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 08034, Spain

[2]Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28049, Spain.

[3]Nostrum Biodiscovery S.L., Barcelona 08028, Spain.

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

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 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 here identified (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 ones 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 point of view[8]. Still, 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 due to 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 to 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 towards 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 low substrate spectrum into one with 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 low substrate spectrum with, a priori, low biotechnological potential, into a highly substrate promiscuous one with higher applied potential.

In this work, we present a deep 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 later (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), CalB (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 (PELE) simulations. 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 20º when the SASA was bigger than

0.15, otherwise it was reduced to 5º. The 6 lowest ANM eigenvectors were linearly combined at random to move the protein. The side chains phase included all residues within 6 Å of the ligand. The Variable Dielectric Generalized-Born Non-Polar (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 as "Serine-substrate distance".

Molecular Dynamics (MD). 4 replicas of 500 ns of MD simulations with OPENMM[39] were performed to analyze the flexibility of the protein and particularly of the catalytic triad on selected systems. 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[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]. Andersen thermostat[42] and Monte Carlo barostat[43,44] were applied for the NPT ensemble (constant pressure and temperature, being 1 bar and 300K, respectively). The NVT equilibration lasted 400 ps and a constraint of 10 kcal/(mol·A^2) was applied to the system, while the NPT equilibration lasted 1 ns and a milder constraint of 5 kcal/(mol·A^2) was used. The Verlet integrator[45] with a 2 fs time step was used, using constraints between H and heavy atoms. For the non-bonded 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 to 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 making 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 $Grid_{phobic}$ (□□□□$_{□h□□□□}$ = □$_{□□□}$ −

$0.3 \cdot$ □□□□□□□□□−□□□□□□) or $Grid_{philic}$ (□□□□□$_{h}$□□□□ = □□□□ + □□□□□□□□□−□□□□□□)

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. The $\Delta\Delta G_{(mut\text{-}WT)}$ 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, 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[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; Ile92GlyTrp96GlyFwd: 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 determinations (k_{cat} and K_M). The vector pET46 Ek/LIC and the host *Escherichia 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 was 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-liter flask, following by cultivation at 37°C and 200 rpm overnight. Afterwards, 50 ml of this culture was used to inoculate 1-liter of LB-Amp medium in a 2.5-liter 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 re-suspended 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 15000 × *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 elution with 50 mM sodium phosphate, pH 8.0, 250 mM imidazole and 300 mM NaCl. Eluted protein was subjected to ultra-filtration through low-adsorption hydrophilic 10000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) to concentrate the protein solution. An extensive dialysis of protein solutions against 40 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 7.0) was then performed using Pur-A-LyzerTM Maxi 1200 dialysis kit (Sigma-Aldrich, MO, US), as follows. 2 ml concentrated protein solution was dialyzed against the 2-liter buffer during 1 hour at room temperature, after which the buffer was changed by another 2-liter buffer and maintained 1 hour 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 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-liter culture.

Substrate fingerprint and kinetic parameters determinations. 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 slights 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 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 pipette (Eppendorf, Hamburg, Germany). The total reaction volume was 44 µl. Ester hydrolysis was measured at 30°C in 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 non-enzymatic 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 4 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.

