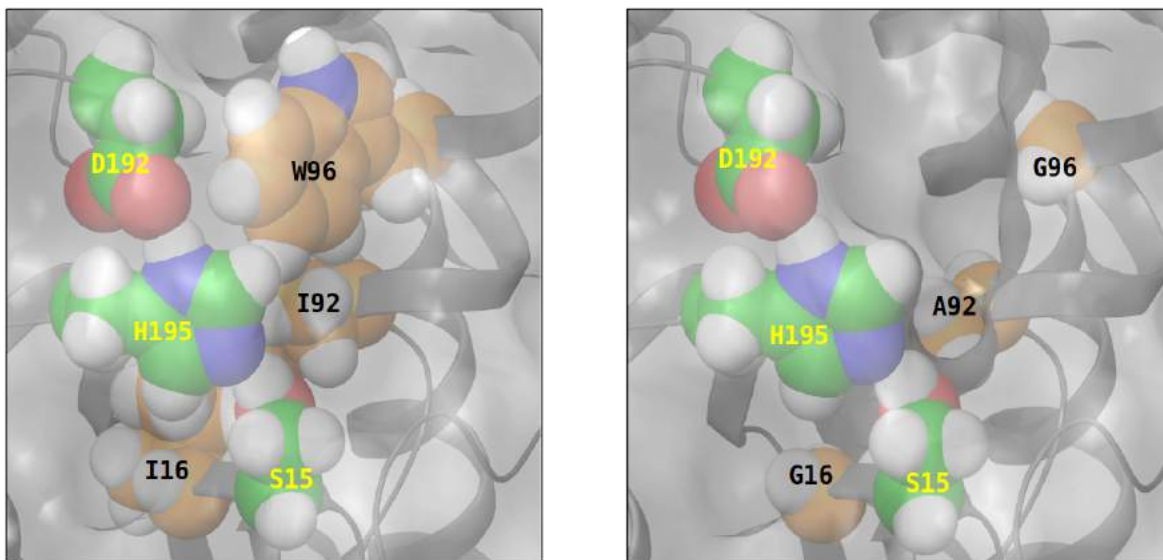
FIGURE 4.11: Active site's cavity of $EH102_{WT}$ (left) and $EH102_{TM1}$ (right). The bulky residues targeted for mutation have the C atoms stained in orange, while the catalytic residues have them in green.

tightness, enclosure, and hydrophobicity of the new cavity increased, while its exposure decreased (*Figure 4.12*).

Also, PELE simulations to study the enzyme-substrate interactions in the designed cavity with two unhydrolyzed substrates (ethyl-3-oxohexanoate and vinyl crotonate) by the WT enzyme were performed. Both substrates showed better accommodation in the variant compared to the WT enzyme with smaller serine-substrate distances, interaction energies, and SASA values (plots regarding the simulation with ethyl-3-oxohexanoate as substrate are displayed in *Figure 4.13*).

The variant was sent to be assayed experimentally at Ferrer's laboratory (CSIC) to verify if the designed mutations accomplished an increase in the substrates hydrolyzed by the enzyme compared to the WT. Indeed, the mutant doubled the number of catalyzed substrates (from 16 in the WT to 35 in $EH102_{TM1}$). However, the overall activity (meaning the average of the k_{cat} against the different substrates) was compromised.

Mutating residues near the active site can alter the overall geometry of the catalytic residues, making the H-bond distances in the catalytic triad worse. This effect of the mutations can be monitored through MD simulations, and the improvement of these distances can boost the overall activity of the biocatalyst (Santiago et al., 2018; Alonso et al., 2019; Roda et al., 2021).

The mutations in $EH102_{TM1}$, which went from bulky hydrophobic residues to smaller ones, could have introduced more flexibility to the catalytic triad. Residue 16 is next to
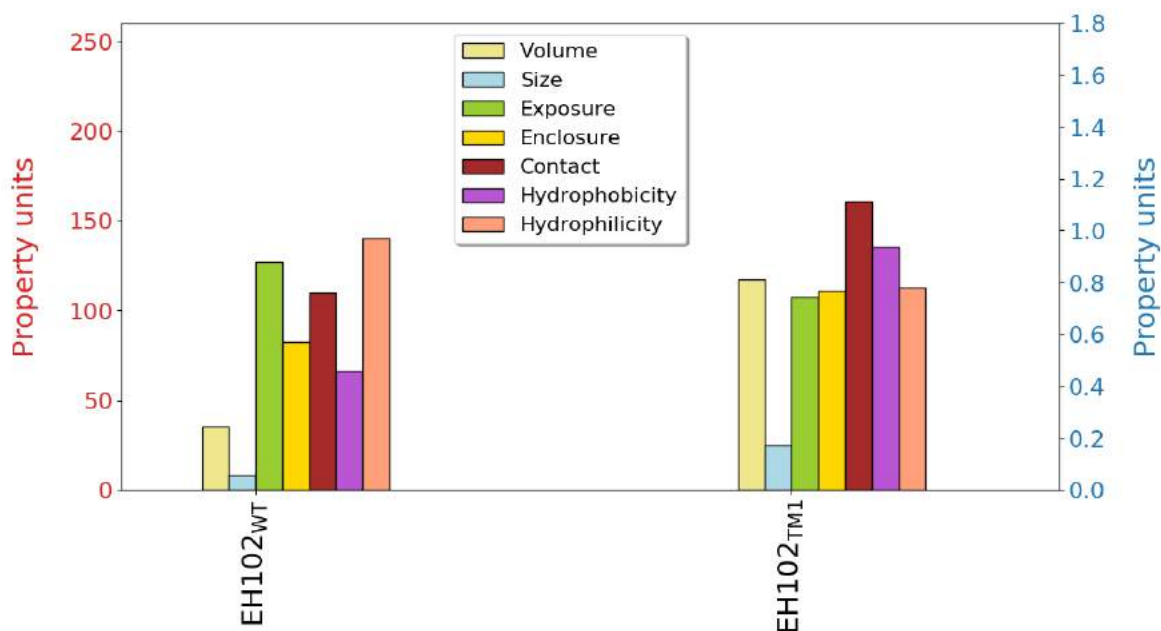
FIGURE 4.12: Bar plot showing the SiteMap features of the active site's cavities of $EH102_{WT}$ in comparison with $EH102_{TM1}$. The Y red axis comprises the volume (in Å^3) and the size. The blue Y axis represents the remaining properties.
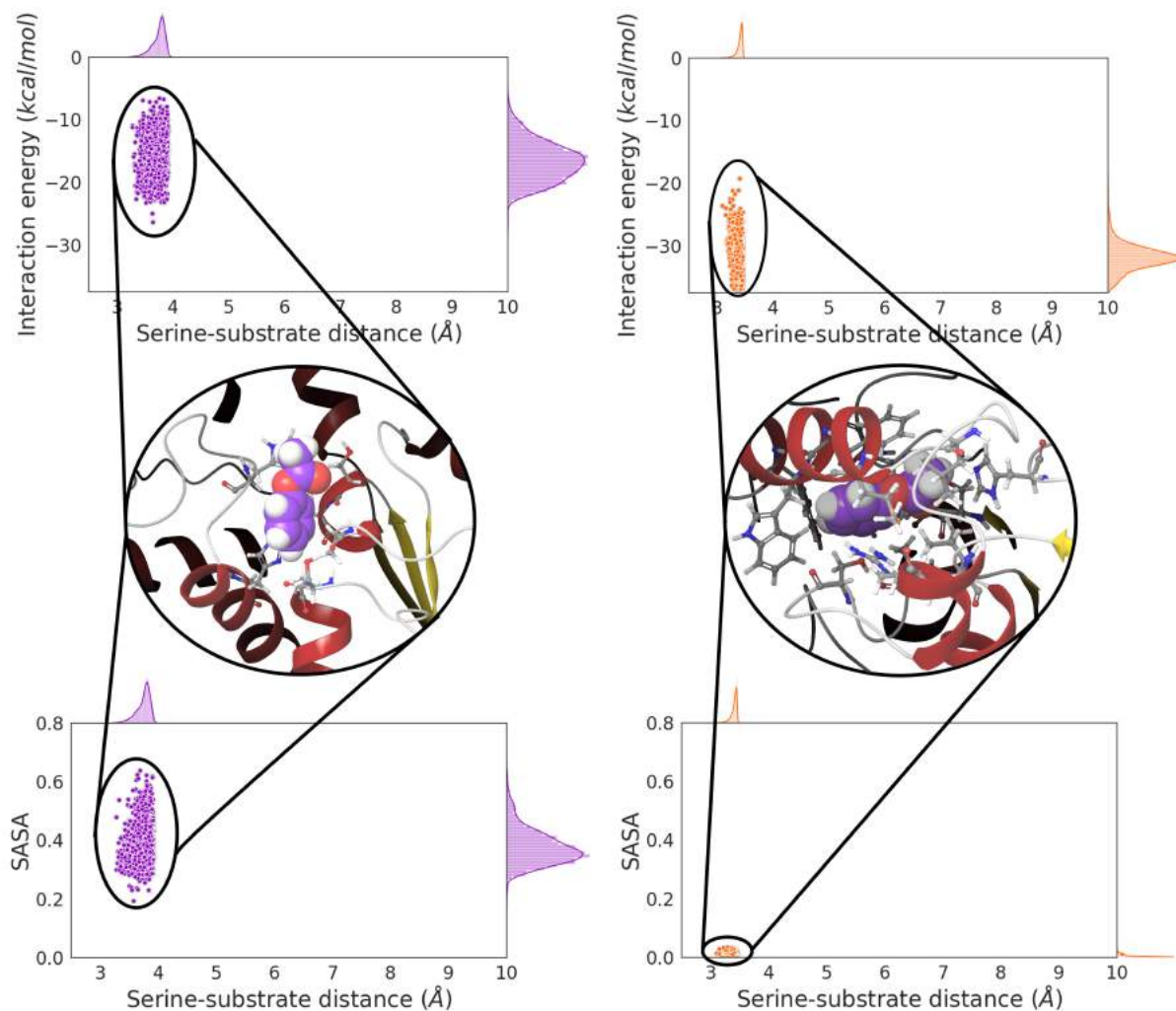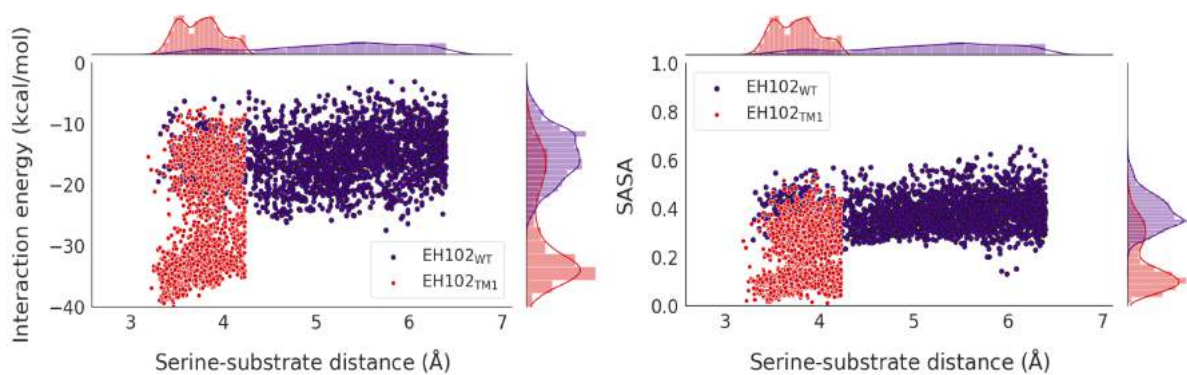

