UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Título
Acyltransferase LovD as a Simvastatin Synthase: Mutational Deconvolution, Design, Substrate Scope and Immobilization

Autor/es
Guillermo García Marquina

Director/es
Francisco Corzana López, Gonzalo Jiménez Osés y Fernando López Gallego

Facultad
Facultad de Ciencia y Tecnología

Titulación

Departamento
Química

Curso Académico

FACULTAD DE CIENCIA Y TECNOLOGÍA

UNIVERSIDAD DE LA RIOJA

Departamento de Química

Centro de Investigación en Síntesis Orgánica

Área de Química Orgánica

TESIS DOCTORAL

Acyltransferase LovD as a Simvastatin Synthase: Mutational Deconvolution, Design, Substrate Scope and Immobilization

Memoria presentada en la Universidad de la Rioja para optar al grado de Doctor en Química por

Guillermo García Marquina

Diciembre 2021

Dr. FRANCISCO CORZANA LÓPEZ, Profesor Titular de Química Orgánica del Departamento de Química de la Universidad de La Rioja,

Dr. GONZALO JIMÉNEZ OSÉS, Investigador asociado Ikerbasque y

Dr. FERNANDO LÓPEZ GALLEGO, Profesor de investigación Ikerbasque

HACEN CONSTAR:

Que la memoria "Acyltransferase LovD as a Simvastatin Synthase: Mutational Deconvolution, Design, Substrate Scope and Immobilization" realizada por Guillermo García Marquina en el Departamento de Química de la Universidad de La Rioja y bajo su inmediata dirección, reúne las condiciones exigidas para optar al grado de Doctor en Química.

Logroño, Diciembre de 2021

Los directores

Francisco Corzana López   Gonzalo Jiménez Osés   Fernando López Gallego

Table of Contents

4

List of Abbreviations

Transition state	TS
Ribonucleic acid	RNA
Deoxyribonucleic acid	DNA
Deoxyribonucleotide	dNTP
Nuclear magnetic resonance	NMR
Polyketide	PK
Polymerase chain reaction	PCR
Error-prone polymerase chain reaction	ep-PCR
Green fluorescence protein	GFP
Site-directed saturation mutagenesis	SdSM
Staggered extension process	StEP
Iterative truncation for the creation of hybrid enzymes	ITCHY
High-performance liquid chromatography	HPLC
Gas chromatography	GC
Mass spectrometry	MS
Ultra-high performance liquid chromatography	UPLC
Fluorescence-activated cell sorting	FACS
Cyclic recombinase	Cre
Sortase A	SrtA
Tobacco etch virus	TEV
In vitro compartmentalization	IVC
DNA binding domain	DBD
Transition state analogues	TSA
Messenger ribonucleic acid	mRNA
Compartmentalized self-replication	CSR
Compartmentalized partnered replication	CPR
Comparative modelling	CM
Protein Data Bank	PDB
Statistical analysis of protein-activity relationships	ProSAR
Molecular dynamics	MD
Degrees of freedom	DOF
Principal component analysis	PCA
Root-mean-square deviation	rmsd
HARvard molecular mechanics	CHARMM
Assisted model building and energy refinement	AMBER
Optimized potential for liquid simulations	OPLS
Groningen molecular simulation	GROMOS

Central processing unit	CPU
Graphics processing unit	GPU
Markov state model	MSM
Accelerated molecular dynamics	aMD
β-amyloid cleaving enzyme 1	BACE1
Metadynamics	MetaMD
Umbrella sampling	US
Weighted histogram analysis method	WHAM
Acetylcholinesterase	AChE
Shortest path map	SPM
Haloalkane dehalogenase	DHlA
Lipase A from *Bacillus subtilis*	BsLipA
Constraint network analysis	CAN
Quantum mechanics	QM
Molecular mechanics	MM
Density functional theory	DFT
Nitrite reductase	NiR
Machine learning	ML
Potential of hydrogen	pH
Immobilized affinity chromatography	IMAC
Cross-linked enzyme aggregate	CLEA
Candida antarctica lipase B	CALB
Polyvinyl alcohol	PVA
Poly (lactic –co glycolic acid	PLGA
Controlled porosity glass	CPG
Nanoparticle	NP
Gold nanoparticle	AuNP
Magnetic nanoparticle	MNP
Poly-N-isopropylacrylamide	polyNIPAM
Metal-organic framework	MOF
Continuous flow system	CFS
Back pressure regulator	BPR
Wall-coated reactor	WCR
Packed-bed reactor	PBR
Coenzyme A	CoA
Acyl carrier protein	ACP
Acyltransferase	AT
Polyhydroxyalkanoate	PHA
Triacylglycerol	TAG
Wax ester	WE
Phospholipid	PL
Glycolipid	GL
Fatty acid alkyl ester	FAAE
Glycerol-3-phosphate	G3P

1,2-diacylglycerol	DAG
Acyl-CoA-diacylglycerol acyltransferase	DGAT
Membrane-bound-O-acyltransferase	MBOAT
Cholesteryl ester	CE
Acyl-CoA:colesterol acyltransferase	ACAT
Polyhydroxyalkanoate synthase	PhaC
Alcohol acyltransferase	AAT
Endoplasmic reticulum	ER
Ketoacid decarboxylase	KDC
Medium-chain fatty acid	MCFA
Lipase A from *Candida antarctica*	CALA
Acyltransferase from *Mycobacterium smegmatis*	MsAcT
Polyketide synthase	PKS
Ketosynthase	KS
Ketoreductase	KR
Dehydratase	DH
Enoylreductase	ER
Methyltransferase	MT
Cysteine lyase	SH
Thioesterase	TE
6-deoxyerythronolide B synthase	DEBS
Statistical coupling analysis	SCA
Pikromycin	Pik
Lovastatin synthase from *Aspergillus terreus*	LovD
Monacolin J acid	MJA
α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate	DMB-SMMP
Hydroxymethyl glutaryl-coenzyme A	HMG-CoA
S-adenosyl-L-methionine	SAM
tert-Butylmethylsilyl chloride	TBSCl
University of California Los Angeles	UCLA
N-Acetylcysteamine	SNAC
Methyl-thioglycolate	SMTG
α-dimethylbutyryl	DMB
S-mercaptopropionic acid	SMPA
Penicillium chrysogenum esterase	PcEST
Dihydromonacolin L acid	DMLA
Cytochrome P450 reductase	CPR
Dimethyl butanoic acid	DMB
Simvastatin acid	SVA
S-methyl-mercaptopropionic acid	SMMP
4'-phosphopantetheine	PPN

Lysogeny broth	LB
Isopropyl-D-thiogalactoside	IPTG
Times gravity	xg
Agarose	AG
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	HEPES
sodium dodecyl sulphate-polyacrylamide gel electrophoresis	SDS-PAGE
Standard molecular weight marker	MW
Dimethyl sulfoxide	DMSO
Photodiod array	PDA
Time-of-flight	TOF
Elecrospray ionization source	ESI
2,2'-dithiodipyridine	DTDP
General Amber forcé field	GAFF2
Michaelis-Menten	MM
Thioesterase activity	TE
Acyltransferase activity	AcT
Total system energy	TSE
Laboratory-directed evolution	LDE
Halohydrin dehalogenase	HHDH
Cryo-electron microscopy	Cryo-EM
Structure-activity relationship	SAR
Weighted Implementation of Suboptimal Paths	WISP
Multiple protein sequence alignement	MPSA
Endo-1,4-β-glucanase from *Reticulitermes speratus*	RsEG
Prolyl endopeptidase from *Sphingomonas capsulate*	ScPEP
Position-Specific Iterative Basic Local Alignment Search Tool	PSI-BLAST
Betweenness centrality	BC
Lovastatin hydrolase from *Penicillium chrysogenum*	PcEST
Vinyl acetate	VA
Vinyl propionate	VP
Vinyl butyrate	VBU
Vinyl benzoate	VBZ
Ethyl 2-methyl butyrate	EMB
p-Nitrophenyl butyrate	pNPB
p-Nitrophenyl 2-methyl butyrate	pNPMB
p-Nitrophenyl 2,2-dimethyl butyrate	pNPDMB
p-Nitrophenol	pNP

N-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride	EDC
Phenol	PH
(+)-Menthol	MT
1-Naphthol	NT
Yeast alcohol dehydrogenase	yADH
Dichloromethane	DCM
4-Dimethylaminopyridine	DMAP
Thin-layer chromatography	TLC
Ultraviolet	UV
Mass selective detector	MSD
Intrinsic reaction coordinate	IRC
Active pharmaceutical ingredient	API
Continuous flow processing	CFP
Immunodeficiency virus	HIV
Ketoreductase	KRED
Nicotinamide adenine dinucleotide phosphate	$NADP^+$/NADPH
Iminodiacetic acid	IDA
Rhodamine B isothiocyanate	RBITC
Fluorescein isothiocyanate	FITC
Confocal laser scanning microscopy	CLSM
Agarose-Co^{2+} epoxidated	AG-Co^{2+}/E
True component analysis	TCA

List of Constants and Parameters

Maximum reaction rate	v_{max}	$\mu mol_{(PRODUCT)}\ mg_{(ENZYME)}^{-1}\ min^{-1}$
Michaelis-Menten constant	K_M	mM^{-1}
Catalytic constant	k_{cat}	s^{-1}
Reynolds number	Re	-
Residence time	τ	min
Space time yield	STY	$mg_{PRODUCT}\ min^{-1}\ mL^{-1}$
Inhibition constant	K_i	mM
Molar absorption coefficient	ε	$M^{-1}\ cm^{-1}$
Inactivation constant	k	-
Melting temperature	T_m	°C
Half-life time	$t_{1/2}$	h
Pearson correlation coefficient	ρ	-
Total electronic units	E_{elec}	Hartree
Enthalpy	H	Hartree
Entropy	S	$cal\ mol^{-1}\ K^{-1}$
Gibbs energy	G	Hartree
Immobilization yield	ψ	%
Recovered activity	RA	$U\ g_{(CARRIER)}^{-1}$
Immobilized specific activity	iSA	$U\ mg_{(ENZYME)}^{-1}$
Relative recovered activity	rRA	%

Resumen

En la presente tesis doctoral se lleva a cabo un estudio sistemático de la enzima aciltransferasa LovD presente en el hongo *Aspergillus terreus*, modificada mediante evolución dirigida para la producción mejorada del fármaco anti-colesterolémico simvastatina a partir de ácido monacolínico J. En primer lugar, se describe un análisis de los factores que influyen en la actividad catalítica de dicha enzima a lo largo de las distintas rondas de evolución dirigida (competición entre actividades tioesterasa, aciltransferasa e hidrolasa, inhibición por sustrato y estabilidad térmica), y se diseñan nuevas variantes con un menor númerode mutaciones y actividades catalíticas comparables. En segundo lugar, se lleva a cabo un diseño racional de la variante nativa de LovD empleando métodos computacionales para diversificación de secuencia y dinámica molecular. A continuación, se describe la utilidad de ésteres de vinilo y para-nitrofenilo como alternativas eficientes al empleo de tioésteres como agentes donadores de grupos acilo. Por último, se describe la inmovilización de una variante diseñada de LovD en distintos soportes sólidos con características diferentes y se analizan en detalle las propiedades de dichos biocatalizadores mediante técnicas biofísicas, optimizándose las condiciones de producción en reactores continuos de flujo para una síntesis de simvastatina eficiente con potencial aplicación a nivel industrial

Abstract

In this doctoral thesis, a systematic study of the acyltransferase enzyme LovD from the fungus Aspergillus terreus, which waspreviously modified by directed evolution for the improved production of the anti-cholesterol drug simvastatin from monacolinic acid J, is performed. First, an analysis of the features affecting the catalytic activity of the enzyme throughout the different rounds of directed evolution (competition between thioesterase, acyltransferase and hydrolase activities, inhibition by substrate and thermal stability), is described, and new variants with fewer mutations and comparable catalytic activities are designed. Second, a rational design of native LovD is performed using computational methods for sequence diversification and molecular dynamics. Then, the value of vinyl and para-nitrophenyl esters as efficient alternatives to thioesters as acyl donor agents is described. Finally, the immobilization of a designed LovD variant on different solid supports with different characteristics is accomplished, and the properties of the resulting biocatalysts are analyzed in detail using biophysical techniques. The conditions for the efficient synthesis of simvastatin using continuous flow reactors with potential interest for industrial production, were optimized.