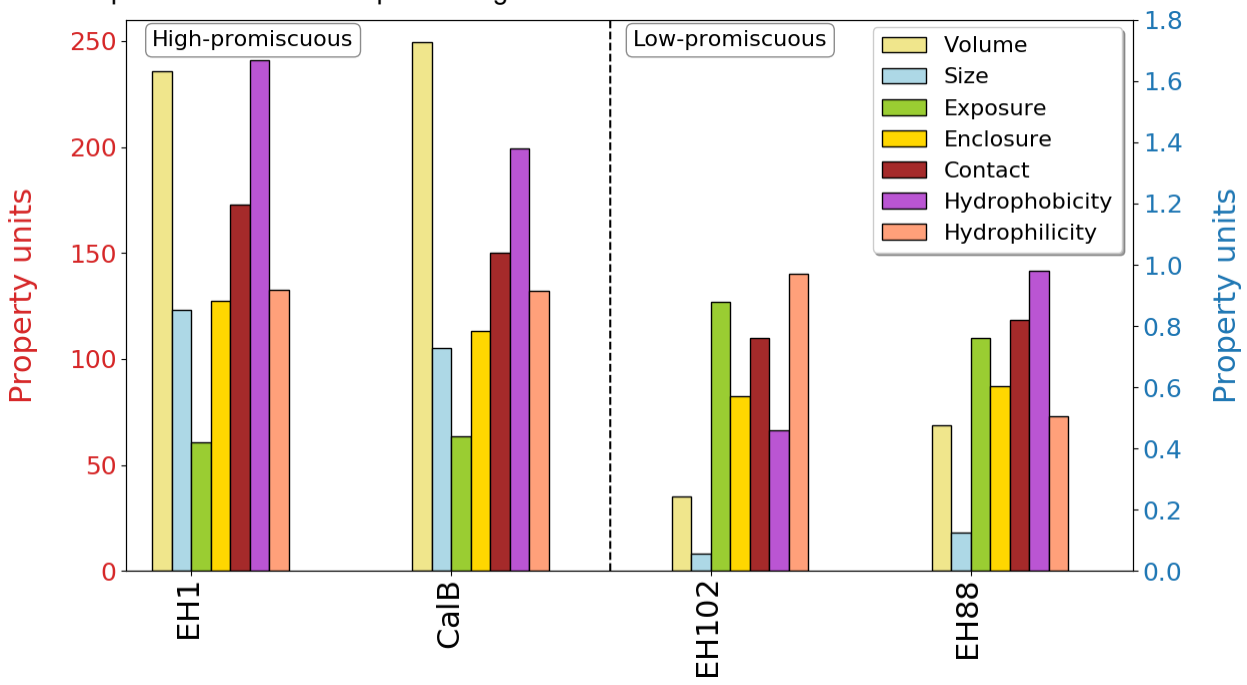
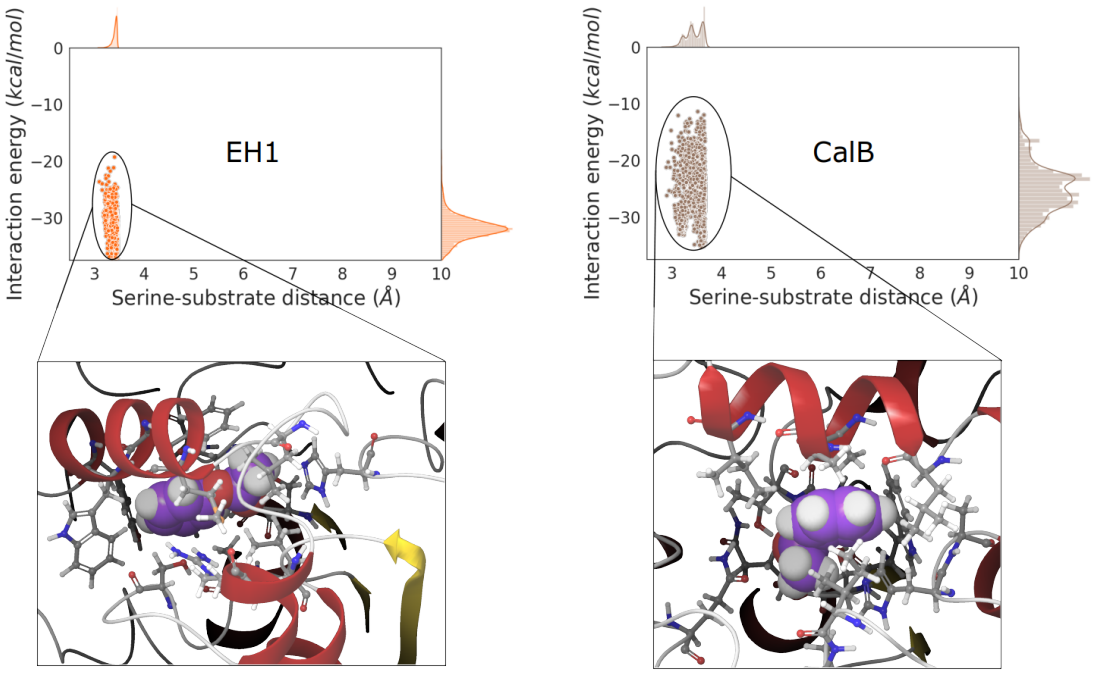


Figure 1. Bar plot representing the Volume, Size, Exposure, Enclosure, Contact, Hydrophobicity, and Hydrophilicity properties of the active site cavity of two high-substrate promiscuous and two low-substrate promiscuous esterases with SiteMap[47,48]. The Y red axis on the left represents the scale for 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].

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 Å^3 for EH1 and 249.70 Å^3 for CalB, while the low-promiscuous have volumes of 68.94 Å^3 for EH88 and 34.99 Å^3 for EH102. Thus, the esterases with broad substrate range have active site cavities 3- to 7-fold bigger 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 higher number of esters have large enough, well defined, hydrophobic, and compact cavities, sheltered from the solvent, whereas less 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 substrate range, and thus, substrate promiscuity in an esterase only hydrolyzing few esters.