FIGURE 4.13: Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy (left) and SASA (right) from the 10 % lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulations for $EH102_{WT}$ and $EH102_{TM1}$ with ethyl-3-oxohexanoate as substrate.
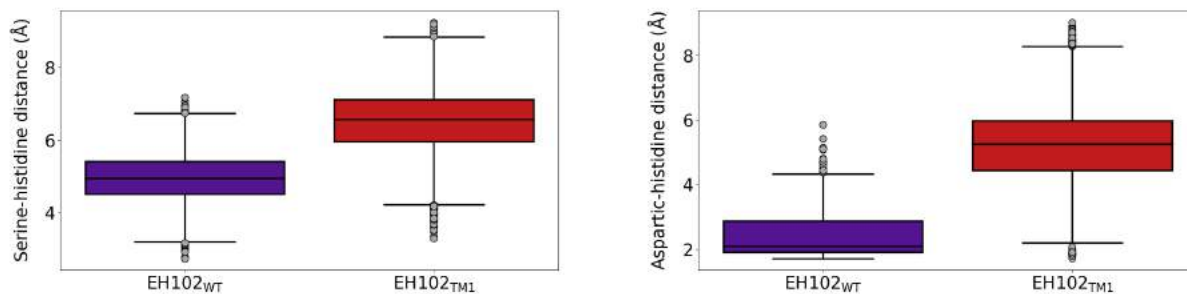
FIGURE 4.14: Box plots representing the Ser-His ($d_{H_\gamma^{Ser}-N_\epsilon^{His}}$) and Asp-His ($d_{O_\delta^{Asp}-H_{\delta N}^{His}}$) distances along the 500 ns of the 4 MD replicas performed for $EH102_{WT}$ and $EH102_{TM1}$.

TABLE 4.3: Recompilation of all the EH102 designed variants

WT	TM1	TM2	TM3	TM4	TM5	TM6
Ile16	Gly16	Val16	Val16	Ala16	Ala16	Gly16
Ile92	Ala92	Ala92	Gly92	Gly92	Ala92	Gly92
Trp96	Gly96	Gly96	Gly96	Gly96	Gly96	Gly96

WT	DM1	DM2	SM1	SM2	SM3
Ile16			Val16	Ala16	Gly16
Ile92	Gly92	Ala92			
Trp96	Gly96	Gly96			

the catalytic Ser residue, which means it could be hindering the H-bond between Ser and His residues. 4 replicas of 500 ns MD simulations were run to confirm this hypothesis. Both Ser-His and Asp-His distances significantly increased in the mutant compared to the WT (*Figure 4.14*).

Based on the MD results, we proceeded by designing the second round of mutants aiming at increasing the substrate promiscuity without compromising the overall activity. Single mutants of position 16 were computationally assayed to see the progressive effect of substituting for a smaller and less-hydrophobic residue. Double mutants considered the other two positions (92 and 96) to reduce the increase of flexibility of the catalytic Ser residue and check whether substrate promiscuity can be increased without changing residue 16. Triple mutants were designed by combining the variants from the single and double mutants (*Table 4.3*). W96 was always mutated to G, since it is the most solvent-exposed and biggest residue of the three residues, significantly increasing the cavity once it is mutated.

Again, the active site's cavity properties were computed with SiteMap for all the variants. The triple and double mutants showed similar cavities to the one from $EH102_{TM1}$ (*Figure*

4.15). On the other hand, single mutants gave smaller improvements of the cavity's properties.

The computational study of these mutants was followed by MD simulations to probe the stability of the catalytic triad over time. Both $EH102_{TM2}$ and $EH102_{TM6}$ improved the catalytic distances and the overall geometry of the catalytic triad (check the plots in *Chapter 3*). $EH102_{TM3}$, $EH102_{TM4}$, $EH102_{DM1}$, and $EH102_{DM2}$ maintained similar values with the WT enzyme. The remaining mutants showed disturbed catalytic triads according to the studied metrics and performed simulations.

$\Delta\Delta G$ estimations were used to see which variants would be more destabilizing than others. According to this calculation, $EH102_{TM2}$ seemed more promising (better $\Delta\Delta G$ value) than $EH102_{TM6}$. The MD results and stability analysis also suggested that residue 16 should be either substituted with a valine residue or preserved due to its role in maintaining the catalytic integrity. Similarly, the results of both $EH102_{DM1}$ and $EH102_{DM2}$ variants seemed to point out that residue 92 must have some side chain (avoiding a Gly residue).

Either way, we sent all mutants to be tested by our collaborators from Ferrer's laboratory (CSIC) for experimental validation to see if we could predict which variants could be better computationally. The outcome was two prominent promiscuous esterases that could hydrolyze 51 and 63 esters ($EH102_{DM2}$ and $EH102_{TM2}$, respectively) from the mentioned library of 96 esters with similar or higher turnover rates in contrast with the WT enzyme.
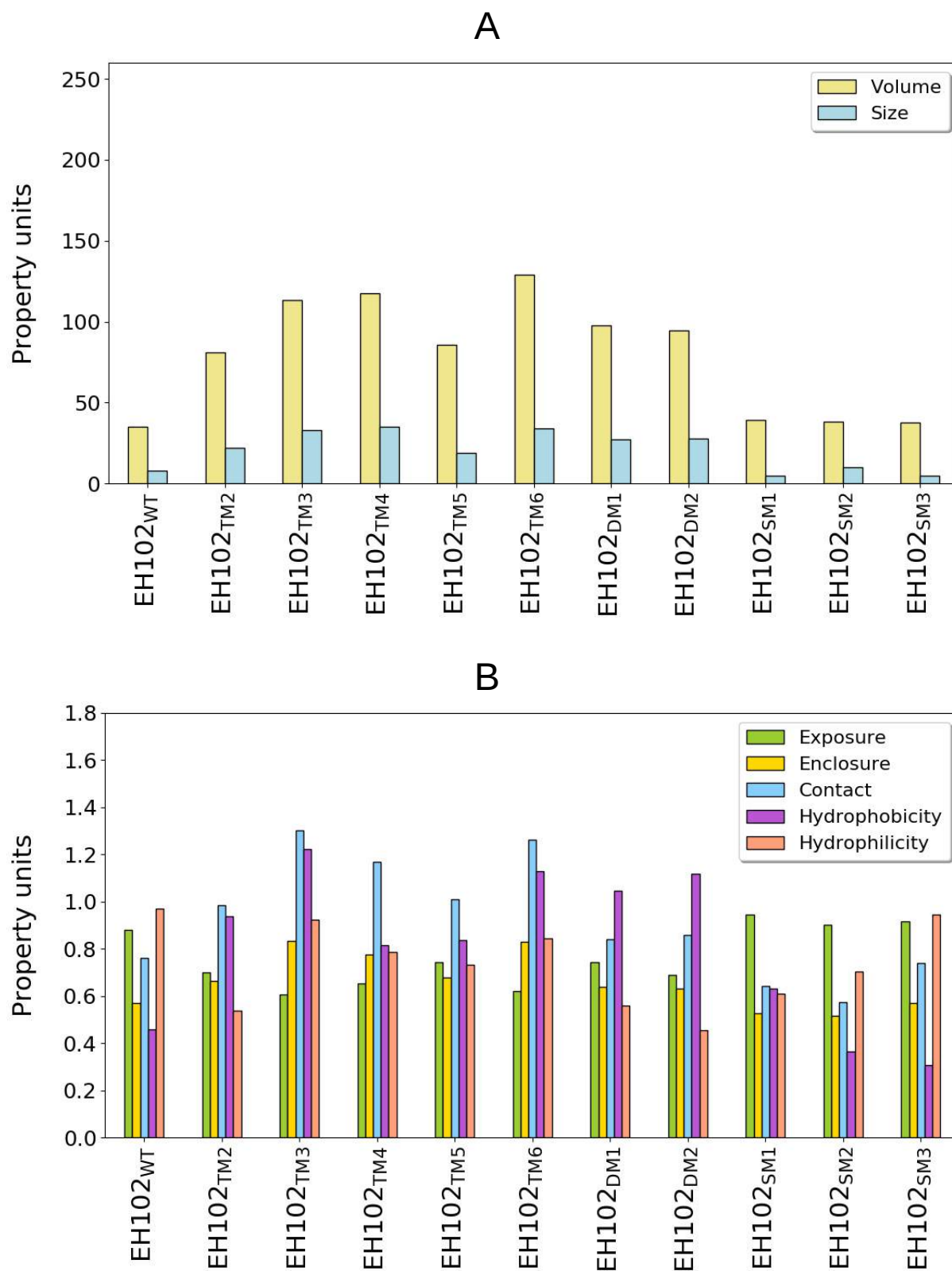
Figure 4.15: Bar plots showing the SiteMap features of the active site's cavities of $EH102_{WT}$ in comparison with the second round mutants. The top plot (A) represents the volume (in Å^3) and the size. The bottom (B) plot represents the remaining properties.