CHAPTER 1:

Introduction

1.1. Biocatalysis

1.1.1. Introduction to Biocatalysis and Enzyme Kinetics

Chemical reactions can be defined as collisions between different chemical compounds that are converted into new ones with different properties.[1] These reactions always involve breaking chemical bonds between reactants and form new bonds between atoms in products, with no alteration in the number of atoms, but a possible change in the number of molecules. The success of a reaction depends on factors as the relative kinetic energy, relative orientation and internal energy of the molecules, but a collision between molecules is not necessarily successful, and the activated complex can form new products or go back to the original reactants. Those chemical reactions occur spontaneously, but usually take place in very long periods of time due to their energy requirements. In this way, the reaction time depends on the **energy activation barrier**, which is determined by the **transition state (TS)** of the reaction, the state corresponding to the highest potential energy along this reaction coordinate. Once the reactants have overcome the transition state, products formation is assured.

A **catalyst** is defined as a substance that modifies the transition state to lower the energy activation barrier, facilitating the reaction and diminishing the reaction time.[1,2] This is possible due to a release of energy that occurs when the reactant binds to the active site of the catalyst. Catalysts can be classified as homogeneous and heterogeneous, depending on whether the reaction components are in the same phase or in different phases, respectively. Organometallic catalysts stand out between **homogeneous catalysts**,[3–5] and they are characterized by containing at least one chemical bond between a carbon atom of an organic molecule and a metal. On the other hand, important **heterogeneous catalysts** are zeolites, alumina, higher-order oxides, graphitic carbon and transition metal oxides, among others.[6–10] Furthermore, organocatalysts have also emerged as a pillar of asymmetric catalysis, offering an efficient way of obtaining chiral molecules that does not rely on transition metals.[11,12]

In nature, **enzymes** are the essential biological catalysts that accelerate catabolic and anabolic reactions occurring in cell metabolism.[1,13,14] Enzymes are usually globular proteins acting alone or in larger complexes, though some Ribonucleic acid (RNA) molecules act as enzymes too (ribozymes).[15,16] The sequence of amino acids specifies the structure, which in turn determines the catalytic activity of the enzyme.[17,18] The broad variety of enzymes existing in nature catalyze very different reactions, and the International Union of Biochemistry and Molecular Biology has used this criterion to establish enzyme classification in six main families: oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases.[19]

The part of the enzyme where the substrate binds is called the **active site**.[20,21] In 1894, the chemist Emil Fischer proposed that the substrate of an enzyme fits into the enzyme's

active site to form an enzyme-substrate complex, using the analogy of a lock and key.[22] The key (substrate) has a specific shape that allows it and no other key to fit into the lock (enzyme) (**Figure 1.a**). This model explained a well-observed phenomenon, namely the high specificity of enzymes, but failed to explain the stabilization of the transition state that enzymes achieve. In 1958, Daniel E. Koshland Jr. modified the lock-and-key model by proposing that binding of the substrate to the enzyme alters the configuration of both, providing a better fit.[23] Thus, the substrate is distorted to resemble the transition state of the reaction and the amino acid side-chains that make up the active site are molded into the precise position that enable the enzyme to perform its catalytic function (**Figure 1.b**). This stabilizes the transition state, accelerating the reaction like any catalyst. Only molecules with the correct functional groups in the correct configuration can be induced to fit in the active site of the enzyme. Moreover, this active site plasticity also supposed an explanation to enzyme promiscuity for more than one substrate, which shows up a very rich catalytic landscape of main and side enzymatic reactions.
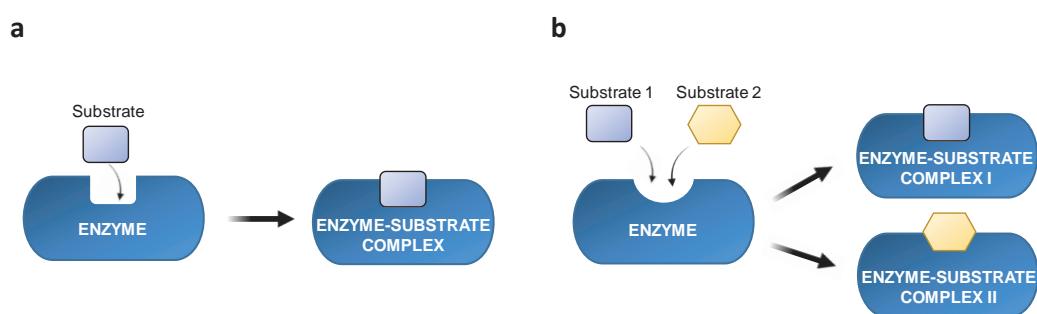
Figure 1. Proposed enzyme-substrate interaction models. **(a)** Fischer "lock-and-key" model. **(b)** Koshland "induced fit" model.

Each enzyme presents a particular efficiency for a particular reaction, which is determined through its kinetic parameters. In 1913, Leonor Michaelis and Maud Menten proposed a quantitative theory of enzyme kinetics (**Michaelis-Menten kinetics**), dividing all enzyme reactions in two stages: first, the reversible binding of the substrate to the enzyme, forming the enzyme-substrate complex; and second, the chemical step of the reaction and the release of the product (**Figure 2**).[24] To find the maximum speed of an enzymatic reaction, the substrate concentration is increased until a constant rate of product formation is seen. The maximum reaction rate (v_{max}) is described as the rate where all the enzyme active sites are bound to substrate. On the other hand, the **Michaelis-Menten constant (K_M)** determines the substrate concentration of an enzyme to reach one-half its maximum reaction rate, and it is often used as a measure of the affinity of the substrate for the enzyme. Finally, the **turnover number (k_{cat})** is defined as the number of substrate molecules handled by one active site per second, being associated with the velocity of the reaction. The **catalytic efficiency** of an enzyme is expressed as k_{cat}/K_M.

Enzymes are not rigid and static, but high dynamic structures composed of different groups of amino acids that organize in more complex frameworks. These motions give rise to a conformational ensemble of slightly different structures that interconvert with one another at equilibrium. These different structural states can directly affect the enzyme function. Apart from the active site, enzymes might also have other pockets, called allosteric sites, where other molecules of the cellular environment can bind. These molecules can change the conformation or dynamics of the enzyme that indirectly influences the catalytic site conformation and thus enzyme activity.[25–27]
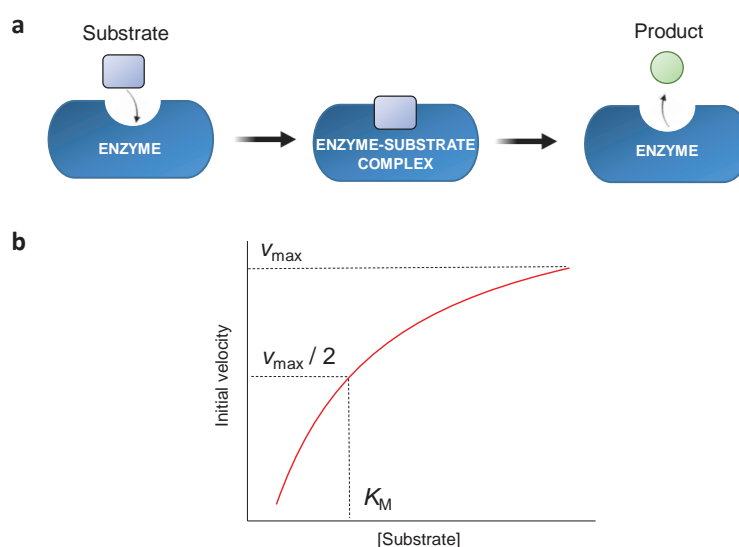
Figure 2. (a) Two-stage Michaelis-Menten enzyme kinetics model. **(b)** Michaelis-Menten plot where v_{max} is the maximum velocity of the reaction and K_M is the Michaelis constant.

1.1.2. Biocatalysis in Chemical Industry

The use of enzymes by humans dates back thousands of years, when **whole-cell bacteria and yeast fermentations** were first used to produce certain foods such as wine, beer, cheese, and bread. However, whole-cell fermentations presented some limitations in terms of product purification, resistance to harsh conditions, by-products toxicity or product diffusion issues. In the 20th century, **isolated enzymes** started to be employed to carry out a wide variety of chemical processes, such as detergents, materials and polymers or fine chemicals.[28,29] Some examples of successful biocatalytic processes include the production of high-fructose corn syrup through the isomerization of D-glucose to D-fructose by the action of xylose isomerase,[30] preparation of semisynthetic penicillins by penicillin amidase[31] and the use of proteases in laundry detergents.[32] Recently, enzyme catalysis is used to synthetize more complex molecules, such as cosmetics, pharmaceuticals, antibody-drug conjugates, food supplements or even new biocatalyst-based processes for the production of biofuels using renewable starting materials as an alternative to fossil fuels.[33–35] Examples include the

lipase-catalyzed resolution of alcohols and amines yielding enantiopure compounds[36] and the hydration of acrylonitrile to acrylamide for polymers catalyzed by a nitrile hydratase.[37]

The rise of enzymatic catalysis is explained by its numerous advantages compared with conventional chemical synthetic processes: they are biocompatible and environmentally friendlier offering safe processes at mild conditions and without generation of waste products in excess; they are capable of catalyze a broad range of chemical reactions and they are regio- stereo- and chemoselective catalysts. Biocatalysis often complements rather than compete with chemocatalysis, which is best used to prepare small enantiomerically pure compounds that reacts with other molecules to synthetize the final product. New biocatalytic processes have the purpose of generating new molecules with commercial interest or replacing or complementing existing non-optimal chemical synthesis in industry. The most desired properties of a biocatalyst to carry out a particular reaction are a high activity, stability under industrial process conditions, environmental sustainability, renewability, low-cost and availability. Moreover, some biocatalysts can be engineered and used to perform other reactions apart from the principal one if they also present a high selectivity, substrate scope and "evolvability".[38–40] Nevertheless, enzymatic bioprocesses still face important challenges. In the first place, the high solubility of enzymes in aqueous solutions limits their recycling. Secondly, their operation is often limited to physiological conditions (pH, temperature, aqueous media and low substrate concentrations). Finally, their high selectivity largely favors reactivity with non-natural substrates, limiting the spectra of possible products.