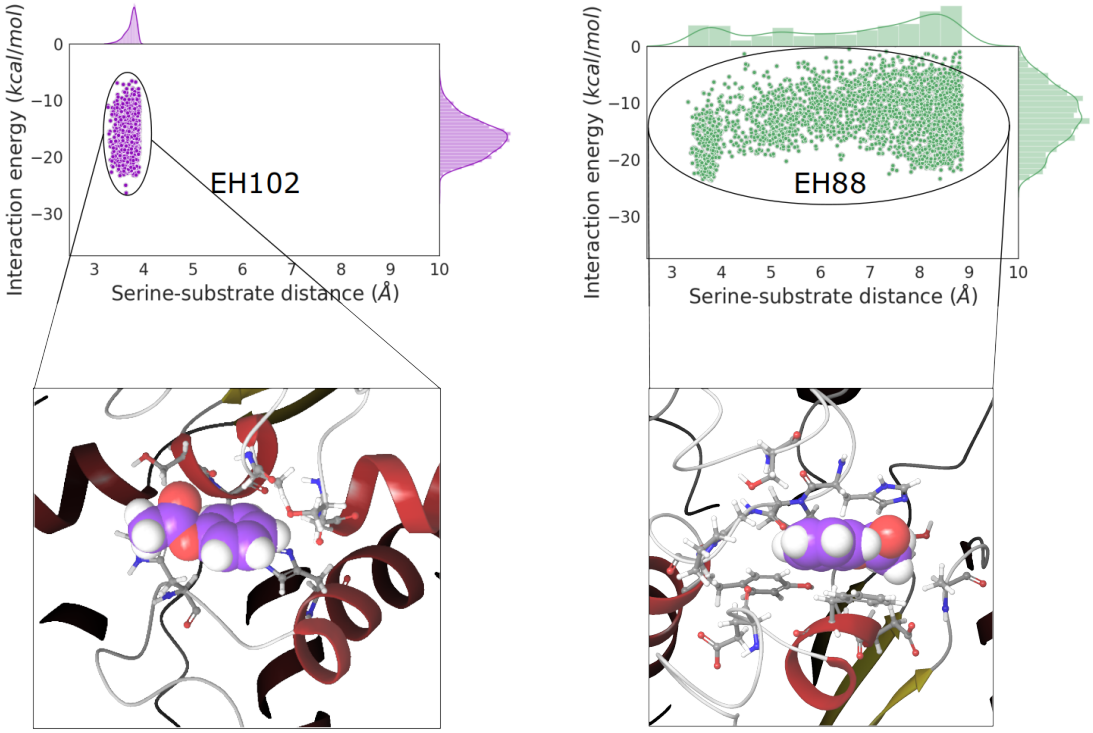


Figure 2. Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulations for the different studied esterases with phenyl acetate as substrate (3226 data points for CalB, 3095 for EH1, 2647 for EH102, and 2734 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[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).

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 due to 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 less 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, since 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 nomenclature).

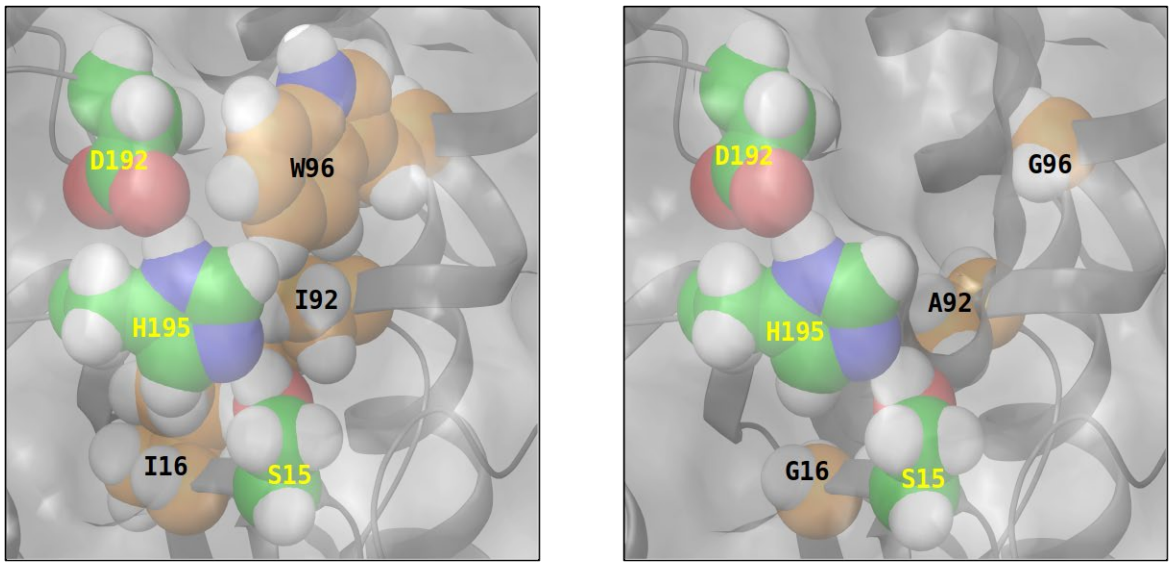


Figure 3. 3D representation of the surface of EH102$_{WT}$ (left) and EH102$_{TM1}$ (right). The mutated residues have the C atoms stained in orange and their labels in black, while the residues of the catalytic triad have them stained in green and the labels in yellow. All shown residues are displayed with the CPK model.