4.4   A tool to automate the design of pluriZymes

PluriZymes developed in our laboratory used chemical intuition to add the catalytic residues in a found alternative binding site to functionalize it. A set of rules/factors were considered, such as the conservation of the mutated residues, prioritization to mutate residues of the same category to the introduced one, choosing those combinations where the catalytic residues are close in sequence and not only spatially, and trying to add the functionality with the fewer mutations possible (reusing some already present residues in the binding site). Although these designs were studied computationally through analysis of the enzyme-substrate interactions in the new functionalized site (with PELE simulations) and the stability of the catalytic architecture and the overall fold (with MD simulations), the developed variants relied on someone coming up with them.

AsiteDesign, a heuristic algorithm capable of identifying the best positions for placing the catalytic residues of a specific active site in the user's protein scaffold, was developed. The reason behind making this method was to democratize the idea of pluriZymes and allow everyone to design an active site in their enzyme of interest with relatively modest computational resources. The algorithm consists in an MC simulation (implemented with pyRosetta) that randomly assigns the catalytic residues to a set of positions defined by the user (being the residues in the alternative binding site) at every iteration, aiming to find the optimal positions for these residues. The correct distances of the catalytic residues are enforced by imposing distance restraints along the simulations.

At each iteration, a random position is mutated to one of the catalytic residues, and the previously mutated position is recovered to the native residue. Simultaneously, the substrate is first sampled by a rigid random translation and rotation. Then, a random set of its rotatable bonds are chosen, followed by a partial optimization of them (instead of assigning a random angle value). This sampling approach consists in an iterative grid search to find the minimum energy conformation(s) (a full explanation on how it works in *Chapter 3*). At last, the obtained ligand conformation is minimized with a gradient descent approach.

An adaptive reinforcement learning protocol was implemented (Lecina et al., 2017) to enhance the sampling in finding the optimal designs. The reason was that the number of possible designs increases substantially with the number of input positions, plus the iterations get more costly with bigger ligands. Basically, the algorithm runs in epochs where the MC sampling of the catalytic residues is performed by multiple processors in parallel. After an epoch ends, all designs are ranked based on a given objective function (for instance, the total energy or the ligand energy), and the top ones are used as starting point in the next epoch.

TABLE 4.4: Top 10 catalytic designs returned by ASiteDesign with PFE's mutated structure. The variants are ranked according to the total energy.

Total Energy	Mutations
-3170.1	A94S/A251H/A222D (WT)
-3160.1	W28S/L29H/T191D (PFE_1)
-3159.6	A94S/V225H/A222D
-3157.8	A94S/A251H/F162D
-3156.8	A94S/A251H/I224D
-3156.1	W28S/L29H/I155D (PFE_1*+I155D)
-3150.8	A94S/V225H/F125D
-3150.1	W28S/L29H/A183D (PFE_1*+A183D)
-3148.5	W28S/V195H/T191D
-3147.7	W28S/M95H/V121D

Moreover, AsiteDesign can also be used to design an active site to enhance productive substrate binding (and thus, catalytic activity). Hence, it can be employed for *in silico* directed evolution of enzymes and to generate a smart library of mutants.

The method was assayed against the *Pseudomonas fluorescens* esterase (PFE) (Khalameyzer et al., 1999) to see if it could design an alternative catalytic triad in its active site's cavity. Likewise, the protocol was employed to enhance the hydrolysis of a bulky chiral substrate (1-phenyl-2-pentyl acetate) by the WT's active site.

The enzyme contains the archetypical Ser-His-Asp catalytic triad (*Figure 4.16*), but we were aiming at designing alternative catalytic triads on the same active site's cavity. The amino acids comprising the WT catalytic triad were mutated to Ala to test the performance of AsiteDesign in identifying optimum positions for placing the catalytic residues.

The mutated structure was used as a starting point for the MC simulation with ethyl acetate as the substrate. All of the positions in the active site's cavity were allowed to be assigned one of the catalytic residues; 27, 28, 29, 30, 34, 57, 69, 93, 94, 95, 120, 121, 125, 135, 139, 140, 143, 147, 154, 155, 158, 162, 183, 191, 195, 198, 199, 222, 224, 225, 230, and 251.

As expected, the results showed that the best found solution was the WT catalytic triad (*Table 4.4*). Additionally, three variants seemed promising as well. In those designs, the catalytic Ser residue is placed in position 28, and the catalytic His residue in position 29. The only difference between the variants resides in the location of the catalytic Asp residue (*Figure 4.16*).

FIGURE 4.16: PFE and its catalytic residues (top). The catalytic triad
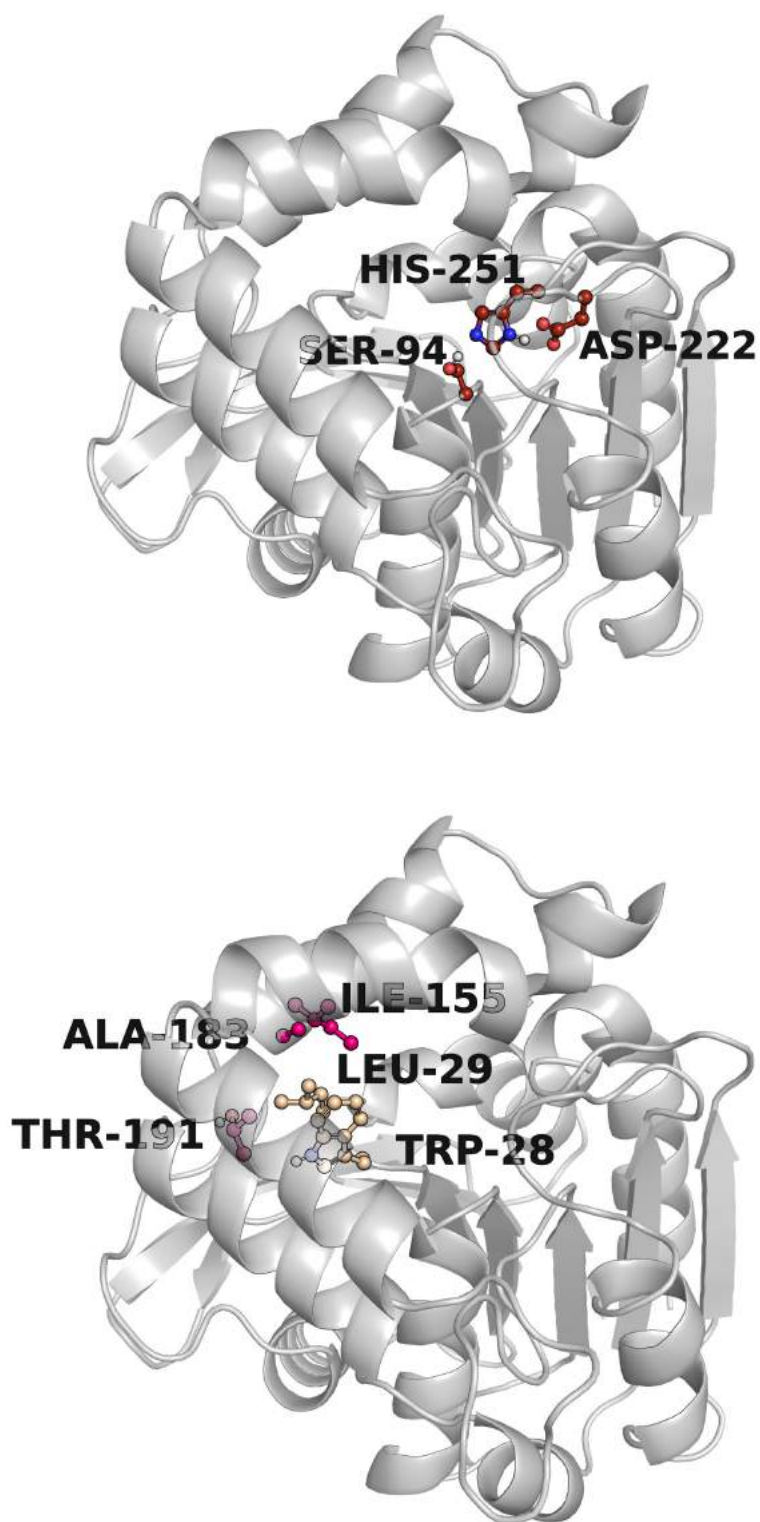residues are colored in red and labeled (PDB ID: 1VA4). Potential positions
for the design of an alternative catalytic triad (bottom). The positions
colored in beige (28 and 29) represent the ones that were mutated to Ser
and His in the promising variants. The residues colored in pink are all the
possible locations of the Asp residue in the promising variants.
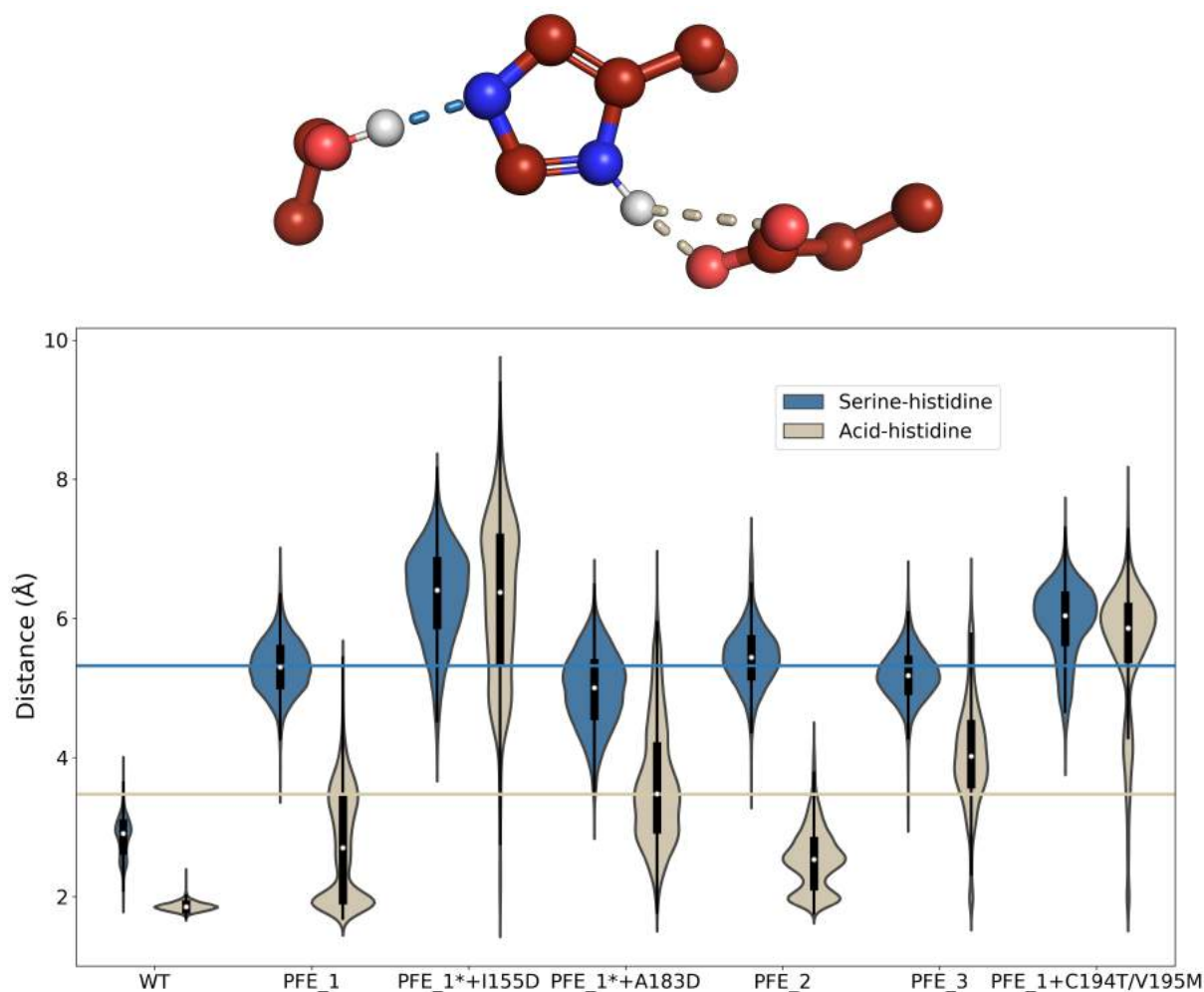
FIGURE 4.17: Violin plots of the catalytic Ser-His ($d_{H_\gamma^{Ser}-N_\epsilon^{His}}$) and Asp-His ($d_{O_\delta^{Asp}-H_{\delta N}^{His}}$) distances along the MD simulations. PFE_1* stands for W28S/L29H/S94A.