All these inconveniences have been overcome by means of different approaches. **Immobilization of enzymes** in solid materials was a major breakthrough to avoid solubility problems and reuse limitations.[41] **Computational enzyme engineering** have emerged as a very valuable approach to obtain novel catalysts with non-natural activities by *de novo* redesigning their active sites.[42] Finally, **directed evolution of enzymes** entails one of the most important advances in catalysis as it achieved the improving of many enzyme features at the same time, considerably increasing the productivity of many processes.[43] All of these approaches are usually combined in a standardized protocol of enzyme engineering, the **biocatalytic cycle (Figure 3)**.[40] The starting key point in the cycle is the selection of a proper organism to produce the catalyst or directly perform the desired reaction. The second step is the characterization of the biocatalyst parameters, such as activity, stability, selectivity, and kinetics. At the same time, obtaining structural information through X-ray crystallography, cryo-electron microscopy or nuclear magnetic resonance (NMR) can be priceless for the next step: engineering the biocatalyst. At this stage, the combination of rational design based on structural information and computational tools along with directed evolution is a winning strategy. Once the biocatalyst is engineered, it must be prepared for its application to the industrial process. Here, key features are solvent system selection, immobilization techniques (in heterogeneous carriers, cross-linking…), recycling for repeated (fed)-batch processes and application in continuous reactions. Finally, in the last step the product must be recovered either through a downstream process or *in situ* (crystallization, distillation, solid- or liquid-phase recovery).
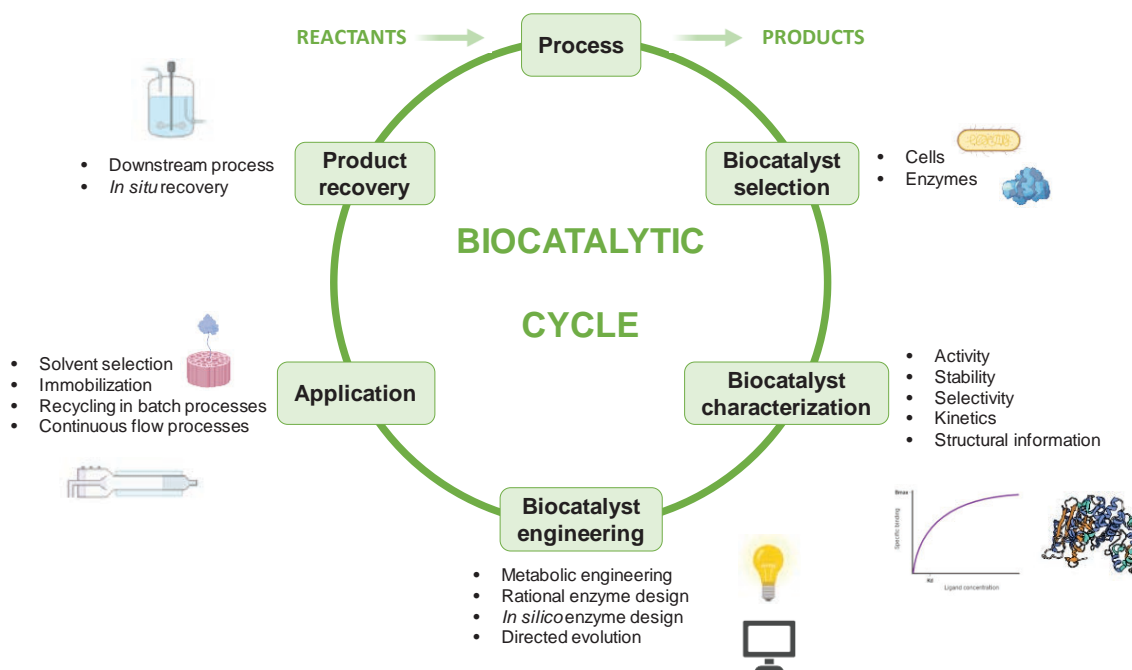
Figure 3. Biocatalytic cycle.[40]

In recent years, acyltransferases have become a quoted target to enzyme engineering due to their ability to catalyze relevant reactions for chemical industry, such as the synthesis of storage and membrane lipids, polyketide (PK)-containing lipids, bacterial toxins or antibiotics.[44,45] Due to the huge impact of cardiovascular diseases in public health, some researches in the last decade have been focused in the engineering of enzymes to improve the synthesis of cholesterol-lowering drugs.[46] A notable example is the characterization of the acyltransferase LovD from *Aspergillus terreus*, responsible of the biosynthesis of some cholesterol-lowering drugs such as lovastatin and simvastatin.[47] In this work, we intend to provide an overview to the state of the art of acyltransferases reactions and protein design for anti-cholesterol drugs synthesis, focusing on the particular case of the lovastatin synthase LovD.

1.2. Directed Evolution of Enzymes

Over many generations, biological evolution provides solutions for challenges that organisms face in the real world through iterated mutation and natural selection. Evolution can be guided to access useful phenotypes of organisms and biomolecules that are sought by humans. Due to the aforementioned limitations for industrial applications of wild-type enzymes, the optimization of some enzyme properties such as substrate specificity and selectivity, high catalytic turnover and high process stability under the conditions of the chemical transformation are often required. Even with a good-resolved crystal structure, rational design of enzymes is complex and not always successful. **Directed evolution** is a

powerful tool to carry out this optimization, consisting in iterative rounds of random gene library generation, expression of genes in a suitable host and screening of libraries of variant enzymes for the property of interest (**Figure 4**).[48,49] Depending on the screening selection, more than one enzymatic property can be improved at the same time.
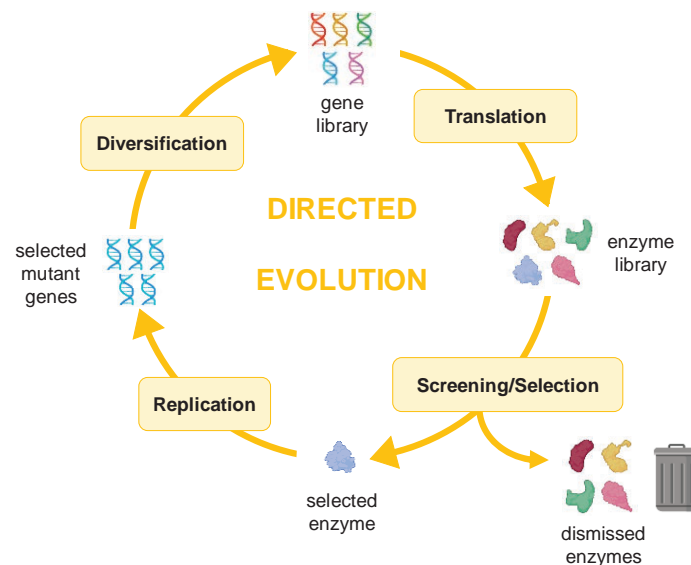
Figure 4. Directed Evolution process.[49]

1.2.1. Methods for gene diversification

For a protein composed of 400 amino acids, the number of possible variants is 20^{400}, a very broad sequence space impossible to access and screen. **Gene diversification methods** are therefore required to perform an optimal sparse sampling of this vast sequence space. DNA libraries are generated using diverse molecular biology techniques as random and site saturation mutagenesis, depending on the availability of structural information and screening capacity. Computational modeling and bioinformatics can be useful guides to select amino acid residues for design more focused libraries (10^2 to 10^4 variants), which are generated using techniques such as saturation mutagenesis or iterative combinatorial active site testing. Other molecular biology techniques such as error-prone polymerase chain reaction (epPCR) or DNA shuffling could be more useful to generate larger libraries (10^5 to 10^8 variants).

In vivo **random mutagenesis** can be carried out using organisms with DNA polymerase III, which introduces a rate of 10^{-10} mutations per replicated base.[50] Some bacteria strains contain deactivated proof-reading and repair enzymes, which help to increase the mutation rate.[51] However, these strains can also induce deleterious mutations in the host genome non compatible with its survival. *In vitro* **random mutagenesis** techniques are preferable due to their higher control and mutation rate (10^{-4} to 10^{-3} mutations per replicated base). With **error-prone polymerase chain reaction (epPCR)**, random mutations can be inserted into any

piece of DNA with a PCR in which the fidelity of a Taq DNA polymerase without proof-reading activity is modulated by alteration of the composition of the reaction buffer (usually increased magnesium concentrations, supplementation with manganese or mutagenic dNTP analogues).[52–54] In this technique, the average number of mutations per clone can be increased with the number of cycles. Normally, the target gene is fused to a green fluorescence protein (GFP) reporter to perform a previous expression screening and discard deleterious mutations before directed evolution experiments.[55]

The main limitation of random mutagenesis is the low efficiency in the introduction of more than one nucleotide change at the same time in a particular codon. To solve this problem, a recurring implement is the use of **site-directed saturation mutagenesis (SdSM)**, which is based on the use of synthetic DNA oligonucleotides containing one or more degenerated codons at the targeted residues positions (**Figure 5.a**).[56] These oligonucleotides are cloned into a gene library as a mutagenic cassette, allowing simultaneous saturation mutagenesis of multiple residues. With this technique, the exploration of epistatic interactions, such as synergistic effects between mutations, is also possible. The method developed by Reetz et al. called CASTing (combinatorial active site testing) selects and organizes active site residues in groups of two or three residues to generate libraries.[57] After those libraries screening, best candidates are combined and further screened in an iterative way.

A prominent implementation of methodology for directed evolution was based on a DNA recombination strategy termed as **"DNA shuffling"**. This technique consists on the fragmentation of a gene with a DNase and the subsequent annealing, extension, denaturation and amplification of the fragments to full-length genes through a PCR without added primers (**Figure 5.b**).[58] This supposes a beneficial way to propagate beneficial mutations while increasing the size of a DNA library. Other alternate techniques of DNA shuffling have emerged from the original one. The **staggered extension process (StEP)** is a modified protocol that consists in priming the template sequences followed by repeated cycles of denaturation and extremely abbreviated annealing/polymerase-catalyzed extension. In each cycle, the growing fragments can anneal to different templates based on sequence complementarity and extend further to create recombinant cassettes. StEP is continued until full-length genes are formed.[59] Other recombination technique is the **iterative truncation for the creation of hybrid enzymes (ITCHY)**, which is based on the generation of N- or C-terminal fragment libraries of two genes by progressive truncation of

the coding sequences with exonuclease III followed by ligation of the products to make a single-crossover hybrid library.[60]
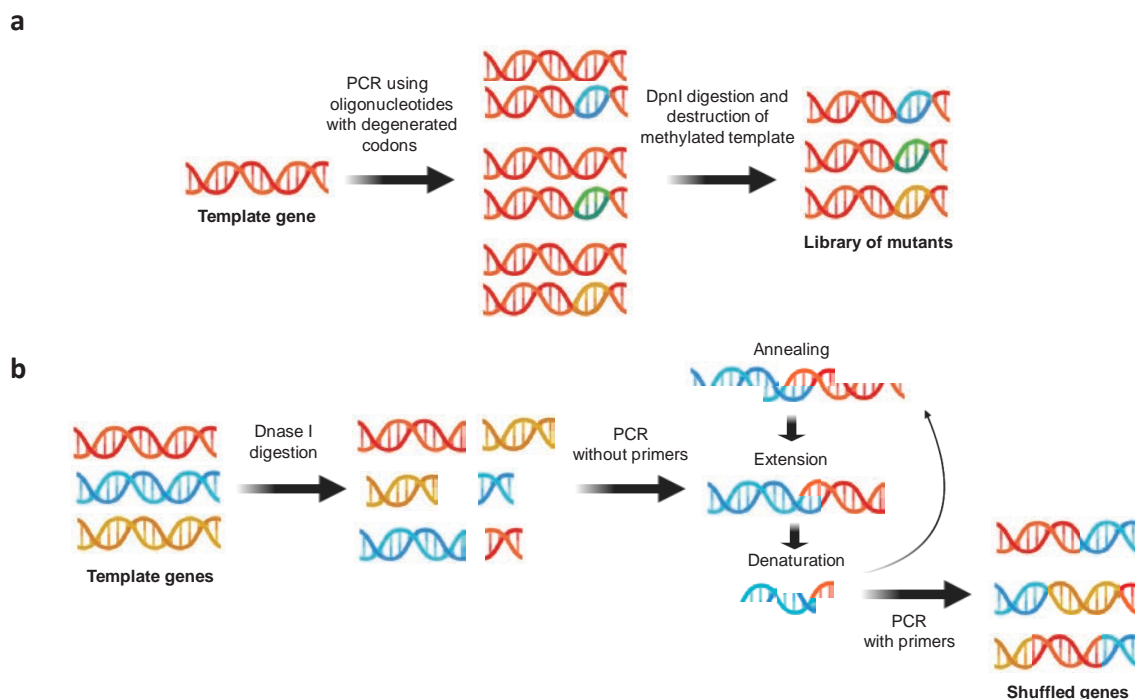
Figure 5. Methods for gene diversification. **(a)** Site-directed saturation mutagenesis (SdSM)[56]. **(b)** DNA shuffling[58].