Table 1. Recopilation 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	-	-	-

To deepen into 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).

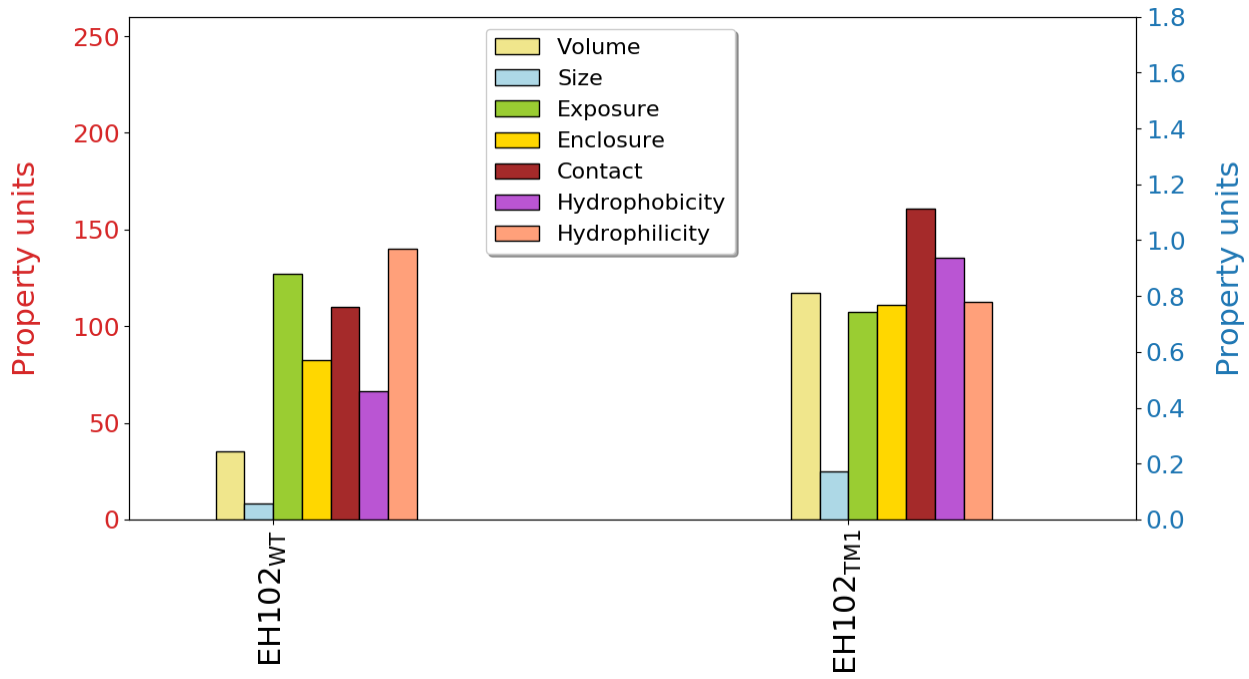


Figure 4. Bar plot representing the Volume, Size, Exposure, Enclosure, Contact, Hydrophobicity, and Hydrophilicity properties of the active site cavity of the $EH102_{WT}$ in comparison with $EH102_{TM1}$ obtained with SiteMap[47,48]. The Y red axis on the left represents the scale for the Volume (in $Å^3$) 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].

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 $Å^3$) 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.

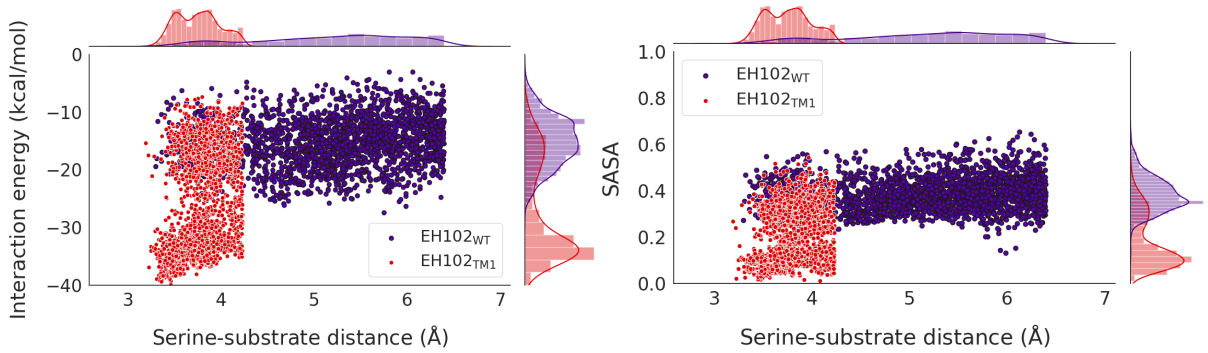


Figure 5. Density plots of the distribution of the catalytic serine-substrate distance against the interaction energy (left) and SASA (right) from the 10% lowest percentile regarding the serine-substrate distance of the accepted steps in the PELE simulation for $EH102_{WT}$ and $EH102_{TM1}$ with ethyl-3-oxohexanoate as substrate (2332 data points for $EH102_{WT}$ and 2225 for $EH102_{TM1}$). The density plots were created with the Matplotlib library[51].