4 replicas of 100 ns MD simulations were performed for the WT enzyme and each variant to predict their stability. These simulations pointed out that the catalytic designs are less stable compared to the WT one (*Figure 4.17*). Still, the PFE_1 variant seemed to have acceptable catalytic distances. Some mutations were introduced to improve them, but did not succeed.

The PFE_1 variant and two others with additional mutations were sent to be experimentally verified by Bornscheuer's laboratory (University of Greifswald). Due to the MD results and the values of the catalytic distances, we expected a lower activity of the alternative catalytic triad than the WT one. Certainly, the experimental results showed PFE_1 had hydrolase activity, but lower than the WT enzyme.

As mentioned earlier in the text, AsiteDesign was assessed for the improvement of the hydrolytic activity of the WT enzyme towards 1-phenyl-2-pentyl acetate. The MC simu-

FIGURE 4.18: Active site of PFE and the used design domain. The catalytic triad residues are colored in red, the mutable residues in yellow, and the only repackable residues in violet.

lation included 11 residues of the active site (excluding the catalytic residues) as mutable, and 20 more were only repacked (*Figure 4.18*). As the enzyme was intended to hydrolyze a hydroph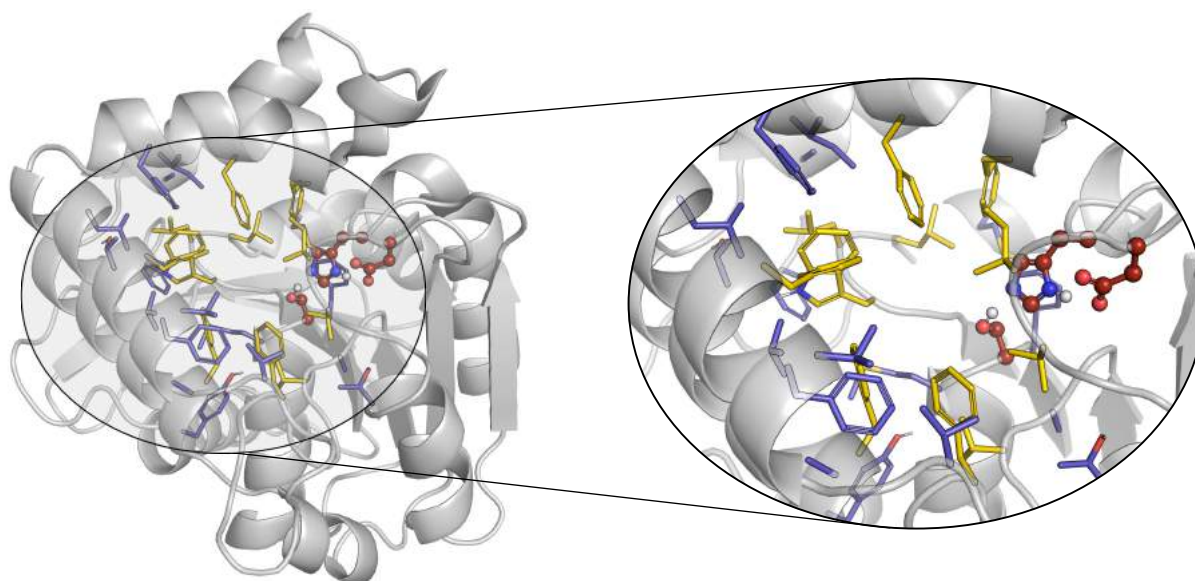obic substrate (with the ester bond as the only polar moiety), the allowed mutations were only to hydrophobic residues (A, F, I, L, P, V, W, Y). Moreover, sequence restraints were placed on the selected mutable residues to avoid considerable differences from the WT enzyme, applying an energy penalty to favor sequence conservation.

Due to the chirality of the substrate, two separate MC simulations were run for the (R)- and (S)-enantiomers of 1-phenyl-2-pentyl acetate. Thus, mutants evolving the active site for each enantiomer alone are generated. For both enantiomers, the top 50 variants were selected, and then, clustered based on the substrate's binding mode. Finally, variants were chosen based on the centroids of each cluster.

The selected variants were further examined through studying the enzyme-substrate interaction with PELE. In these simulations, the substrate was placed outside the active site's cavity and the binding was followed by counting catalytic poses and computing the average interaction energy once the ligand is inside the active site (*Figure 4.19*).

Overall, the variants showed more catalytic poses, as well as better average interaction energies than the WT enzyme (all plots and results can be found in *Chapter 3*). The 8 variants with the highest number of catalytic poses were selected for experimental testing by Bornscheuer's laboratory. Their validation showed that 3 mutants improved the hydrolysis of 1-phenyl-2-pentyl acetate with reversed enantioselectivity (from (R) to (S) enantiomer).
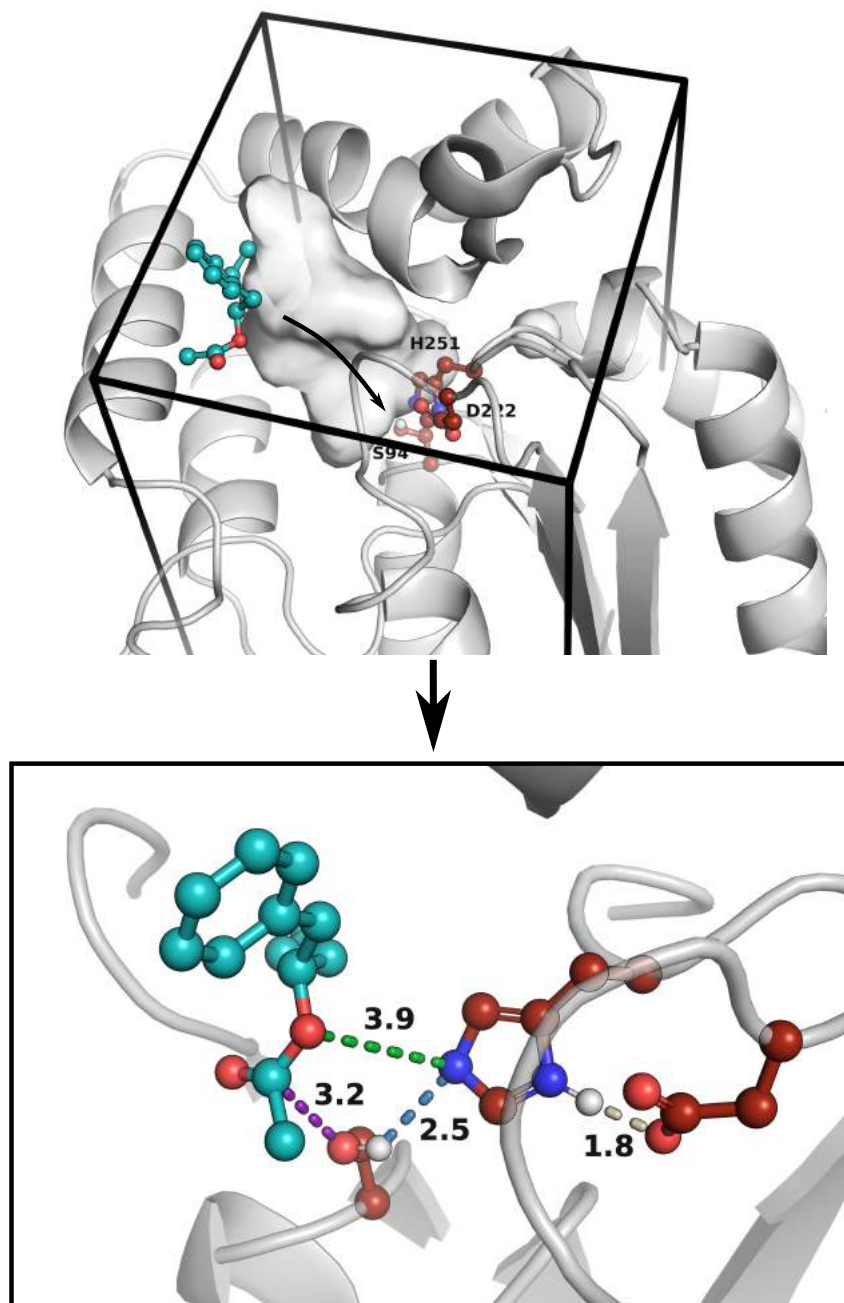
FIGURE 4.19: Initial setup for PELE simulations and the representation of a catalytic pose. The substrate is placed outside the active site and allowed to explore around the drawn box (top). A catalytic pose is represented with every key distance highlighted in a different color (blue for serine-histidine, beige for aspartate-histidine, violet for serine-substrate, and green for histidine-substrate) (bottom). The catalytic triad residues are colored in red, and the substrate in cyan.