1.2.2. Screening Methods for Enzyme Evolution

Whatever strategy is followed for library generation, the development of a robust high-throughput **screening method** for the desired function of the catalyst is vital. Normally, enzyme characteristics implemented in the screening process are catalytic activity, selectivity (enantioselectivity, diastereoselectivity and regioselectivity) and stability to temperature and organic solvents.

Expression of gene variants in a unicellular model organism such as *E. coli* allows the screening of colonies when the bacterial culture is transferred into solid media or liquid culture. Despite the limitation in the library members screening, this **spatial separation** (**Figure 6**) is very useful for its compatibility with many different analytical techniques, such as nuclear magnetic resonance (NMR), high-performance liquid chromatography (HPLC), gas chromatography (GC) or mass spectroscopy (MS), which can monitor substrate consumption or product formation. These **low-throughput** screens require a certain knowledge about the structure-activity relationship of the target protein to maximize the probability of obtaining an improved variant. A good example where this screening strategy was successfully used is the evolution of cytochromes P450 by Arnold et al.[61]

When this structural knowledge is lacking, **high-throughput** techniques are a more useful approach. The 96-well plate is the most widely used format to incubate reaction components along with either crude cell extracts or purified proteins. Colony pickers and liquid handling robots allow to array 10^3 colonies per hour in 96-well plates in an accelerated way of library generation and protein production,[62] whereas ultra-high performance liquid chromatography systems (UPLC) are efficient procedures to screen the reaction, with an efficiency of 10^3 clones per instrument per day. Colorimetric and fluorometric assays are the most convenient microtiter plate-based enzyme activity assays.[63–65]

Rather than spatially separating clones, a bulk population can be screened at the level of individual cells using **fluorescence-activated cell sorting (FACS)**. This technique relies on a fluorescent reporter such as GFP that is coupled with the target enzyme and was successfully used for the screening of cyclic recombinase (Cre) mutants with altered site specificity and to monitor the folding efficiency mediated by chaperonin mutants.[66] FACS technology can be coupled to different screening systems:

- For **product entrapment**, a fluorescent substrate that interacts with the target enzyme is employed both into and out of the cell. This substrate can be washed off without being retained within the cell, while the fluorescent product remains retained inside the cell.[67]

- **Cell surface display** can also be coupled to FACS using fluorescent-labelled antibodies.[68] This approach became more broadly used with a yeast display screening method for protein-protein interactions.[69] Bond-forming or cleaving enzymes, such as sortase A (SrtA), a sequence-specific transpeptidase, or tobacco etch virus (TEV) protease, can be evolved using the yeast display framework.[70]

- When cell-constrained fluorescent reporters implementation for a particular gene is challenging, *in vitro* **compartmentalization (IVC)** is another alternative for high-throughput screening.[71] IVC uses a compartmentalized system consisting on water-in-oil or water-in-oil-in-water emulsion droplets that act as independent reactors that contain isolated individual DNA molecules and are able to synthetize proteins and carry out enzymatic reactions as a cell-free system. With IVC, the possibility that the evolved phenotype arises from mutations not related to the target gene is removed. Moreover, library size is not limited to host cell transformation efficiency anymore. IVC have been employed to screen the activity of [FeFe] hydrogenases.[72]
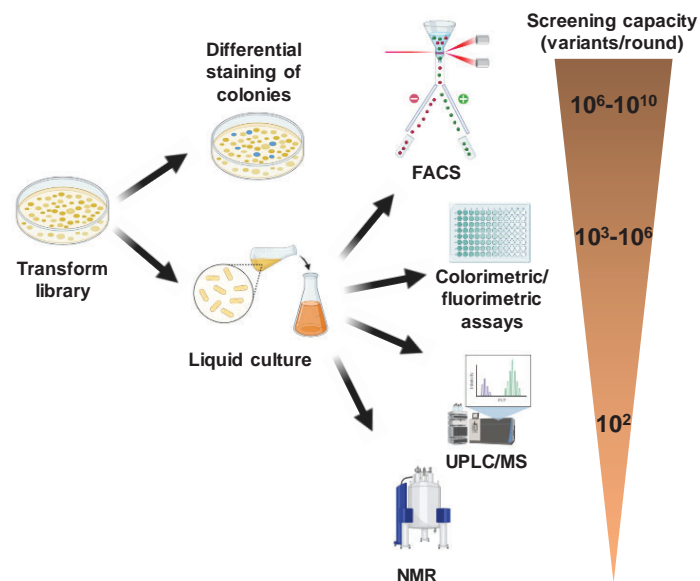
Figure 6. Screening methods for enzyme evolution: Fluorescence Activated Cell Sorting (FACS), differential staining of colonies, colorimetric and fluorometric 96-well plates assays, Ultra-high Performance Liquid Chromatography (UPLC) coupled with Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR).

1.2.3. Selection Methods for Enzyme Evolution

Screening methods are based on the generation and analysis of a huge set of enzyme variants, with the subsequent discard of a vast majority and the selection of a few variants. **Selection methods** directly eliminate unwanted enzyme variants through applying a selective pressure to the mutant library. In this way, only positive variants are carried out onto the next round of directed evolution. These methods allow the approach of much larger libraries feasible. Nevertheless, selecting techniques often require more creativity and a strong molecular intuition.

A useful selecting method is **plasmid display**, which is based on the fusion of a DNA binding protein with the target protein.[73] The fusion protein is expressed in the cell and binds its recognition sequence of the DNA binding domain (DBD) in the target plasmid. After the cell lysis, the protein-plasmid complex is selected and subjected to another round of evolution. In **cell-surface or phage display** methods,[74,75] Protein members are expressed and fused with surface proteins of the cell or of the coat of the bacteriophage (**Figure 7**). Different strategies for enzyme selection can be assessed with phage display as the indirect selection with inhibitors as transition state analogues (TSA) attached to solid supports, or the double attachment of both enzyme and substrate to the phage and the adsorption to a product binding support.[76] Despite filamentous phages' ability to infect cells, the main limitation of these selecting methods is the transformation step, which reduces considerably the library size to 10^9 to 10^{10} transformants per experiment. mRNA and ribosome display overcome this limitation. **mRNA display** produces proteins that are covalently bound to its encoding mRNA through a puromycin analogue.[77] **Ribosome display** is conceptually similar to mRNA display,

but instead of a covalent bond between the mRNA and the protein, the absence of a stop codon under carefully controlled conditions allows the stable bound of ribosome to both mRNA and the growing polypeptide. In this way, proteins are coupled with their encoding genes.[78] All of these techniques are useful to evolve binding affinity, but present substantial limitations to evolve other important enzyme properties, such as activity or catalytic efficiency, which reduce the number of evolved enzyme cases to some β-lactamases and RNA ligases.[79]
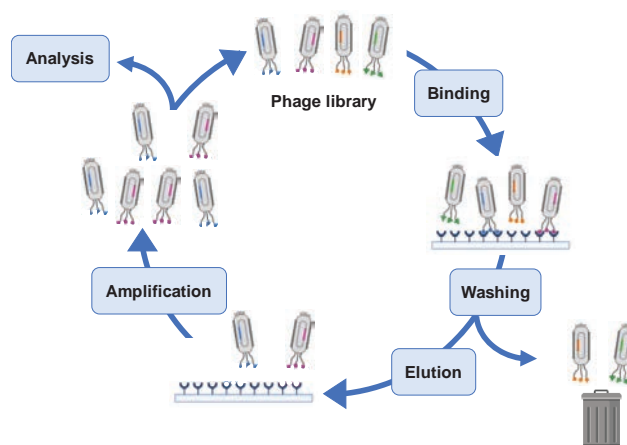
Figure 7. Phage display selection for enzyme evolution.[75]

Another category of selecting methods is based on organismal survival as a basis for selection. To this purpose, the desired enzyme property is correlated with the survival of the cell in which the enzyme is expressed through an imposed selective pressure. A clear example is the evolution of enzymes that confer **antibiotic resistance**,[80] but it is also possible to carry out a reporter-based selection and evolve other proteins by linking the reporter activity to the activity of the target enzyme. This **reporter-based selection** can be performed through transcriptional-regulator-based strategies, such as the AraC-based three hybrid system, which uses arabinose operon activator AraC for selection.[81] Another reporter-based technique is riboswitch/ribozyme-based strategy, which consists on the binding of a specific small molecule, often the product of an enzymatic reaction, to the riboswitch/ribozyme, and the reporter gene is turned on and can be selected.[82] Thermostable antibiotic resistance reporters can be used as a selecting method of thermostable variants that would permit de proper folding of the reporter and hence the growth on antibiotic-containing plates.[83] Selection for complementation of amino acid auxotrophy can also be applied for those bacteria whose growth require the production of a particular amino acid that is catalyzed by the triggered enzyme.[84]

Selection can also be carried out with *in vitro* compartments. Once again, the use of *in vitro* techniques has some advantages compared to *in vivo*, such as overcoming transformation efficiency bottlenecks and host genome mutations that unexpectedly

influence selection survival. **Compartmentalized self-replication (CSR)** is *in vitro* selection for DNA and RNA polymerases that replicate their encoding gene in an emulsion PCR.[85] In **compartmentalized partnered replication (CPR)**, the evolving enzymatic activity triggers the expression of Taq polymerase. Higher Taq expression levels drive to better PCR amplification of the active library members.[86]

1.3. Computational Enzyme Modeling and Design

As mentioned before, directed evolution of enzymes are giving promising results, but the vast number of possible combination of mutations supposes the main limitation for the generation and screening of the desired chemical space. The generation of "smart libraries" of smaller size that target specific residues has proven to cover functional space more effectively. However, the generation of these smart libraries often requires structural and functional knowledge that is not evident for many systems. Many computational approaches for enzyme design and engineering are focused on the design of those smart libraries, the identification of mutational "hotspots" or reshaping of the catalytic active site.

1.3.1. Knowledge-based Methods

Some key positions of structural or functional relevance can be identified with the **alignment of multiple homologous genes**[87]. A better computational capacity allows the alignment of more distantly-related proteins. High conservation of a residue correlates mainly with protein stability, but can also be relevant to other enzyme properties, such as reactivity or selectivity. These consensus positions are often preserved to prevent a loss of function, but sometimes their modification can suppose a huge increase in enzyme properties. Sequence alignment may fail to identify homologues due to the low conservation of protein sequence throughout evolution. Structural conservation is much more frequent than sequence conservation and thus is a better target for protein alignment. The majority of methods used in structural homology modelling can be further grouped into two types: **comparative modelling (CM)**,[88] characterized by a high homologous template (> 30% sequence identity); and the **threading approach**,[89] with a better performance by modelling distant templates. However, over 3% of protein structures of the protein data bank (PDB) still have an unknown function, and the identification and alignment of very distant structural homologues is computationally demanding.