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 the k_{cat} and $K_M$against the set of 96 structurally diverse esters (Figure 6), corroborated that substrate range was enhanced, going from hydrolyzing only 16 esters in the $EH102_{WT}$ to 35 esters in the $EH102_{TM1}$, 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 the main 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.

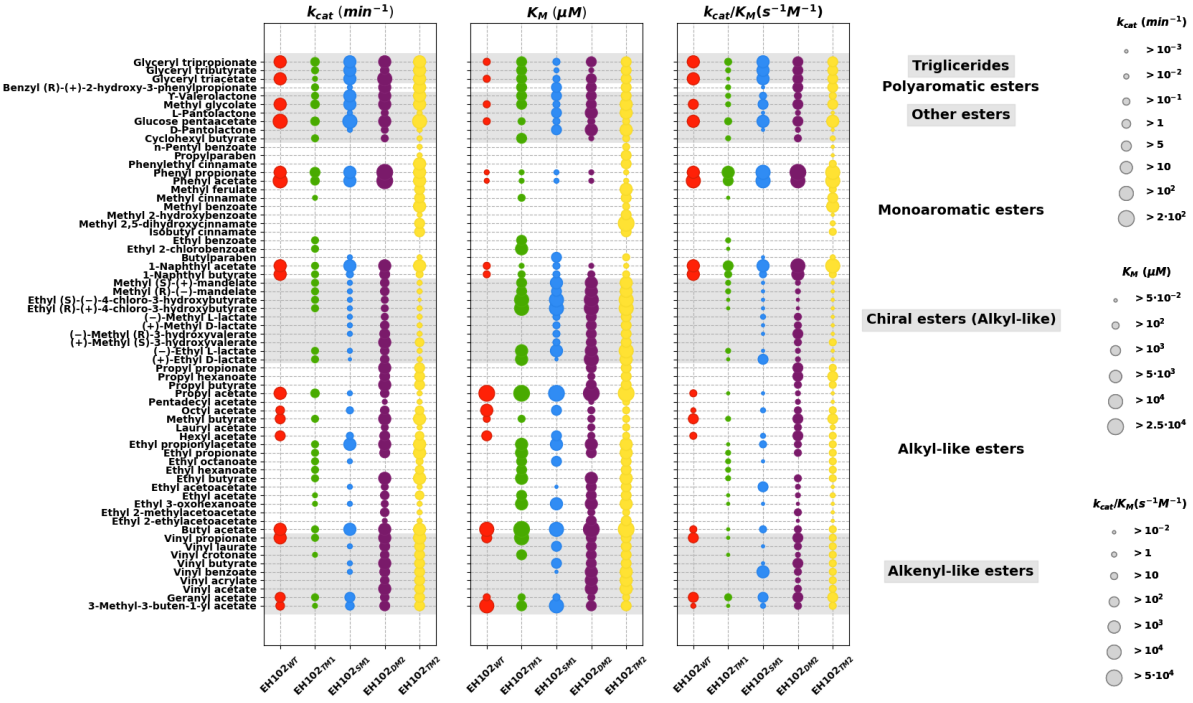


Figure 6. Substrate spectra of the $EH102_{WT}$, $EH102_{TM1}$, $EH102_{SM1}$, $EH102_{DM2}$, and $EH102_{TM2}$. The identification code of each variant can be found at the X axis, while the ester names are placed at the Y axis. The size of the points indicates a higher value for the represented property 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.

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 catalytic triad have significant larger values as well in the proposed variant (Figures S5-6). However, the RMSD of the protein backbone converges to similar values (Figure S7), indicating that the mutations do not destabilize considerably the fold. 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.