5 | Conclusions

To sum up, this thesis has proven the potential of pluriZymes and the value of computational methods for enzyme engineering. Enzymes with multiple active sites could be a promising idea to be used in the industry soon to make processes more sustainable and cheaper. Moreover, the shown results in the thesis ease the process of developing an extra active site and allow for its enhancement according to the user's needs.

Therefore, the main conclusions of the thesis are:

- **The design of an enzyme with more than two active sites is feasible**: A protease site was successfully designed in a pluriZyme, having the three sites functioning simultaneously. The new active site allowed the serine esterase to catalyze both the hydrolysis of ester and peptide bonds, having a potential interest for the detergent industry. Likewise, both protease and esterase sites were combined for the one-pot cascade synthesis of L-histidine methyl ester, an intermediate for the design of Schiff base ligands.

- **A pluriZyme for the pharmaceutical industry**: A hydrolase site was incorporated in a metagenomic ω-TA, allowing the synthesis of enantiopure β-amino acids from β-keto esters in a single protein scaffold. Thus, we proved that one enzyme could perform compelling cascade reactions by synthesizing 3-ATfBA, a crucial intermediate to manufacture a family of antidiabetic drugs. Besides, this design proved that pluriZymes can be conceived in other enzyme families.

- **Substrate promiscuity of serine hydrolases can be rationalized and customized**: The computational study allowed us to accurately identify what makes a serine hydrolase highly promiscuous or not. Subsequently, this knowledge was used to transform a low-promiscuous esterase into a prominent one without compromising its average turnover rates. This example shows how artificial catalytic triads of pluriZymes could be enhanced toward hydrolyzing a broader range of substrates.

- **AsiteDesign: a new approach to automate and ease the design of active sites**: A semi-rational algorithm was developed to generate and rank possible vari-

ants to functionalize a site. In addition, the method was capable of generating a smart library of mutants to enhance the property of an existing active site. Benchmarking the protocol experimentally with an esterase allowed a demonstration of its potential.

Bibliography

Agarwal, P. K. (2006). Enzymes: An integrated view of structure, dynamics and function. *Microbial Cell Factories*, 5(1):1–12.

Agresti, J. J., Antipov, E., Abate, A. R., Ahn, K., Rowat, A. C., Baret, J.-C., Marquez, M., Klibanov, A. M., Griffiths, A. D., and Weitz, D. A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences*, 107(9):4004–4009.

Alonso, S., Santiago, G., Cea-Rama, I., Fernandez-Lopez, L., Coscolín, C., Modregger, J., Ressmann, A. K., Martínez-Martínez, M., Marrero, H., Bargiela, R., Pita, M., Gonzalez-Alfonso, J. L., Briand, M. L., Rojo, D., Barbas, C., Plou, F. J., Golyshin, P. N., Shahgaldian, P., Sanz-Aparicio, J., Guallar, V., and Ferrer, M. (2019). Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. *Nature Catalysis*, 3(3):319–328.

Althoff, E., Wang, L., Jiang, L., Giger, L., Lassila, J., Wang, Z., Smith, M., Hari, S., Kast, P., Herschlag, D., Hilvert, D., and Baker, D. (2012). Robust design and optimization of retroaldol enzymes. *Protein Science*, 21(5):717–726.

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*, 80(1):505–515.

Austin, H. P., Allen, M. D., Donohoe, B. S., Rorrer, N. A., Kearns, F. L., Silveira, R. L., Pollard, B. C., Dominick, G., Duman, R., Omari, K. E., Mykhaylyk, V., Wagner, A., Michener, W. E., Amore, A., Skaf, M. S., Crowley, M. F., Thorne, A. W., Johnson, C. W., Woodcock, H. L., McGeehan, J. E., and Beckham, G. T. (2018). Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proceedings of the National Academy of Sciences*, 115(19):E4350–E4357.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., Dijk, A. A. v., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, page eabj8754.

Banks, J. L., Beard, H. S., Cao, Y., Cho, A. E., Damm, W., Farid, R., Felts, A. K., Halgren, T. A., Mainz, D. T., Maple, J. R., Murphy, R., Philipp, D. M., Repasky, M. P., Zhang, L. Y., Berne, B. J., Friesner, R. A., Gallicchio, E., and Levy, R. M. (2005). Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of computational chemistry*, 26(16):1752–1780.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011). The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*, 50(21):4402–4410.

Barrozo, A., Borstnar, R., Marloie, G., and Kamerlin, S. C. L. (2012). Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution. *International Journal of Molecular Sciences*, 13(10):12428–12460.

Bayly, C. I., Merz, K. M., Ferguson, D. M., Cornell, W. D., Fox, T., Caldwell, J. W., Kollman, P. A., Cieplak, P., Gould, I. R., and Spellmeyer, D. C. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19):5179–5197.

Bell, E. L., Finnigan, W., France, S. P., Green, A. P., Hayes, M. A., Hepworth, L. J., Lovelock, S. L., Niikura, H., Osuna, S., Romero, E., Ryan, K. S., Turner, N. J., and Flitsch, S. L. (2021). Biocatalysis. *Nature Reviews Methods Primers*, 1(1):1–21.

Benavente, R., Lopez-Tejedor, D., Morales, M. d. P., Perez-Rizquez, C., and Palomo, J. M. (2020). The enzyme-induced formation of iron hybrid nanostructures with different morphologies. *Nanoscale*, 12(24):12917–12927.

Benavente, R., Lopez-Tejedor, D., and Palomo, J. (2018). Synthesis of a superparamagnetic ultrathin FeCO_{3} nanorods-enzyme bionanohybrid as a novel heterogeneous catalyst. *Chemical Communications*, 54(49):6256–6259.

Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56.

Bezborodov, A. M. and Zagustina, N. A. (2016). Enzymatic biocatalysis in chemical synthesis of pharmaceuticals (Review). *Applied Biochemistry and Microbiology 2016 52:3*, 52(3):237–249.

Bornscheuer, U. T., Altenbuchner, J., and Meyer, H. H. (1998). Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epothilones. *Biotechnology and Bioengineering*, 58(5):554–559.

Bos, J., Browne, W. R., Driessen, A. J. M., and Roelfes, G. (2015). Supramolecular Assembly of Artificial Metalloenzymes Based on the Dimeric Protein LmrR as Promiscuous Scaffold. *Journal of the American Chemical Society*, 137(31):9796–9799.

Bos, J., Fusetti, F., Driessen, A. J. M., and Roelfes, G. (2012). Enantioselective Artificial Metalloenzymes by Creation of a Novel Active Site at the Protein Dimer Interface. *Angewandte Chemie International Edition*, 51(30):7472–7475.

Branneby, C., Carlqvist, P., Magnusson, A., Hult, K., Brinck, T., and Berglund, P. (2003). Carbon-carbon bonds by hydrolytic enzymes. *Journal of the American Chemical Society*, 125(4):874–875.

Breaker, R. R. (1997). DNA enzymes. *Nature biotechnology*, 15(5):427–431.

Breaker, R. R. and Joyce, G. F. (2014). The Expanding View of RNA and DNA Function. *Chemistry & Biology*, 21(9):1059–1065.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217.

Buchholz, P. C., Feuerriegel, G., Zhang, H., Perez-Garcia, P., Nover, L. L., Chow, J., Streit, W. R., and Pleiss, J. (2022). Plastics degradation by hydrolytic enzymes: The plastics-active enzymes database—PAZy. *Proteins: Structure, Function, and Bioinformatics*, 90(7):1443–1456.

Cadwell, R. C. and Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *Genome Research*, 2(1):28–33.

Casas-Godoy, L., Duquesne, S., and Bordes, F. (2012). Lipases and Phospholipases. *Methods Mol Biol*, 861:3–30.

Chen, K. and Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences*, 90(12):5618–5622.

Chen, R. (1999). A general strategy for enzyme engineering. *Trends in biotechnology*, 17(9):344–345.

Chen, R. (2001). Enzyme engineering: rational redesign versus directed evolution. *Trends in biotechnology*, 19(1):13–14.

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

Chu, C.-T., Parris, D. S., Dixon, R. A., Farber, F. E., and Schaffer, P. A. (1979). Hydroxylamine mutagenesis of HSV DNA and DNA fragments: Introduction of mutations into selected regions of the viral genome - ScienceDirect. *Virology*, 98(1):168–181.

Cohen, S. N., Chang, A. C. Y., Boyer, H. W., and Helling, R. B. (1973). Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 70(11):3240.

Coscolín, C., Katzke, N., García-Moyano, A., Navarro-Fernández, J., Almendral, D., Martínez-Martínez, M., Bollinger, A., Bargiela, R., Gertler, C., Chernikova, T. N., Rojo, D., Barbas, C., Tran, H., Golyshina, O. V., Koch, R., Yakimov, M. M., Bjerga, G. E., Golyshin, P. N., Jaeger, K. E., and Ferrer, M. (2019). Bioprospecting Reveals

Class III ω-Transaminases Converting Bulky Ketones and Environmentally Relevant Polyamines. *Applied and environmental microbiology*, 85(2).

Cox, E. C. (1976). Bacterial mutator genes and the control of spontaneous mutation. *Annual Review of Genetics*, 10:135–156.

Craik, C., Largman, C., Fletcher, T., Roczniak, S., Barr, P., Fletterick, R., and Rutter, W. (1985). Redesigning trypsin: alteration of substrate specificity. *Science*, 228(4697):291–297.

Cramer, C. J. and Truhlar, D. G. (1999). Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chemical Reviews*, 99(8):2161–2200.