Further increasing the degree of complexity, beneficial mutations within a library can be identified through **statistical analysis of protein-activity relationships (ProSAR)**. Codexis Inc. has used this statistical correlation approach to improve the reaction ratio of halohydrin dehalogenase (> 4,000-fold).[90] They performed random amino-acid substitutions in the dehalogenase and measured the rate of catalysis by the variants to use statistical methods for the identification of a particular pattern (such as beneficial mutations) common to improved variants.

The computational methods described above can give a good guidance to identify mutational "hotspots", but they do not directly consider the substrate of interest. Substrate-specific knowledge is fundamental for many systems with a naturally broad substrate scope or for enzymes that are engineered to bind a new substrate or inhibitor. Virtual substrate screening can be carried out using **molecular docking** tools such as AutoDock.[91] By providing a structure of a protein and ligand, molecular docking will predict possible binding poses and help to identify mutational "hotspots", as Notonier et al. did with cytochrome P450 CYP153A.[92]

1.3.2. Molecular Dynamics Simulations for Enzyme Modeling and Design

All the aforementioned computational methods treat the protein with a static structural approach. Nevertheless, enzymes can adopt a broad range of **conformational states** (or sub-states) in thermal equilibrium separated by small energy barriers. Statistical thermodynamic distributions are used to represent all the conformational states and sub-states in the **free energy landscape**.[93] Regions with high populations of specific conformers depict local or global energy minima, while the velocity of the conformational transition depends on the height of the energy barriers. These conformational landscapes can be characterized in a limited way through experimental techniques such as X-ray crystallography, cryo-electron microscopy, NMR or biophysical techniques (fluorescence, circular dichroism or Raman spectroscopy).[94] However, computational methods are particularly useful in reconstructing the free energy landscape of enzymes.

Molecular dynamics (MD) techniques are based on the integration of Newton's laws of motion to sample the population distribution of atoms or biomolecules. During MD simulations, the interactions between atoms are used to calculate the energy of the molecule, and an initial velocity is assigned to all the atoms in the molecule, making possible a fluctuation in the spatial position of atoms. The extent of the fluctuation depends on the interaction forces and influences the potential energy.[95,96]

The huge dimensional space generated by the large number of atoms that compose a protein makes the analysis of enzyme thermodynamic properties particularly difficult. **Dimensionality reduction** is a necessary approach to solve this problem and it can be addressed in a rational way, focusing in global or collective **degrees of freedom (DOF)**, or in an automated way with a **Principal Component Analysis (PCA)**. DOFs are explicit functions relevant for the protein, such as coordinates, distances between catalytic residues, backbone dihedral angles or the root-mean-square deviation (rmsd) of selected substructures. PCA is a way of automatically reduce the dimensionality of the data while preserving as much relevant information as possible, grouping in a principal component all the strongly correlated atomic motions.[97]

The potential function of a protein is simulated using accurate **force fields** that are divided into different energy terms.[98] Different force fields are available, the most commonly used ones being HARvard molecular mechanics (CHARMM),[99] assisted model building and

energy refinement (AMBER),[100] optimized potential for liquid simulations (OPLS)[101] and Groningen molecular simulation (GROMOS),[102] Most of them are developed using *ab initio* quantum mechanical and molecular mechanics calculations. Knowledge-based force fields like Rosetta[103] use extra potential energy terms obtained after refitting of statistical and experimental knowledge-based data.

Atomistic MD simulations use empirical force fields with a time step in the order of femtoseconds (10^{-15} seconds), which limits the total simulation time to the scale of microsecond (10^{-6} seconds) for an experiment that lasts days with the current technology. However, interesting allosteric transitions occur in the millisecond to second timescale (**Figure 8.a**). Thus, the main bottleneck of MD is the simulation time required to navigate all the conformational landscape (**Figure 8.b**).[93] Various approaches have been developed in the last years to overcome the sampling problem.

Unbiased MD methods can be a useful tool to reduce the conformational sampling map:

- **Central Processing Unit (CPU) and particularly Graphics Processing Unit (GPU) parallelization** increases the timescale through breaking the system down into smaller entities.[104]

- The special-purpose **Anton supercomputer** was designed to perform micro- to milliseconds single MD simulations with extreme computing efficiency.[105]

- **Replica exchange or parallel tempering** is based on running several copies of the same system at different temperatures in order to facilitate configurations exchange and barrier crossing.[106] Then, all the simulations are combined to recover the associated conformational free energy landscape.

- **Markov State Models (MSMs)** are based on dimensionality reduction techniques and short molecular simulation trajectories to identify the kinetically relevant slow states and their conversion rates.[107]

- Most recently, **crowd-sourced structure prediction** (such as Folding@home, Rosetta@home or Foldit)[108] through online multiplayer games have demonstrated their effectiveness to statistical and deterministic search algorithms. The Foldit project has shown that basic spatial recognition, intuition and decision making can out-compete the stochastic component of conformational search when applied to problems of protein-structure prediction.

Conformational sampling issue can also be dealt with **biased MD methods (Figure 8.c-e)**:

- **Accelerated MD (aMD)**[109] is a useful method to explore conformations without initial structural knowledge. aMD is based in the use of a boost potential to raise energy wells that are below a certain threshold level, with none alteration of those above this level. The reduction of adjacent energetic barriers allows the system to sample a more extended conformational space than the one sampled with classical MD. aMD is applied when few structural information is available and only one conformational state is known, but there is no clear information about the transition state (TS). aMD have been recently used by Chen et al. to decode the effect of disulfide bonds in the β-amyloid cleaving enzyme 1 (BACE1).[110]

- **Metadynamics (MetaD)**[111] is another approach to speed up slow processes, forcing a system to sample all available conformations and generating a complete free energy landscape for chosen characteristics of the system. MetaD uses a biased potential to force the system to leave local minima by adding a Gaussian hill to the potential energy of the current region of state space. MetaD is useful when conformational states are known but no information about TS is available. MetaD has been used by Saladino et al. to prove the effect of mutations in the conformational dynamics of some protein kinases related with cancer.[112]

- When both conformational states along with intermediate conformations are known, **umbrella sampling (US)**[113] is one the best options. In US, bias potentials (often harmonic potentials) along with one or more dimensional reaction coordinate intermediate steps are covered by a series of windows for which an MD simulation is performed. After the simulations, weighted histogram analysis method (WHAM) could recover the unbiased free energy by the umbrella integration.[114] Recently, US have been used to estimate the affinity between acetylcholinesterase (AChE) and its inhibitor, a key target to prevent Alzheimer's disease.[115]

An enzyme in solution can adopt multiple conformational states that define the free energy landscape. A population shift between all these conformational states can occur when a ligand is bound to the catalytic site of the enzyme or when the enzyme is mutated in allosteric sites. One of the main goals in enzyme design is the prediction of target residues to mutate for activity improvement or novel functionality. MD simulations are sometimes utilized for predicting beneficial point mutations during re-design of active sites, both by *in silico* studying the dynamics of the catalytic residues and by predicting the role of substitutions within the active site cavity. This approach was used by Monza et al. to engineer the active site of a fungal laccase to improve substrate oxidation.[116] Active site and distal positions that by mutations can provoke a population shift can also be identified by exploring correlated movements and distances between all protein residues through MD simulations.[117] The obtained **shortest path map (SPM)** is constructed with pairs of residues that have a high contribution to the communication pathway. DynaComm.py is a python code developed by Osuna et al. that implements this approach to identify mutational "hotspots".[118]

a

loop motions

allosteric transitions

α-helix folding

β-hairpin folding

bond vibrations

side-chain rotation

protein folding

| 10 fs | 1 ps | 1 ns | 1 μs | 1 ms | 1 s |

Molecular Dynamics

b

Folded state A

PC2

Folded state B

Unfolded state

PC1

c

Free energy

Energy threshold

A

B

Conformations

d

Free energy

A

B

∧ Potential function

Conformations

e

Free energy

A

B

◿ Harmonic potential

Conformations

Figure 8. Conformational sampling through biased Molecular dynamics (MD) methods. **(a)** Time scales of different types of protein motions.[119] **(b)** Conformational protein landscape after a Principal component dimensionality reduction analysis. PC1: Principal component 1. PC2: Principal component 2. Red areas represent low energy conformations; blue areas represent high energy conformations. **(c)** Accelerated MD approach. **(d)** Metadynamics approach. **(e)** Umbrella sampling approach.[93]

Another important aim in protein engineering is the increment of protein stability against temperature, solvent and other harsh conditions. One of the used techniques to increase the stability of proteins is the introduction of disulfide bonds to the protein structure, which lowers the entropy of the unfolded state and kinetically stabilizes the protein[120]. Nevertheless, in some cases the reverse effect has been observed. By this way, MD simulations can be useful to determine flexible regions proper to introduce disulfide bonds and reduce the trial-and-error examination. MD simulations of the haloalkane dehalogenase (DHlA) revealed high mobility in a helix-loop-helix region in the cap domain, and the mutated protein D16C/A201C demonstrated significant changes in thermal stability[121]. Thermostability of lipase A from *Bacillus subtilis* (BsLipA) and some of its variants has been thoroughly studied through experimental techniques,[122–125] but also from a computational approach that applied the rigidity theory-based Constraint Network Analysis (CAN), which correlates the structural rigidity and the thermodynamic thermostability of proteins. Robustness of rigidity analysis was improved by generating an ensemble of structures with MD simulations.[126]

1.3.3. QM/MM Tools for Enzyme Modelling

The modelling of an enzymatic reaction is challenging due to changes in the electronic structure of reactants during catalysis. **Quantum mechanical methods (QM)** are considered the most accurate description of these electronic changes. However, these methods are not suitable for molecules with many atoms, such as proteins. On the other hand, theoretical methods such as **molecular mechanics (MM)** can model systems of hundreds of atoms but cannot describe electronic changes. Enzyme catalytic mechanisms are modelled through the combination of QM and MM methodologies in an approach called **hybrid QM/MM**. This technique was applied for the first time to a reaction catalyzed by lysozyme.[127] Hybrid QM/MM methods carry out a two-scale division to describe the energetics of a macromolecule: the region where chemical reactions or electronic rearrangements occur, which is treated with QM, and the rest of the enzyme, which is treated with MM. Interaction terms are required between QM and MM regions, normally calculated with computational methods like density functional theory (DFT) and *ab initio* methods.[128]

Although QM/MM has not been widely applied to enzyme design, important advances with this methodology have been performed for the family of metalloenzymes. In a recent study, the structure and energetics of different hydrolysis paths of the matrix metalloproteinase-2, a Zn protease that is a promising drug target, was elucidated through QM/MM.[129] Vidal-Limón et al. also used the QM/MM methodology to find electron pathways involved in the suicide inactivation, a common process in heme peroxidases.[130] Other illustrative examples of QM/MM application in enzyme modeling are the quantification of the electrostatic factors governing the enantioselectivity in *Candida antarctica* lipase A, a serine hydrolase, during the hydrolysis of a model ester[131] and the engineer of a glycerol dehydrogenase with an expanded substrate scope.[132]

QM/MM and MD simulations can be performed together to investigate the reaction mechanisms of enzymes and study the protein-inhibition interactions facilitating the search of suitable active enzymes. An example is the study by Silva et al. of *Mycobacterium tuberculosis* L,D-transpeptidase which catalyzes the cell wall formation of peptidoglycans that help in the resistance against β-lactams.[133]

1.3.4. *Ab initio* Enzyme Design

In recent years, computational algorithms have become increasingly reliable to both predict protein sequences that would produce a stable fold and create functional proteins that can promote non-natural chemical reactions. The main challenge in the design of new enzyme functions relies in the rational prediction of mutations required to favor the catalytically active conformation.