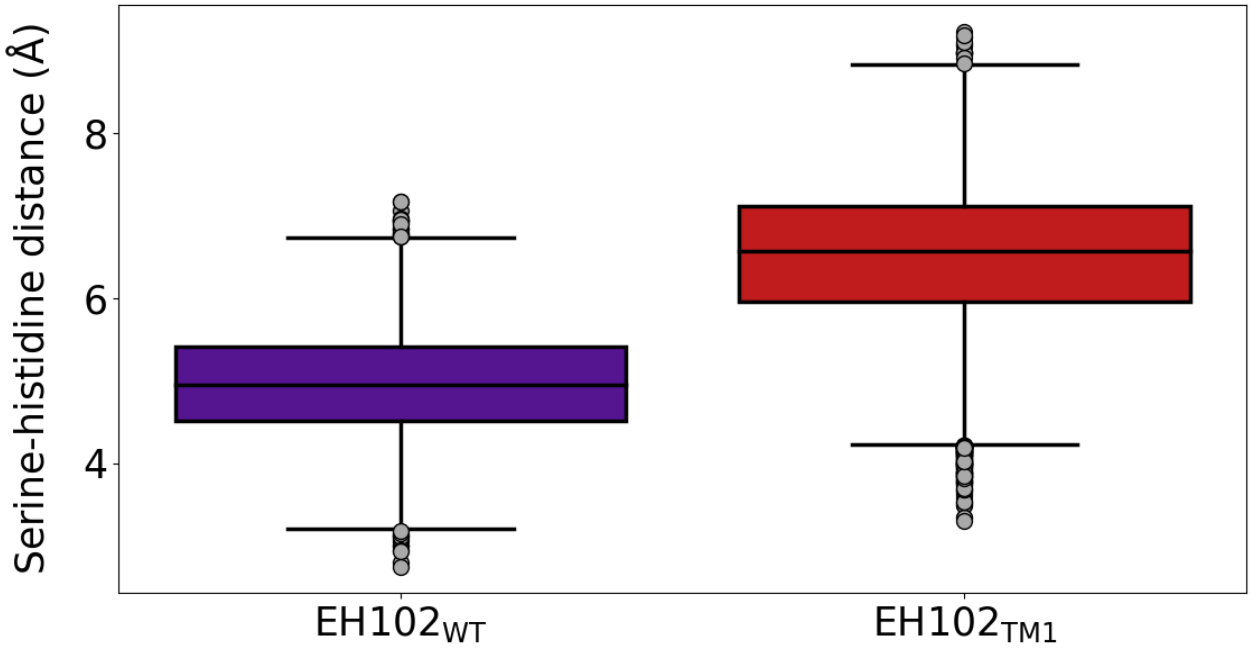


Figure 7. Box plot representing the serine-histidine distance ($d_{H_{\gamma}^{Ser}-N_{\varepsilon}^{His}}$) along the 500 ns of the 4 MD replicas performed for $EH102_{WT}$ and $EH102_{TM1}$. The figure was created with the Matplotlib library[51].

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 mutants-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 6 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, since 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 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 S9-11). 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 single mutation stability analysis with the MD results (Figures S9-11), suggests that residue 16 should be either substituted with a valine residue ($EH102_{SM1}$) or preserved, due to 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 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 volume lower than 200 $Å^3$, but one higher than 200 $Å^3$ which was vinyl laurate (k_{cat} 0.028 min^{-1}). Opposite, $EH102_{TM1}$ gained the capacity to convert 11 esters not hydrolyzed by the $EH102_{SM1}$, all of them being molecules with volume below 200 $Å^3$. In addition, differences were also observed at the level of catalytic turnover 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 $8.2 \cdot 10^{-5}$ 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}) is approaching 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 the k_{cat} for a few substrates (from 1.3 to 2.5-fold for 3 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 towards 11 esters that were not hydrolyzed neither by $EH102_{SM1}$ nor $EH102_{TM1}$. They all represent molecules with 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-methylacetoacetate, (+)-methyl (S)-3-hydroxyvalerate) and alkenyl- (vinyl acetate, 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 to hydrolyze 63 esters (Figure 6, Table S1), which represents 47, 28, 27 and 12 more esters than $EH102_{WT}$, $EH102_{TM1}$, $EH102_{SM1}$, and $EH102_{DM2}$, in the same order. In addition, the 3 mutations gained activity for 8 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, 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 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 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 for a hydrolase capable of hydrolyzing only 16 esters out of 96 tested, we have engineered 2 hydrolases which exhibit an impressive 51- ($EH102_{DM2}$) and 63- ($EH102_{TM2}$) substrate repertoire. Such 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 engineering 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 4 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, methyl ferulate) rarely hydrolyzed together, e.g. can only be observed in 3 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 similar substrate range. This was confirmed by selecting 6 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 model substrate (Table 2).