Cui, Y., Chen, Y., Liu, X., Dong, S., Tian, Y., Qiao, Y., Mitra, R., Han, J., Li, C., Han, X., Liu, W., Chen, Q., Wei, W., Wang, X., Du, W., Tang, S., Xiang, H., Liu, H., Liang, Y., Houk, K. N., and Wu, B. (2021). Computational Redesign of a PETase for Plastic Biodegradation under Ambient Condition by the GRAPE Strategy. *ACS Catalysis*, 11(3):1340–1350.

Dayanandan, A., Kanagaraj, J., Sounderraj, L., Govindaraju, R., and Suseela Rajkumar, G. (2003). Application of an alkaline protease in leather processing: an ecofriendly approach. *Journal of Cleaner Production*, 11(5):533–536.

Degani, O., Gepstein, S., and Dosoretz, C. G. (2002). Potential use of cutinase in enzymatic scouring of cotton fiber cuticle. *Applied Biochemistry and Biotechnology 2002 102:1*, 102(1):277–289.

DeGrado, W., Prendergast, F., Wolfe, H., and Cox, J. (1985). The design, synthesis, and characterization of tight-binding inhibitors of calmodulin. *Journal of cellular biochemistry*, 29(2):83–93.

Delagrave, S., Murphy, D. J., Pruss, J. L. R., Maffia, A. M., Marrs, B. L., Bylina, E. J., Coleman, W. J., Grek, C. L., Dilworth, M. R., Yang, M. M., and Youvan, D. C. (2001). Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase. *Protein Engineering, Design and Selection*, 14(4):261–267.

Devine, P. N., Howard, R. M., Kumar, R., Thompson, M. P., Truppo, M. D., and Turner, N. J. (2018). Extending the application of biocatalysis to meet the challenges of drug development. *Nature Reviews Chemistry*, 2(12):409–421.

Diáz, L., Soler, D., Tresadern, G., Buyck, C., Perez-Benito, L., Saen-Oon, S., Guallar, V., and Soliva, R. (2020). Monte Carlo simulations using PELE to identify a protein–protein inhibitor binding site and pose. *RSC Advances*, 10(12):7058–7064.

Díaz-Caballero, M., Navarro, S., Nuez-Martínez, M., Peccati, F., Rodríguez-Santiago, L., Sodupe, M., Teixidor, F., and Ventura, S. (2020). pH-Responsive Self-Assembly of Amyloid Fibrils for Dual Hydrolase-Oxidase Reactions. *ACS Catalysis*, 11(2):595–607.

Dill, K. A. and MacCallum, J. L. (2012). The Protein-Folding Problem, 50 Years On. *Science*, 338(6110):1042–1046.

Dodson, G. and Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends in biochemical sciences*, 23(9):347–352.

Doudna, J. A. and Lorsch, J. R. (2005). Ribozyme catalysis: not different, just worse. *Nature Structural & Molecular Biology*, 12(5):395–402.

Eliot, A. C. and Kirsch, J. F. (2004). Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annual review of biochemistry*, 73:383–415.

Estell, D. A., Graycar, T. P., and Wells, J. A. (1985). Engineering an Enzyme by Site-directed Mutagenesis to Be Resistant to Chemical Oxidation. *Journal of Biological Chemistry*, 260(11):6518–6521.

Eyring, H. (1935). The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics*, 3(2):107–115.

Fasan, R., Meharenna, Y. T., Snow, C. D., Poulos, T. L., and Arnold, F. H. (2008). Evolutionary History of a Specialized P450 Propane Monooxygenase. *Journal of Molecular Biology*, 383(5):1069–1080.

Field, M. J., Bash, P. A., and Karplus, M. (1990). A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, 11(6):700–733.

Filice, M., Marciello, M., Morales, M. d. P., and Palomo, J. M. (2013). Synthesis of heterogeneous enzyme–metal nanoparticle biohybrids in aqueous media and their applications in C–C bond formation and tandem catalysis. *Chemical Communications*, 49(61):6876–6878.

Filice, M., Romero, O., Gutiérrez-Fernández, J., Rivas, B. d. l., Hermoso, J. A., and Palomo, J. M. (2015). Synthesis of a heterogeneous artificial metallolipase with chimeric catalytic activity. *Chemical Communications*, 51(45):9324–9327.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*, 27(3):2985–2993.

Friesner, R. A. and Guallar, V. (2005). Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annual Review of Physical Chemistry*, 56:389–427.

Ghislieri, D. and Turner, N. J. (2014). Biocatalytic approaches to the synthesis of enantiomerically pure chiral amines. *Topics in Catalysis*, 57(5):284–300.

Gilabert, J. F., Grebner, C., Soler, D., Lecina, D., Municoy, M., Gracia Carmona, O., Soliva, R., Packer, M. J., Hughes, S. J., Tyrchan, C., Hogner, A., and Guallar, V. (2019). PELE-MSM: A Monte Carlo Based Protocol for the Estimation of Absolute Binding Free Energies. *Journal of Chemical Theory and Computation*, 15(11):6243–6253.

Gilabert, J. F., Lecina, D., Estrada, J., and Guallar, V. (2018). Monte Carlo Techniques for Drug Design: The Success Case of PELE. In *Biomolecular Simulations in Structure-Based Drug Discovery*, pages 87–103. John Wiley & Sons, Ltd.

Gilbert, W. (1986). Origin of life: The RNA world. *Nature*, 319(6055):618–618.

Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. D. (1979). Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America*, 76(1):106.

Goldsmith, M. and Tawfik, D. (2017). Enzyme engineering: reaching the maximal catalytic efficiency peak. *Current opinion in structural biology*, 47:140–150.

Gonzalez-Perez, D., Molina-Espeja, P., Garcia-Ruiz, E., and Alcalde, M. (2014). Mutagenic Organized Recombination Process by Homologous In Vivo Grouping (MORPHING) for Directed Enzyme Evolution. *PLOS ONE*, 9(3):e90919.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 Pt 2):849–857.

Halgren, T. (2007). New Method for Fast and Accurate Binding-site Identification and Analysis. *Chemical Biology & Drug Design*, 69(2):146–148.

Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. *Journal of Chemical Information and Modeling*, 49(2):377–389.

Hammer, S. C., Kubik, G., Watkins, E., Huang, S., Minges, H., and Arnold, F. H. (2017). Anti-Markovnikov alkene oxidation by metal-oxo–mediated enzyme catalysis. *Science*, 358(6360):215–218.

Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F. H. (2009). A family of thermostable fungal cellulases created by structure-guided recombination. *Proceedings of the National Academy of Sciences*, 106(14):5610–5615.

Hellinga, H. W., Caradonna, J. P., and Richards, F. M. (1991). Construction of new ligand binding sites in proteins of known structure: II. Grafting of a buried transition metal binding site into *Escherichia coli* thioredoxin. *Journal of Molecular Biology*, 222(3):787–803.

Hoogkamp-Korstanje, J. A. and Roelofs-Willemse, J. (2000). Comparative in vitro activity of moxifloxacin against Gram-positive clinical isolates. *The Journal of antimicrobial chemotherapy*, 45(1):31–39.

Howard, A. E. and Kollman, P. A. (1988). An Analysis of Current Methodologies for Conformational Searching of Complex Molecules. *Journal of Medicinal Chemistry*, 31(9):1669–1675.

Huffman, M. A., Fryszkowska, A., Alvizo, O., Borra-Garske, M., Campos, K. R., Canada, K. A., Devine, P. N., Duan, D., Forstater, J. H., Grosser, S. T., Halsey, H. M., Hughes, G. J., Jo, J., Joyce, L. A., Kolev, J. N., Liang, J., Maloney, K. M., Mann, B. F., Marshall, N. M., McLaughlin, M., Moore, J. C., Murphy, G. S., Nawrat, C. C., Nazor, J., Novick, S., Patel, N. R., Rodriguez-Granillo, A., Robaire, S. A., Sherer, E. C., Truppo, M. D., Whittaker, A. M., Verma, D., Xiao, L., Xu, Y., and Yang, H. (2019). Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science*, 366(6470):1255–1259.

Hughes, D. L. (2018). Biocatalysis in Drug Development—Highlights of the Recent Patent Literature. *Organic Process Research & Development*, 22(9):1063–1080.

Hurley, J. H., Baase, W. A., and Matthews, B. W. (1992). Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. *Journal of Molecular Biology*, 224(4):1143–1159.

Hutchison 3rd, C. A., Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smithy, M. (1978). Mutagenesis at a Specific Position in a DNA Sequence. *Journal of Biological Chemistry*, 253(18):6551–6560.

Ihara, Y., Nango, M., Kimura, Y., and Kuroki, N. (1983). Multifunctional Micellar Catalysis as a Model of Enzyme Action. *Journal of the American Chemical Society*, 105(5):1252–1255.

Illergård, K., Ardell, D., and Elofsson, A. (2009). Structure is three to ten times more conserved than sequence–a study of structural response in protein cores. *Proteins*, 77(3):499–508.

Jakočiūnas, T., Pedersen, L. E., Lis, A. V., Jensen, M. K., and Keasling, J. D. (2018). CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metabolic Engineering*, 48:288–296.

Jemli, S., Ayadi-Zouari, D., Hlima, H., and Bejar, S. (2016). Biocatalysts: application and engineering for industrial purposes. *Critical reviews in biotechnology*, 36(2):246–258.

Jemli, S., Ayadi-Zouari, D., Hlima, H. B., and Bejar, S. (2014). Biocatalysts: application and engineering for industrial purposes. *Critical reviews in biotechnology*, 36(2):246–258.

Jensen, F. (2007a). 12. Optimization techniques. In *Introduction to computational chemistry*, pages 380–420. John Wiley & Sons, Chichester, England [etc.].

Jensen, F. (2007b). 14. Simulation Techniques. In *Introduction to computational chemistry*, pages 445–486. John Wiley & Sons, Chichester, England [etc.].

Jensen, F. (2007c). 2. Force Field Methods. In *Introduction to computational chemistry*, pages 22–79. John Wiley & Sons, Chichester, England [etc.].

Jensen, F. (2007d). 3. Electronic Structure Methods: Independent-Particle Models. In *Introduction to computational chemistry*, pages 80–132. John Wiley & Sons, Chichester, England [etc.].

Jensen, F. (2007e). 6. Density Functional Methods. In *Introduction to computational chemistry*, pages 232–267. John Wiley & Sons, Chichester, England [etc.].