With that aim, Baker and Houk groups have developed the **"inside-out" protocol**,[134] which relies on the construction of a **theozyme** (a theoretical active site) with the appropriate catalytic functionality (**Figure 9**) and the determination of the geometric arrangement of the transition state (TS) through QM calculations. This theozyme model is then grafted in protein scaffolds selected from the protein data bank (PDB)[135] with **RosettaMatch**. This software relies on hashing techniques and prunes the majority of potential active centers at a very high speed but very little cost. Then, **RosettaDesign** mutates and optimizes amino acid residues surrounding the QM theozyme to ensure good packing and fold stability and complements the geometric and electronic features of the TS.[103,136] After optimizing all unique matches, designs with a predicted optimal performance are selected for experimental validation. As it is unlikely that a design has the highest score for each factor, extensive examinations that can be guided by MD simulations are the most effective way to predict the catalytic potentiality of designs. The design of Kemp eliminases,[137–139] retro-aldolases[140,141] and Diels-Alderases[142,143] for example, was carried out with this "inside-out" protocol. The enhancement of target reactions by designed enzymes was assessed by the ratio of the catalytic rate constant and uncatalyzed rate constant k_{cat}/k_{uncat}. In the above cases, the values of k_{cat}/k_{uncat} ranged from 10^2 to 10^5 for the most active designs, indicating the effectiveness of such design strategies.
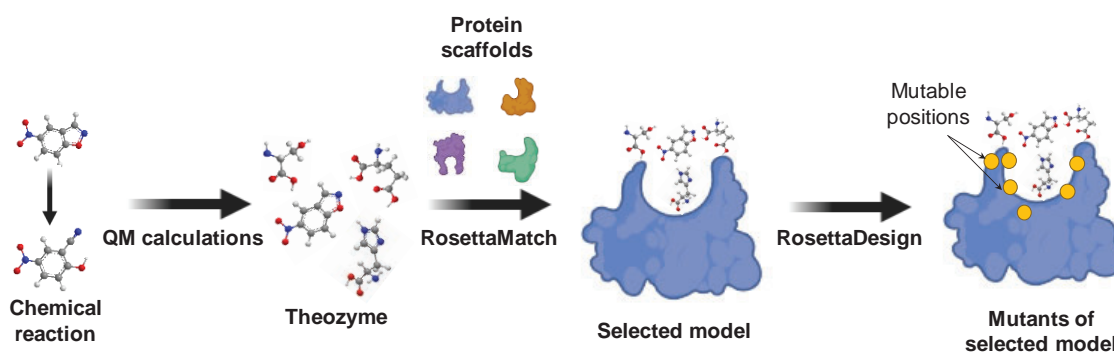
Figure 9. Steps of the "inside-out" protocol for *ab initio* enzyme design[134].

The RosettaMatch algorithm has also been used by Kuhlman and co-workers for the design of a zinc-mediated homodimer protein interface cleft with catalytic activity for carboxyester and phosphoester hydrolysis.[144] Pecoraro group has approached *de novo* redox active metalloprotein design of a Cu nitrite reductase (NiR) using three stranded coiled coil scaffolds, obtaining the most active water-soluble and stable artificial NiR catalysts yet produced.[145]

1.3.5. Machine Learning in Enzyme Engineering

As seen before, directed evolution and rational design can complement each other for the identification of mutable "hostspots". While the combination of both strategies shows remarkable results, it still requires a vast amount of computational or experimental effort. **Machine learning (ML)** is a third approach that can contribute to the reduction of this effort by identifying patterns in the existing data to predict properties of the previously unseen but similar input. ML-based design can generate new variants based on the patterns in the collected data[146,147].

Different ML algorithms are strong in predicting a particular enzyme property: random forests have been used to predict protein solubility, support vector machines and decision trees are good describing enzyme stability changes upon mutations, K-nearest-neighbor classifiers can elucidate enzyme function and mechanisms, and various scoring and clustering algorithms are the best option for rapid functional sequence annotation. The main strength, and at the same time weakness of ML in enzyme engineering, is the high dependence for the success of the prediction on the training data set. This strong dependence allows the algorithm to potentially make predictions about new variants almost immediately after training. However, the great diversity of enzyme mechanisms and reactions requires a high degree of representativeness on the training data set to assure its quality, which requires a large amount of experimental data that are not always available.

1.4. Enzyme Immobilization

Directed evolution and computational approaches explained above trigger the engineering of the enzyme sequence to improve its properties as a biocatalyst. Despite the achievements obtained through these protein engineering approaches, the industrial application of engineered enzymes still often hampered by their lack of long-term operational stability, shelf-storage life and their hard going recovery and reuse. The biocatalyst can be then optimized to its industrial application through enzyme immobilization techniques. **Enzyme immobilization** refers to the state of enzyme molecules confined into or attached to solid materials where enzymes are in different phase from the reactants, but substrates and products can diffuse between the solid and liquid phases. That is why these enzymatic systems can also be referred as **heterogeneous biocatalysts**.

Immobilization of enzymes guarantees a suitable physical form of the catalyst to enable its recovery at the end of the biotransformation, as well as its reuse or even its integration into continuous flow systems, reducing the cost of the final product. Moreover, the solid-liquid composition of heterogeneous biocatalysts facilitates the handling and separation of the product from the biocatalyst, minimizing the contamination of the product and simplifying the downstream processing. Finally, the enhanced stability conferred to the catalyst through the immobilization process lowers enzyme inactivation by external agents such as temperature, pH and organic solvents and increases its operational time.

1.4.1. Enzyme Immobilization Techniques

Enzymes may be immobilized by a variety of methods, which can be broadly classified depending on the nature of the interaction between the enzyme and the support as physical or chemical, and as reversible or irreversible. **Reversible methods** are characterized by weaker, non-covalent interactions such as hydrogen bonds, hydrophobic interactions, van der Waals forces, ionic binding or physical entrapment of the enzyme within the support. On the contrary, **irreversible methods** involve the formation of covalent bonds. Given the inherently complex nature of protein structure, no single method is ideal for all purposes. Basically, the most used immobilization methodologies are the following (**Figure 10**):

- **Physical adsorption**: enzyme is bound to the support through weak non-covalent forces, such as van der Waals, hydrophobic interactions, hydrogen bonds or salt linkages. Ionic adsorption is the most common one due to the abundance of charged residues in the enzyme surface, which can easily interact with a support activated with complementary charged groups. The major advantage of this technique is that neither additional coupling reagents nor modification of the protein of interest is required. Moreover, the enzyme can be reversibly removed from the support under gentle conditions (by modifying the ionic strength of pH in case of ionic adsorption), which is a great advantage to easily regenerate the support when the enzyme activity has decayed. This weak binding also implies an operational disadvantage, as enzyme can easily be leaked out of the matrix. Another relevant drawback is that adsorption of proteins onto surfaces often results in conformational changes or even denaturation. Finally, the lack of control over the packing density of the immobilized enzymes can be traduced in an excessive crowding that diminish their activity.[148–152]

- **Affinity immobilization**: it is based on the high affinity of protein domains with small organic molecules (or even macromolecules) displayed at the surface of the carrier. A widely exploited interaction is the one between the imidazole ring of histidine and metal ions (Co^{2+}, Cu^{2+}, Zn^{2+} or Ni^{2+}). A very common strategy implemented in the last decades is the genetic modification of the enzyme through the introduction of a polyhistidine tag coupled to its protein sequence and the activation of the support with a metal to favor chelation with the polyhistidine tag. This immobilization method is broadly also used for protein purification and is often termed as immobilized metal affinity chromatography (IMAC). Control over the orientation of the immobilized

enzyme and the minimal conformational changes resulting in a high retained activity, are key advantages of this method.[153–156]

- **Covalent binding**: enzyme binds trough the side chain of one of its residues, such as lysine (ε-amino group), cysteine (thiol group) and aspartic and glutamic acids (carboxylic group), which performs a nucleophilic attack to the reactive groups (epoxides, aldehydes, alkenes, cyanate esters or imidocarbonates) of the support. Covalent binding can be reversible (imine and disulfide bonds) and irreversible, although reversible covalent unsaturated bonds are frequently reduced chemically to become irreversible. Covalent bonds provide powerful linkages between the enzyme and the matrix and prevent enzyme release, which considerably improves the biocatalyst reuse. The method also increases half-life and thermal stability of enzymes. Nonetheless, the irreversible nature of the immobilization does not allow recycling the support when the enzyme is inactivated.[153,157–159]

- **Physical entrapment/encapsulation**: the enzyme is entrapped in a support or inside of fibers, either in the grating structure of a material or in polymer membranes. The encapsulating material can be modified in order to generate an ideal microenvironment for the enzyme (optimal pH, polarity or amphilicity). Many materials can be used to that goal, including mostly polymers and sol-gels, but also vesicles built with biomolecules, such as liposomes and proteinosomes. The simplest and most common technique of enzyme entrapment is the gelation of polyanionic or polycationic polymers by the addition of multivalent counter-ions. Enzyme mobility inside the material without direct contact with the carrier is one of the main advantages of this technique. However, this method is considerably constrained by mass transfer limitations of the substrate to the enzyme active site, enzyme leakage from the support when the pores of the matrix are too large and mechanical damage during usage.[160–163]

- **Cross-linked enzyme aggregates (CLEAs)**: it is another irreversible carrier-free method for enzyme immobilization. The enzyme acts as its own carrier, forming cross-linkages through the free amino groups of lysine residues with a cross-linking agent (commonly glutaraldehyde), resulting from inter- and intramolecular aldol condensations. Cross-linking can occur through both Schiff's base formation and Michael-type conjugate addition to α,β-unsaturated aldehyde moieties, and it is pH dependent. CLEAs are produced by simple precipitation of the enzyme from aqueous solution by the addition of salts or water miscible organic solvents or non-ionic polymers. Non-covalent interactions maintain these physical aggregates together without denaturation. Cross-linking of these physical aggregates allows their insolubility and at the same time the preservation of their pre-organized superstructure and hence of their catalytic activity. Unfortunately, industrial application is limited by their mechanical instability and the severity of mass transport, so their uses have been restricted for biosensing applications.[164–167]
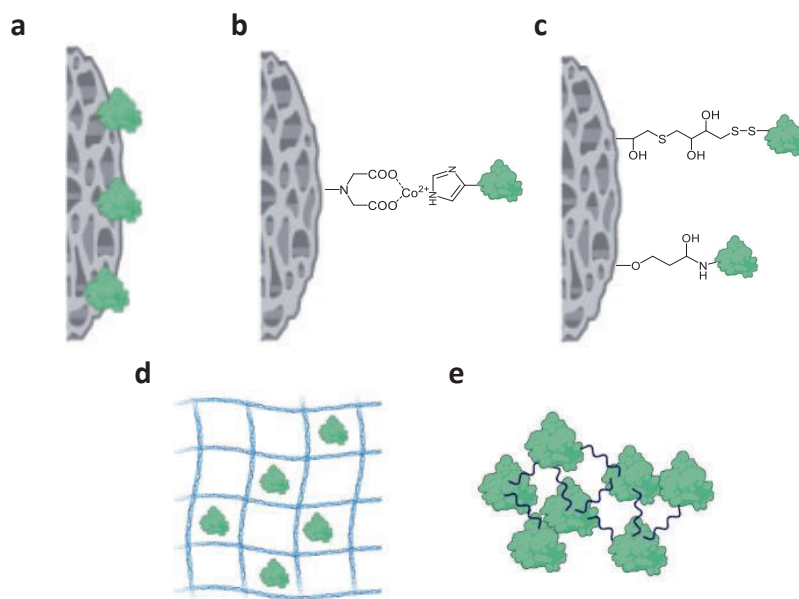
Figure 10. Enzyme immobilization techniques. **(a)** Physical adsorption. **(b)** Affinity immobilization. **(c)** Covalent binding. **(d)** Physical entrapment/encapsulation. **(e)** Cross-linked enzyme aggregates (CLEAs).