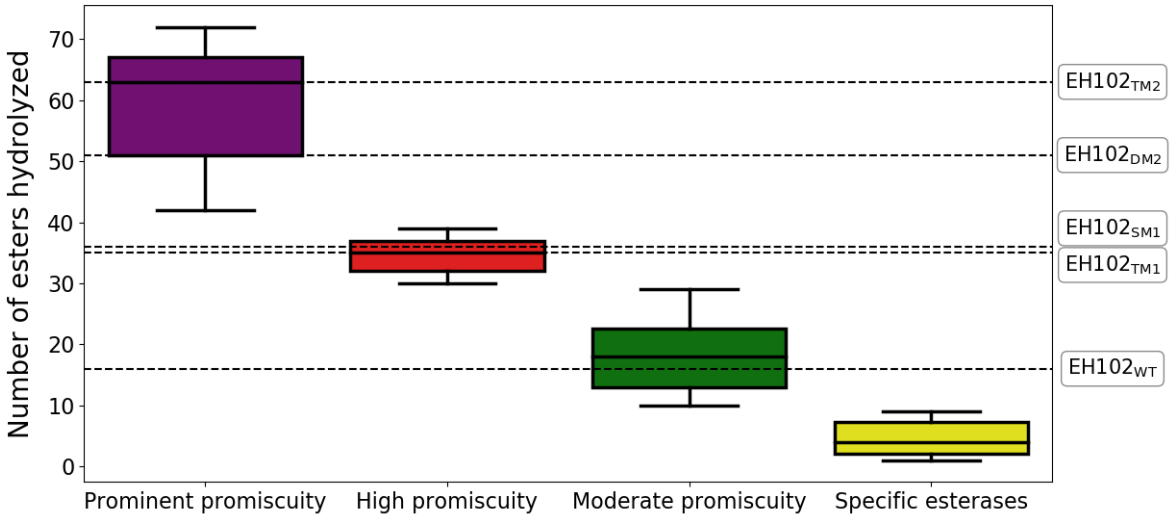


Figure 8. Comparison of the substrate promiscuity of EH102 variants against the published dataset of esterases[13]. The plot shows the number of esters hydrolyzed by $EH102_{WT}$, $EH102_{TM1}$,

$EH102_{SM1}$, $EH102_{DM2}$, and $EH102_{TM2}$ in comparison with the other 146 ester hydrolases. The figure was created with the Matplotlib library[51].

Table 2. Kinetic parameters determined for the model ester 1-naphthyl acetate.

Hydrolase	**Number of esters hydrolyzed**	k_{cat} **(min^{-1})**	K_M **(□M)**	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	22103
$EH102_{TM2}$	63	67.777	58	19466
EH1	72	2247.0	272	40126
EH3	69	173.260	231	33706
EH5	67	416.670	56	32477
EH7	64	1683.80	249	29056
EH12	51	3360.0	657	25240
EH37	28	59.840	35	23827

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, e.g. 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, etc; 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.

AUTHOR INFORMATION

Corresponding Author

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

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

ORCID

Sergi Roda: 0000-0002-0174-7435

Rubén Cañadas: 0000-0002-9063-4984

Gerard Santiago: 0000-0002-0506-3049

Manuel Ferrer: 0000-0003-4962-4714

Victor Guallar: 0000-0003-3274-2482

Author Contributions

⊥These authors contributed equally.

Notes

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge.

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 ΔΔG 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 (PDF)

ACKNOWLEDGMENTS

This work was funded by grant ‘INMARE’ from the European Union’s Horizon 2020 (grant agreement no. 634486), grants PCIN-2017-078 (within the Marine Biotechnology ERA-NET), and BIO2017-85522-R and PID2019-106370RB-I00 grants from the Spanish Ministry of Science and Innovation, Ministerio de Economía y Competitividad, Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER) and European Union (EU). This work has also been supported by a predoctoral fellowship from the Spanish Ministry of Science and Innovation (FPU19/00608).

ABBREVIATIONS

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* (3), e201209017.

(2) Brannigan, J. A.; Wilkinson, A. J. Protein Engineering 20 Years on. *Nature Reviews Molecular Cell Biology* **2002**, *3* (12), 964–970.

(3) Chen, R. A General Strategy for Enzyme Engineering. *Trends in Biotechnology* **1999**, *17* (9), 344–345.

(4) Chen, R. Enzyme Engineering: Rational Redesign versus Directed Evolution. *Trends Biotechnol.* **2001**, *19* (1), 13–14.

(5) Kuchner, O.; Arnold, F. H. Directed Evolution of Enzyme Catalysts. *Trends Biotechnol.* **1997**, *15* (12), 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* (22), 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* (2), 157–167.

(9) Hult, K.; Berglund, P. Enzyme Promiscuity: Mechanism and Applications. *Trends in Biotechnology* **2007**, *25* (5), 231–238.

(10) Copley, S. D. An Evolutionary Biochemist's Perspective on Promiscuity. *Trends in Biochemical Sciences* **2015**, *40* (2), 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. *Proceedings of the National Academy of Sciences* **2013**, *110* (45), 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* (9), 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. Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chem. Biol.* **2018**, *13* (1), 225–234.

(14) Holmquist, M. Alpha/Beta-Hydrolase Fold Enzymes: Structures, Functions and Mechanisms. *Curr. Protein Pept. Sci.* **2000**, *1* (2), 209–235.

(15) Nardini, M.; Dijkstra, B. W. Alpha/beta Hydrolase Fold Enzymes: The Family Keeps Growing. *Curr. Opin. Struct. Biol.* **1999**, *9* (6), 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* (Database issue), 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* (1), 16–30.