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., Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008). De Novo Computational Design of Retro-Aldol Enzymes. *Science*, 319(5868):1387–1391.

Johnson, K. A. and Goody, R. S. (2011). The original Michaelis constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry*, 50(39):8264–8269.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935.

Jorgensen, W. L. and Tirado-Rives, J. (2005). Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *Journal of Computational Chemistry*, 26(16):1689–1700.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, pages 1–11.

Kahana, A. and Lancet, D. (2021). Self-reproducing catalytic micelles as nanoscopic protocell precursors. *Nature Reviews Chemistry 2021 5:12*, 5(12):870–878.

Kamerlin, S. and Warshel, A. (2010). At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins*, 78(6):1339–1375.

Kamp, M. W. v. d. and Mulholland, A. J. (2013). Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry*, 52(16):2708–2728.

Kan, S. B. J., Lewis, R. D., Chen, K., and Arnold, F. H. (2016). Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science*, 354(6315):1048–1051.

Kelly, S. A., Pohle, S., Wharry, S., Mix, S., Allen, C. C., Moody, T. S., and Gilmore, B. F. (2018). Application of ω-Transaminases in the Pharmaceutical Industry. *Chemical Reviews*, 118(1):349–367.

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

Khalameyzer, V., Fischer, I., Bornscheuer, U. T., and Altenbuchner, J. (1999). Screening, nucleotide sequence, and biochemical characterization of an esterase from *Pseudomonas fluorescens* with high activity towards lactones. *Applied and environmental microbiology*, 65(2):477–482.

Khersonsky, O., Kiss, G., Röthlisberger, D., Dym, O., Albeck, S., Houk, K. N., Baker, D., and Tawfik, D. S. (2012). Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proceedings of the National Academy of Sciences*, 109(26):10358–10363.

Khobragade, T. P., Sarak, S., Pagar, A. D., Jeon, H., Giri, P., and Yun, H. (2021). Synthesis of Sitagliptin Intermediate by a Multi-Enzymatic Cascade System Using Lipase and Transaminase With Benzylamine as an Amino Donor. *Frontiers in bioengineering and biotechnology*, 9.

Kim, G. H., Jeon, H., Khobragade, T. P., Patil, M. D., Sung, S., Yoon, S., Won, Y., Choi, I. S., and Yun, H. (2019). Enzymatic synthesis of sitagliptin intermediate using a novel ω-transaminase. *Enzyme and Microbial Technology*, 120:52–60.

Kim, H. J., Ruszczycky, M. W., Choi, S.-h., Liu, Y.-n., and Liu, H.-w. (2011). Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature*, 473(7345):109–112.

Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery 2004 3:11*, 3(11):935–949.

Knowles, J. R. (1991). Enzyme catalysis: not different, just better. *Nature*, 350(6314):121–124.

Kohen, A. (2014). Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Accounts of Chemical Research*, 48(2):466–473.

Kontkanen, H., Tenkanen, M., Fagerström, R., and Reinikainen, T. (2004). Characterisation of steryl esterase activities in commercial lipase preparations. *Journal of Biotechnology*, 108(1):51–59.

Koshland, D. E. J. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 44(2):98.

Kraut, J. (1988). How do enzymes work? *Science*, 242(4878):533–540.

Kruger, K., Grabowski, P., Zaug, A., Sands, J., Gottschling, D., and Cech, T. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31(1):147–157.

Kuchner, O. and Arnold, F. H. (1997). Directed evolution of enzyme catalysts. *Trends in Biotechnology*, 15(12):523–530.

Labrou, N. (2010). Random Mutagenesis Methods for In Vitro Directed Enzyme Evolution. *Current Protein & Peptide Science*, 11(1):91–100.

Lane, T. J., Shukla, D., Beauchamp, K. A., and Pande, V. S. (2013). To milliseconds and beyond: challenges in the simulation of protein folding. *Current Opinion in Structural Biology*, 23(1):58–65.

Lecina, D., Gilabert, J. F., and Guallar, V. (2017). Adaptive simulations, towards interactive protein-ligand modeling. *Scientific Reports*, 7(1):8466.

Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., and 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):423–429.

Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., Qiu, J.-L., Chen, Y.-H., Li, J., and Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nature Biotechnology*, 38(7):875–882.

Lin, Y.-W., Nie, C.-M., and Liao, L.-F. (2012). Rational design of a nitrite reductase based on myoglobin: a molecular modeling and dynamics simulation study. *Journal of Molecular Modeling*, 18(9):4409–4415.

Liu, B., He, L., Wang, L., Li, T., Li, C., Liu, H., Luo, Y., and Bao, R. (2018). Protein Crystallography and Site-Direct Mutagenesis Analysis of the Poly(ethylene terephthalate) Hydrolase PETase from *Ideonella sakaiensis*. *ChemBioChem*, 19(14):1471–1475.

Lonsdale, R., N. Harvey, J., and J. Mulholland, A. (2012). A practical guide to modelling enzyme-catalysed reactions. *Chemical Society Reviews*, 41(8):3025–3038.

Lu, H., Diaz, D. J., Czarnecki, N. J., Zhu, C., Kim, W., Shroff, R., Acosta, D. J., Alexander, B. R., Cole, H. O., Zhang, Y., Lynd, N. A., Ellington, A. D., and Alper, H. S. (2022). Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature 2022 604:7907*, 604(7907):662–667.

Lutz, S. and Bornscheuer, U. T. (2011). *Protein Engineering Handbook, Volume 1 & Volume 2*. Wiley-VCH.

Lutz, S. and Iamurri, S. (2018). Protein Engineering: Past, Present, and Future. *Methods in molecular biology*, 1685:1–12.

Ma, Y., Yao, M., Li, B., Ding, M., He, B., Chen, S., Zhou, X., and Yuan, Y. (2018). Enhanced Poly(ethylene terephthalate) Hydrolase Activity by Protein Engineering. *Engineering*, 4(6):888–893.

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. *Journal of Computer-Aided Molecular Design*, 27(3):221–234.

Malik, M. S., Park, E. S., and Shin, J. S. (2012). Features and technical applications of ω-transaminases. *Applied Microbiology and Biotechnology*, 94(5):1163–1171.

Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Eguizabal, E. R., Guerrero, M. d. C., Yakunin, A. F., and Ferrer, M. (2013). Biochemical Diversity of Carboxyl Esterases and Lipases from Lake Arreo (Spain): a Metagenomic Approach. *Applied and Environmental Microbiology*, 79(12):3553.

Martínez-Martínez, M., Coscolín, C., Santiago, G., Chow, J., Stogios, P. J., 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., Pérez-García, P., Hai, T., Del Pozo, M. V., Stokke, R., Steen, I. H., Cui, H., Xu, X., Nocek, B. P., Alcaide, M., Distaso, M., Mesa, V., Peláez, A. I., Sánchez, J., Buchholz, P. C., Pleiss, J., Fernández-Guerra, A., 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., and Ferrer, M. (2018). Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chemical Biology*, 13(1):225–234.

Maurer, K. H. (2004). Detergent proteases. *Current Opinion in Biotechnology*, 15(4):330–334.

Mazurenko, S., Prokop, Z., and Damborsky, J. (2019). Machine Learning in Enzyme Engineering. *ACS Catalysis*, 10(2):1210–1223.

McCullum, E., Williams, B., Zhang, J., and Chaput, J. (2010). Random mutagenesis by error-prone PCR. *Methods in molecular biology*, 634:103–109.

McIntosh, J. A. and Owens, A. E. (2021). Enzyme engineering for biosynthetic cascades. *Current Opinion in Green and Sustainable Chemistry*, 29:100448.

Miao, Y., Metzner, R., and Asano, Y. (2017). Kemp Elimination Catalyzed by Naturally Occurring Aldoxime Dehydratases. *ChemBioChem*, 18(5):451–454.

Moore, J. C. and Arnold, F. H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature*, 14(4):458–467.

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American*, 262(4):56–65.

Municoy, M., Roda, S., Soler, D., Soutullo, A., and Guallar, V. (2020). AquaPELE: A Monte Carlo-Based Algorithm to Sample the Effects of Buried Water Molecules in Proteins. *Journal of Chemical Theory and Computation*, 16(12):7655–7670.

Nardini, M. and Dijkstra, B. W. (1999). Alpha/beta hydrolase fold enzymes: the family keeps growing. *Current opinion in structural biology*, 9(6):732–737.

Nazor, J., Liu, J., and Huisman, G. (2021). Enzyme evolution for industrial biocatalytic cascades. *Current Opinion in Biotechnology*, 69:182–190.

Neet, K. (1998). Enzyme catalytic power minireview series. *Journal of Biological Chemistry*, 273(40):25527–25528.

Nelson, D. L. and Cox, M. M. (2008a). 1. The foundaments of biochemistry. In *Lehninger Principles of Biochemistry 5th ed.*, pages 1–43. W.H. Freeman.

Nelson, D. L. and Cox, M. M. (2008b). 2. Water. In *Lehninger Principles of Biochemistry 5th ed.*, pages 47–74. W.H. Freeman.

Nelson, D. L. and Cox, M. M. (2008c). 3. Amino Acids, Peptides and Proteins. In *Lehninger Principles of Biochemistry 5th ed.*, pages 71–112. W.H. Freeman.

Nelson, D. L. and Cox, M. M. (2008d). 4. The three-dimensional structure of proteins. In *Lehninger Principles of Biochemistry 5th ed.*, pages 116–156. W.H. Freeman.

Nelson, D. L. and Cox, M. M. (2008e). 6. Enzymes. In *Lehninger Principles of Biochemistry 5th ed.*, pages 190–237. W.H. Freeman.

Ogawa, R., Sunatsuki, Y., and Suzuki, T. (2018). Schiff Base Ligands Derived from l-Histidine Methyl Ester: Characterization, Racemization, and Dimerization of Their Transition-Metal Complexes. *European Journal of Inorganic Chemistry*, 2018(16):1733–1742.

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., and Schrag, J. (1992). The alpha/beta hydrolase fold. *Protein engineering*, 5(3):197–211.

Palomo, J. M. (2019). Nanobiohybrids: a new concept for metal nanoparticles synthesis. *Chemical Communications*, 55(65):9583–9589.

Perez, C., Soler, D., Soliva, R., and Guallar, V. (2020). FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design. *Journal of chemical information and modeling*, 60(3):1728–1736.