1.4.2. Support Materials Used for Enzyme Immobilization

Chemical, biochemical, mechanical and kinetic properties of an immobilized enzyme directly depend on the properties of the enzyme, but also on the properties of the material in which the enzyme is immobilized. Most employed supports can be synthetic organic polymers, biopolymers, inorganic materials, smart polymers and hydrogels. No universal material has been described for all different applications in immobilized enzymes.

- **Synthetic organic polymers**: most representative synthetic organic polymers used for enzyme immobilization are methacrylic resins, such as Purolite and Relizyme.[168] They are highly hydrophilic and both mechanically and chemically stable. Amberlite XAD-7 is another acrylic resin that can be used to covalently attach enzymes, and it was notably used for *C. antarctica* lipase B (CALB) immobilization by Novozymes.[169] However, the most commercially exploited methacrylic resin for CALB immobilization *via* interfacial activation is Lewatit VC OP 1600, which gave rise to the most ever commercialized biocatalyst, Novozyme435.[170] The main issue of acrylic resins are diffusion limitations. Plastic materials such as nylon[171] or polyurethanes[172] are other synthetic organic polymers relevant for enzyme immobilization.

- **Biopolymers**: a wide range of polymers of natural origin, such as polysaccharides including cellulose, agarose, starch, pectins, chitin, carragenans and chitosan, as well as some proteins, such as albumin, gelatin and collagen are mainly used as supportive materials for enzyme immobilization. These materials are disposed in matrixes that form very inert aqueous gels characterized by a high mechanical strength. Their

chemical structure enables their easy activation to bind proteins both reversibly and irreversibly, mostly with aldehyde, carbodiimide, epoxide, hydrazide or active ester groups[173]. A clear advantage of these polymers is their natural availability: agro-industrial wastes have been used to develop efficient cellulose matrices for lipases or transaminases immobilization,[174,175] and a lignin-based matrix have been successfully implemented in flow biocatalysis.[176]

- **Hydrogels**: polyvinyl alcohol (PVA) cryogels have been widely used for immobilization of whole cells.[177] However, enzymes' size is too small and they can diffuse out of the gel matrix. Nevertheless, with a cross-linking immobilization strategy, the use of PVA cryogels increases its effectiveness, and they can be easily prepared and degraded.[178] Poly (lactic –co-glycolic acid) (PLGA) is another relevant polymer utilized for enzyme immobilization through cross-linking in this type of hydrogels.[179]

- **Inorganic materials**: a huge variety of inorganic materials can be used for enzyme immobilization, such as alumina, silica, glass, zeolites and ceramics. Silica-based supports are the most suitable matrices due to their high mechanical strength, their easy functionalization with chemical groups and their easy fabrication to provide desirable morphology, pore structures and micro-channels. Furthermore, silica gels are chemically inert and therefore environmentally friendly for chemical manufacturing. Enzymes can be easily immobilized in silica by absorption.[180–182] Controlled porosity glass (CPG) is a silica glass composed by pores with a particular size distribution and a wide variety of geometric forms.[183]

- **Nanoparticles (NPs)**: one of the main advantages of nanostructured materials is the control over size at the nanometer scale, such as the pore size in nanopores, thickness of nanofibers or nanotubes and the particle size of NPs. Gold NPs (AuNPs) are particularly interesting due to their attractive electronic, optical, thermal and catalytic properties.[184] Magnetic NPs (MNPs) have also gained prominence as they can be easily separated and recovered by applying an external magnetic field. The most common metal for MNPs synthesis is iron oxide (Fe_3O_4), due to its high biocompatibility, non-toxicity and ease with which it may be used to bind enzymes. MNPs have been used to immobilize enzymes in order to enhance their stability, to facilitate their separation from the reaction mixture, to the manufacturing of biosensors and to facilitate diagnosis and treatment of diseases.[185–187]

- **Smart polymers**: covalent attachment can also be achieved through stimulus-responsive or smart polymers that undergo dramatic conformational changes in response to small changes in their environment. The most studied example is the thermoresponsive and biocompatible polymer poly-*N*-isopropylacrylamide (polyNIPAM).[188] Its main advantage is the easy recovery of the enzyme by raising the temperature.

- **Metal-organic frameworks (MOFs)**: these crystalline porous materials are composed of metal ions or clusters linked by organic ligands. Their main strengths are their high surface area, their tunable ultra-high porosity, their designable functionality and their excellent thermal stability. Moreover, it is possible to customize the pore size and optimize it for a given enzyme, which allows the compact structure of the MOF to tightly confine the encapsulated enzyme, hugely increasing its stability.[189]

1.5. Flow Bio-processes

As aforementioned, immobilization of enzymes in solid materials represents a valuable tool for pharmaceutical and synthetic chemistry. Normally, immobilized enzymes are implemented in batch processes, due to their fast process and to their product traceability. However, **continuous flow systems (CFS)** present several advantages when compared with batch reactors, as they can synthetize more product in a given time, they are more easily transferrable to large-scale production and their modularity permits a much better control of the setup of the reaction (**Figure 11**).[190] CFS requirements are not particularly demanding and allow for the automation of the process. CFS strategies also guarantee the control in conditions such as pressure, temperature or reactivity in a much safer way than batch. In addition, process intensification and feedback loop strategies can be implemented with continuous operations and usually improve the cost-efficiency. In particular, multiphasic reactions benefit from flow systems as for example gaseous reagents often present poor interfacial mixing due to their low solubility, and pressurized flow reactors can increase this solubility.[191] Another key advantage is the easy integration of analytical devices to real-time monitoring *in situ* the reaction. Additionally, the continuous removal of product and substrate in flow reduces enzyme inhibition, improving the reaction rate of the enzyme.[192] Finally, continuous flow systems can also facilitate the performance of enzymatic and chemical cascade reactions by combining more than one reactor in series.[193–195]
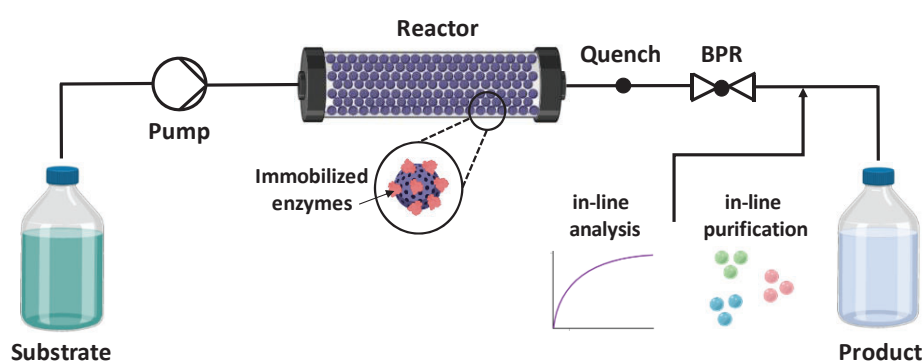
Figure 11. Classical Continuous Flow System (CFS) composed of the substrate delivery pump, the reactor, a quenching system, a back pressure regulator (BPR), and in-line collection, analysis and purification systems.

Flow reactors are normally miniaturized in a laboratory scale to utilize devices constituted by channels or tubes with different diameters from μm (microreactors) to mm

(mesoreactors). **Microreactors** are characterized by a laminar flow and they are more effective in heat/mass transfer and temperature control, but mixing is limited by diffusion, while channel obstruction can easily occur.[196,197] **Mesoreactors** can present a turbulent flow when they operate at high flow rates, and they usually overcome issues related with mixing efficiency, that is why these kind of reactors is more extendedly used.[198,199] Depending on how the enzyme immobilization is carried out, reactors can be classified in two types: **wall-coated reactors (WCR)**, in which the enzyme is directly immobilized on the reactor's walls or on a membrane;[200,201] and **packed-bed reactors (PBR)**, in which the supported-enzyme particles are packed in a column.[202] WCR are often applied for microfluidic systems where solid carriers could cause obstructions, while PBR are more popular for any mesoreactor design.

When a biotransformation is translated from a batch to a flow system, there are several parameters that need to be considered:[203]

- The **Reynolds number (Re)** is the ratio of inertial forces to viscous forces within a fluid, and it is calculated from the density of the fluid (ρ), the flow speed (υ), the diameter of the tube (d) and the dynamic viscosity of the fluid (μ) *via* the following equation:

$$Re = \frac{(\rho \, \upsilon \, d)}{\mu}$$

- The **residence time (τ)** is the time the reactants spend inside the reactor, and it is calculated from the reactor volume (V) and the volumetric flow rate (Q) *via* the simple equation:

$$\tau = \frac{V}{Q}$$

- The **space time yield (STY)** describes the amount of product (m_P) formed within the residence time (τ) in the reaction volume (V):

$$STY = \frac{m_P}{\tau \, V}$$

- The **specific productivity (SP)**, which informs of the time-averaged efficient use of the enzyme and is calculated from the amount of product (m_P), the amount of immobilized enzyme (m_e) and the residence time (τ) *via* the equation:

$$SP = \frac{m_P}{m_e \, \tau}$$

- The **total turnover number (TTN)** that means the enzyme operational stability under operation conditions, and is calculated with the moles of product (n_P) and the moles of enzyme (n_e) with the following equation:

$$TTN = \frac{n_P}{n_e}$$

Flow bio-transformations have been utilized for many synthetic processes, but pharmaceutical industry has been the first field in which flow biocatalysis has been applied and still leads the utilization of this technique for manufacturing fine chemicals with high conversion yields. Many types of enzymes have been implemented in flow processes, such as lipases,[204] dehydrogenases,[205] lysases,[206] phosphatases and aldolases,[207] among others.

1.6. The Role of Acyltransferases in Lipids Metabolism and their Applications in Chemical and Pharmaceutical Industry

Long-chain hydrophobic acyl esters are essential molecules that conform a huge variety of biological structures and take part in vital biological processes. To name only a few relevant functions, they constitute the inner hydrophobic layers of cell membranes, they have a storage function inside the cell, and they can act as membrane anchors that modify protein properties or functions. Acyl-thioesters are obtained through the esterification of fatty acids by the attack of the thiol group of coenzyme A (CoA) or of an acyl carrier protein (ACP) on to their carboxyl group. These reactions are catalyzed by the enzyme family of acyltransferases (AT). AT catalyze reactions of great use for the chemical industry to obtain valuable compounds such as polyhydroxyalkanoates (PHA), polyketides with pharmacological properties and many esters with flavor and fragrance properties. Thus, this family of enzymes is a promising target for novel biocatalysts design.