(18) Kawai, F.; Kawabata, T.; Oda, M. Current State and Perspectives Related to the PET Hydrolases Available for Biorecycling. *ACS Sustainable Chemistry & Engineering* **2020**, *8* (24), 8894–8908.

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

(20) Ferreira-Dias, S.; Sandoval, G.; Plou, F.; Valero, F. The Potential Use of Lipases in the Production of Fatty Acid Derivatives for the Food and Nutraceutical Industries. *Electronic Journal of Biotechnology* **2013**, *16* (3), 12–12.

(21) Guerrand, D. Lipases Industrial Applications: Focus on Food and Agroindustries. *OCL* **2017**, *24* (4), D403.

(22) Panda, T.; Gowrishankar, B. S. Production and Applications of Esterases. *Applied Microbiology and Biotechnology* **2005**, *67* (2), 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* (6), 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* (22), 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* (3), 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* (2), 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* (11), 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* (1), 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* (7), 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* (6), 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 Catalysis* **2020**, *10* (1), 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* (2), 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* (3), 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* (2), 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* (16), 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. *International Journal of Quantum Chemistry* **2013**, *113* (18), 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* (6), 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* (7), 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. *The Journal of Chemical Physics* **1983**, *79* (2), 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* (3), 712–725.

(42) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. *The Journal of Chemical Physics* **1980**, *72* (4), 2384–2393.

(43) Chow, K.-H.; Ferguson, D. M. Isothermal-Isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling. *Computer Physics Communications* **1995**, *91* (1-3), 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. *Chemical Physics Letters* **2004**, *384* (4-6), 288–294.

(45) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **1967**, *159* (1), 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* (7), 1739–1749.

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

(48) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49* (2), 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* (W1), 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. *Nature Catalysis* **2020**, *3* (3), 319–328.

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

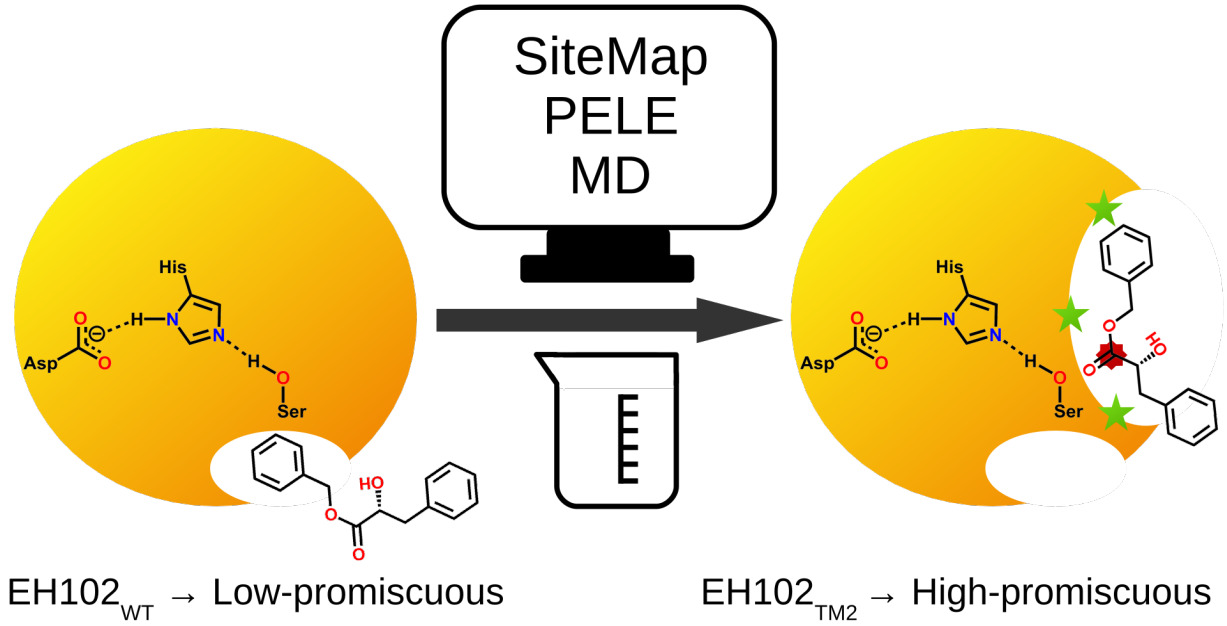
(52) Santiago, G.; Martínez-Martínez, M.; Alonso, S.; Bargiela, R.; Coscolín, C.; Golyshin, P. N.; Guallar, V.; Ferrer, M. Rational Engineering of Multiple Active Sites in an Ester Hydrolase. *Biochemistry* **2018**, *57* (15), 2245–2255.

(53) Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial Biocatalysis Today and Tomorrow. *Nature* **2001**, *409* (6817), 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. *Biocatalysis and*

Biotransformation **2015**, *33* (5-6), 235–249.

Table of Contents / Abstract Graphics



For Table of Contents only