Petrović, D., Risso, V. A., Kamerlin, S. C. L., and Sanchez-Ruiz, J. M. (2018). Conformational dynamics and enzyme evolution. *Journal of the Royal Society Interface*, 15(144).

Prakash, D., Nawani, N., Prakash, M., Bodas, M., Mandal, A., Khetmalas, M., and Kapadnis, B. (2013). Actinomycetes: A repertory of green catalysts with a potential revenue resource. *BioMed Research International*, 2013.

Ranaghan, K. E. and Mulholland, A. J. (2010). Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods. *International Reviews in Physical Chemistry*, 29(1):65–133.

Rauwerdink, A. and Kazlauskas, R. J. (2015). How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS catalysis*, 5(10):6153–6176.

Robinson, P. (2015). Enzymes: principles and biotechnological applications. *Essays in biochemistry*, 59:1–41.

Roda, S., Robles-Martín, A., Xiang, R., Kazemi, M., and Guallar, V. (2021). Structural-Based Modeling in Protein Engineering. A Must Do. *The Journal of Physical Chemistry B*, 125(24):6491–6500.

Roda, S., Santiago, G., and Guallar, V. (2020). Mapping enzyme-substrate interactions: its potential to study the mechanism of enzymes. *Advances in Protein Chemistry and Structural Biology*, 122:1–31.

Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004). Protein Structure Prediction Using Rosetta. *Methods in Enzymology*, 383:66–93.

Romero, P. A. and Arnold, F. H. (2009). Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12):866–876.

Rosenboom, J. G., Langer, R., and Traverso, G. (2022). Bioplastics for a circular economy. *Nature Reviews Materials 2022 7:2*, 7(2):117–137.

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering, Design and Selection*, 12(2):85–94.

Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008). Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–195.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5463.

Santiago, G., de Salas, F., Lucas, M. F., Monza, E., Acebes, S., Martinez, A. T., Camarero, S., and Guallar, V. (2016). Computer-Aided Laccase Engineering: Toward Biological Oxidation of Arylamines. *ACS Catalysis*, 6(8):5415–5423.

Santiago, G., Martínez-Martínez, M., Alonso, S., Bargiela, R., Coscolín, C., Golyshin, P. N., Guallar, V., and Ferrer, M. (2018). Rational Engineering of Multiple Active Sites in an Ester Hydrolase. *Biochemistry*, 57(15):2245–2255.

Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. (2010). Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science*, 329(5989):305–309.

Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Lerardi, D. J., Kolossváry, I., Klepeis, J. L., Layman, T., McLeavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y., Spengler, J., Theobald, M., Towles, B., and Wang, S. C. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*, 51(7):91–97.

Shaw, D. E., Grossman, J. P., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., Deneroff, M. M., Dror, R. O., Even, A., Fenton, C. H., Forte, A., Gagliardo, J., Gill, G., Greskamp, B., Ho, C. R., Ierardi, D. J., Iserovich, L., Kuskin, J. S., Larson, R. H., Layman, T., Lee, L. S., Lerer, A. K., Li, C., Killebrew, D., Mackenzie, K. M., Mok, S. Y. H., Moraes, M. A., Mueller, R., Nociolo, L. J., Peticolas, J. L., Quan, T., Ramot, D., Salmon, J. K., Scarpazza, D. P., Ben Schafer, U., Siddique, N., Snyder, C. W., Spengler, J., Tang, P. T. P., Theobald, M., Toma, H., Towles, B., Vitale, B., Wang, S. C., and Young, C. (2014). Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*, 2015-January(January):41–53.

Shen, J. W., Qi, J. M., Zhang, X. J., Liu, Z. Q., and Zheng, Y. G. (2018). Significantly increased catalytic activity of *Candida antarctica* lipase B for the resolution of cis -(±)-dimethyl 1-acetylpiperidine-2,3-dicarboxylate. *Catalysis Science & Technology*, 8(18):4718–4725.

Shen, R., Crean, R. M., Johnson, S. J., Kamerlin, S. C. L., and Hengge, A. C. (2021). Single Residue on the WPD-Loop Affects the pH Dependency of Catalysis in Protein Tyrosine Phosphatases. *JACS Au*, 1(5):646–659.

Shortle, D. and Botstein, D. (1983). Directed mutagenesis with sodium bisulfite. *Methods in enzymology*, 100(C):457–468.

Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St.Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. (2010). Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science*, 329(5989):309–313.

Silverman, S. K. (2004). Deoxyribozymes: DNA catalysts for bioorganic chemistry. *Organic & Biomolecular Chemistry*, 2(19):2701–2706.

Simon, G. M. and Cravatt, B. F. (2010). Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. *Journal of Biological Chemistry*, 285(15):11051–11055.

Smith, G. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705):1315–1317.

Smith, G. and Petrenko, V. A. (1997). Phage Display. *Chemical Reviews*, 97(2):391–410.

Son, H. F., Cho, I. J., Joo, S., Seo, H., Sagong, H. Y., Choi, S. Y., Lee, S. Y., and Kim, K. J. (2019). Rational Protein Engineering of Thermo-Stable PETase from *Ideonella sakaiensis* for Highly Efficient PET Degradation. *ACS Catalysis*, 9(4):3519–3526.

Sousa, S., Ribeiro, A., Coimbra, J., Neves, R., Martins, S., Moorthy, N., Fernandes, P., and Ramos, M. (2013). Protein-ligand docking in the new millennium–a retrospective of 10 years in the field. *Current medicinal chemistry*, 20(18):2296–2314.

Sperl, J. M. and Sieber, V. (2018). Multienzyme Cascade Reactions—Status and Recent Advances. *ACS Catalysis*, 8(3):2385–2396.

Steiner, E. (2004). Charge Densities in Atoms. *The Journal of Chemical Physics*, 39(9):2365.

Steiner, K. and Schwab, H. (2012). Recent advances in rational approaches for enzyme engineering. *Computational and structural biotechnology journal*, 2:e201209010.

Sumbalova, L., Stourac, J., Martinek, T., Bednar, D., and Damborsky, J. (2018). HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Research*, 46(W1):W356–W362.

Sutcliffe, M. and Scrutton, N. (2000). Enzymology takes a quantum leap forward. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 358(1766):367–386.

Takahashi, R., Gil, V. A., and Guallar, V. (2014). Monte carlo free ligand diffusion with markov state model analysis and absolute binding free energy calculations. *Journal of Chemical Theory and Computation*, 10(1):282–288.

Taylor, J. W., Eghtesadi, S. A., Points, L. J., Liu, T., and Cronin, L. (2017). Autonomous model protocell division driven by molecular replication. *Nature Communications 2017 8:1*, 8(1):1–6.

Tournier, V., Topham, C. M., Gilles, A., David, B., Folgoas, C., Moya-Leclair, E., Kamionka, E., Desrousseaux, M.-L., Texier, H., Gavalda, S., Cot, M., Guémard, E., Dalibey, M., Nomme, J., Cioci, G., Barbe, S., Chateau, M., André, I., Duquesne, S., and Marty, A. (2020). An engineered PET depolymerase to break down and recycle plastic bottles. *Nature*, 580(7802):216–219.

Truhlar, D. G. (2015). Transition state theory for enzyme kinetics. *Archives of biochemistry and biophysics*, 582:10.

Vertommen, M. A., Nierstrasz, V. A., Veer, M. V. D., and Warmoeskerken, M. M. (2005). Enzymatic surface modification of poly(ethylene terephthalate). *Journal of Biotechnology*, 120(4):376–386.

Warshel, A. and Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249.

Watt, A. P., Morrison, D., Locker, K. L., and Evans, D. C. (2000). Higher Throughput Bioanalysis by Automation of a Protein Precipitation Assay Using a 96-Well Format with Detection by LC-MS/MS. *Analytical Chemistry*, 72(5):979–984.

Wilkinson, A. J., Fersht, A. R., Blow, D. M., and Winter, G. (1983). Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. *Biochemistry*, 22(15):3581–3586.

Wilson, C., Mace, J., and Agard, D. (1991). Computational method for the design of enzymes with altered substrate specificity. *Journal of molecular biology*, 220(2):495–506.

Wilson, D. and Keefe, A. (2001). Random mutagenesis by PCR. *Current protocols in molecular biology*, Chapter 8.

Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M., and Smith, M. (1982). Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature*, 299(5885):756–758.

Wolfenden*, R. (2006). Degrees of Difficulty of Water-Consuming Reactions in the Absence of Enzymes. *Chemical Reviews*, 106(8):3379–3396.

Wolfenden*, R. and Snider, M. J. (2001). The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Accounts of Chemical Research*, 34(12):938–945.

Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K., and Bornscheuer, U. T. (2021). Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angewandte Chemie International Edition*, 60(1):88–119.

Xiao, H., Bao, Z., and Zhao, H. (2015). High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Industrial & Engineering Chemistry Research*, 54(16):4011.

Yang, G. and Withers, S. G. (2009). Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem*, 10(17):2704–2715.

Yeung, D. K. J., Gao, T., Huang, J., Sun, S., Guo, H., and Wang, J. (2013). Organocatalytic 1,3-dipolar cycloaddition reactions of ketones and azides with water as a solvent. *Green Chemistry*, 15(9):2384–2388.

Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., and Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, 351(6278):1196–1199.

Zanghellini, A. (2014). de novo computational enzyme design. *Current Opinion in Biotechnology*, 29(1):132–138.

Zhang, R. K., Chen, K., Huang, X., Wohlschlager, L., Renata, H., and Arnold, F. H. (2018). Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp 3 C–H functionalization. *Nature*, 565(7737):67–72.

Zhao, H. and Arnold, F. H. (1997). Optimization of DNA Shuffling for High Fidelity Recombination. *Nucleic Acids Research*, 25(6):1307–1308.

Zhao, L., Cai, J., Li, Y., Wei, J., and Duan, C. (2020). A host–guest approach to combining enzymatic and artificial catalysis for catalyzing biomimetic monooxygenation. *Nature Communications*, 11(1):1–11.

Zhou, Z. and Roelfes, G. (2020). Synergistic catalysis in an artificial enzyme by simultaneous action of two abiological catalytic sites. *Nature Catalysis*, 3(3):289–294.

Zinovjev, K. and Tuñón, I. (2017). Quantifying the limits of transition state theory in enzymatic catalysis. *Proceedings of the National Academy of Sciences*, 114(47):12390–12395.