1.6.1. Acyltransferases Involved in Long-Chain Storage Lipid Synthesis

Lipids are molecules with ideal storage properties, such as high caloricity, water insolubility and inert osmotic behavior. Storage lipids in living organisms are usually esterified fatty acids such as poly-(3-hydroxyalkanoic acids) (PHA), triacylglycerols (TAG), wax esters (WE), phospholipids (PL) or glycolipids (GL). PHA is the main energy storage system for prokaryotes, whereas TAG is the one for eukaryotes (although they are also present in prokaryotes) and WE rather fulfill more specialized purposes.

Triacylglycerols (TAG) have obvious applications in the synthesis of oils and fats, but they are also progressively being used for the lipases-catalyzed synthesis of biodiesel or **fatty acid alkyl esters (FAAE)**, consisting of TAG-derived fatty acids esterified with short-chain alcohols, as substitutes for diesel fuel.[208] TAG biosynthesis is known as the Kennedy

pathway,[209] and its starting point comes with successive acylation steps from glycerol-3-phosphate (G3P) to 1,2-diacylglycerol (DAG), which is finally re-acylated with acyl-CoA to obtain TAG in a reaction catalyzed by an **acyl-CoA:diacylglycerol AT (DGAT) (Scheme 1)**.[44] DGAT belongs to the **membrane-bound-O-acyltransferase (MBOAT)** enzyme family, which is composed of multispan membrane enzymes that utilize long-chain or medium-chain fatty acyl-CoA and hydrophobic molecules as their substrates. An invariant histidine and a near-invariant asparagine within a long stretch of hydrophobic residues are characteristic features of those enzymes.[210]

Wax esters (WE) list of applications is also extensive, as they can be employed in the commercial production of cosmetics, candles, printing inks, lubricants, food additives and coatings. WE can be obtained through the esterification of long-chain fatty alcohols and CoA-activated fatty acids (acyl-CoA) catalyzed by **wax synthases (WS/DGAT) (Scheme 1)**. WS/DGAT allows WE synthesis from cheap renewable sources, such as fatty acids or carbohydrates. **AtfA** from *A. baylyi* is a model enzyme from the WS/DGAT family which catalyzes the synthesis of TAG or WE from acyl-CoA and DAG or fatty alcohols.

As WS/DGAT can accept ethanol as a substrate, they have the ability of synthesize FAAE. In the "microdiesel project", a recombinant *E. coli* strain was engineered to contain the required genes for ethanol production and the *atfA* gene from *A. baylyi* ADP1 on one plasmid in order to directly synthesize FAAE.[211–213] The major bottleneck was found to be the low specificity of the enzyme towards ethanol, but this limitation could be perfectly solved with several tools such as directed evolution. It is precisely with this technique with which *T. curvata* WS/DGAT has been successfully evolved towards the improvement of TAG and WE production by introducing mutations in the outer surface of the enzyme that increase the surface hydrophobicity, which could stabilize lipoprotein complexes allowing the storage of a greater amount of lipoprotein aggregates.[214]

Cholesterol is a lipid molecule present in membranes of all mammalian cells and is required for their growth and viability. The excess of cholesterol is stored as cholesteryl esters (CE), which conversion is catalyzed by **acyl-CoA:cholesterol AT (ACAT) (Scheme 1)**, an enzyme of the MBOAT family.[215] The accumulation of CE in macrophages is one of the causes of atherosclerosis, which makes ACAT an appropriate target for therapeutic treatment of the abovementioned disease.[216] This target could be an alternative to cholesterol synthesis inhibition through statins, which monotherapy cannot be efficient in all cases an even provoke side effects.
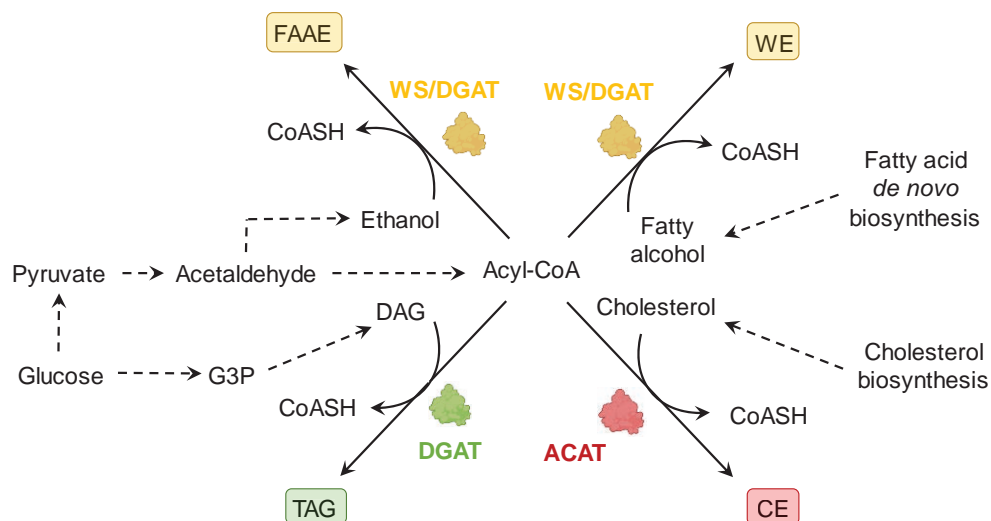
Scheme 1. Biosynthetic pathways of Long-Chain Storage Lipids from acyl-CoA. G3P: glycerol-3-phosphate; DAG: diacylglycerol; TAG: triacylglycerol; CoA: coenzyme A; FAAE: fatty acid alkyl esters; WE: wax esters; CE: cholesteryl esters; DGAT: acyl-CoA:diacyl glycerol acyl transferase; WS: wax synthase; ACAT: acyl-CoA:cholesterol acyl transferase.

In the past years, **poly (3-hydroxyalkanoic acids) (PHA)** has attracted public interest as a promising solution for the increasing plastic waste problem. PHA can be used as substitutes for conventional plastics as they are water insoluble, nontoxic and thermoplastically deformable. **PHA synthase (PhaC)** is the key enzyme involved in PHA biosynthesis and functions by polymerizing hydroxyalkanoate substrates.[217] Due to their high applied value, efforts have been made to improve the performance of PhaC by broadening their substrate specificity through enzyme engineering. Successful enzyme evolution approaches have been applied to PhaC from *Ralstonia eutropha, Aeromonas caviae* and *Pseudomonas* sp, obtaining new engineered PhaC capable of incorporating man-made substrates into the polymer backbone.[218]

1.6.2. Acyltransferases Involved in Short and Medium Chain Esters

Alcohol acyltransferases (AATs) are the principal source of ester production in plants, yeast, filamentous fungi and some bacteria and they act by transferring the acyl moiety from an acyl-CoA molecule to an alcohol, with a wide ream of specificities for their alcohol and acyl-CoA substrates, giving rise to a plethora of esters. The most studied ester-producing AATs are derived from yeasts, mostly from *Saccharomyces cerevisiae*, being Atf1, Atf2, Eht1, Eeb1 and Eat1.[45,219]

Atf1 and **Atf2** are involved in acetate ester production and sterol metabolism, respectively, and they are localized in the endoplasmic reticulum (ER) and in lipid droplets in the cytosol. Atf1 and Atf2 share common characteristics with the already mentioned AtfA from *A. baylyi* and other AT from plants and bacteria, as the conserved HXXXD motif that

participates in transferring the acyl group from the acyl-CoA donor to the alcohol acceptor without forming a covalent acyl-enzyme intermediate. For some AATs, thioesterase activity with different substrate specificities has been observed. For example, Aft1 only accepts acetyl-CoA as an acyl donor *in vitro*, but can hydrolyze longer acyl-CoA substrates as a thioesterase. Regarding the acyl acceptor, both Atf1 and Atf2 have a broad specificity towards primary alcohols *in vitro*.[220,221] Atf1 has been utilized in metabolic engineering experiments to develop pathways that enable the production of complex esters. For example, the 2-keto pathway is used to produce such alcohols and is naturally present in yeast and some lactic acid bacteria. When the key enzyme of this pathway, 2-ketoacid decarboxylase (2-KDC) of *Lactococcus lactis* is introduced in *E. coli* along with *S. cerevisiae* Atf1, high rates of acetate esters such as isobutyl acetate can be produced directly from glucose.[222]

The second paralog pair of *S. cerevisiae* is **Eht1** and **Eeb1**, that produce medium-chain fatty acid (MCFA) ethyl esters utilized to detoxify MCFAs that accumulate during fatty acid synthesis.[223] Eht1 is located in the ER, the outer mitochondrial membrane and lipid bodies, while its paralog is located in the mitochondria. On the other hand, the role of **Eat1** is ethyl acetate synthesis from acetyl-CoA, and is key to avoid acetyl-CoA accumulation in mitochondria and to regenerate the free CoA pool in the cell.[224] All three Eht1, Eeb1 and Eat1 are defined by an α/β hydrolase fold and a Ser-Asp-His catalytic triad,[225] a structure common for a large group of hydrolytic enzymes such as proteases, esterases, lipases and peroxidases. In Eht1, Eeb1 and Eat1, a covalent acyl-enzyme intermediate is formed and transferred to the substrate alcohol. The specific environment of the catalytic pocket influences the preference of the AAT for the acyl-transfer over hydrolysis reaction through the exclusion of water from the active site and an increased binding to alcohols over water. Nevertheless, the α/β hydrolase fold confers these enzymes relatively high hydrolytic activities, such as thioesterase and esterase capabilities.[226] Few information is available for the alcohol specificities of Eht1 and Eeb1 towards both alcohols and acyl-CoAs. Besides accepting alcohols and thioesters as acyl donors, Eat1 exhibits a high alcoholysis activity using a broad range of alcohols and (to a minor extent) thiolysis.[227] The discovery of the high alcoholysis side activity of Eat1 sheds new light on the production of short chain esters in food and beverage products and might open new research lines for the industrial production of sustainable short chain esters.

Other enzymes of the α/β hydrolase family also present promiscuous acyltransferase and hydrolase activities. This is the case of **lipase A from *Candida antarctica* (CALA)**, that exhibits acyltransferase activity even in aqueous media. Biocatalytic synthesis of fatty acid esters in an aqueous media using lipases notably reduces the economic and environmental costs of the process. As CALA is naturally a lipase and thus presents a hydrolase side-activity, this enzyme is a convenient target for protein engineering strategies to reduce the production of undesired free fatty acids, which has already been performed through *in silico* rational design and enzyme immobilization.[228]

Another acyl transferase of the α/β hydrolase superfamily that has been found to preferentially catalyze ester synthesis over hydrolysis through acyl transfer reactions in aqueous solutions is the **acyltransferase from *Mycobacterium smegmatis* (MsAcT)**. The explanation for this catalytic performance against thermodynamics seems to be the hydrophobic microenvironment around the active site, but also the hydrophobicity of both the surface of the protein and the substrate entrance channel.[229] A hydrophobic tunnel might repel water and prevent it from accessing the active site, a phenomenon earlier observed for some thioesterases involved in the polyketide biosynthesis.[230] MsAcT has reported to catalyze a wide range of acyl-transfer reactions, as transesterification of primary and secondary alcohols with ethyl acetate or kinetic resolution of cyanohydrins and secondary alkynols.[231] Moreover, MsAcT displays a secondary activity in the acylation of some primary amines using methoxyacetate, ethyl and vinyl esters as acyl donors.[232] Recently, this enzyme has been rationally modified through the substitution of the catalytic serine by a cysteine with the aim of constructing a catalytic site more susceptible to nucleophilic attacks that could accept thiols and secondary amines that are not accepted by the wild-type.[233]

Table 1. Principal acyl transfer reactions for the synthesis of short and medium chain esters and amides catalyzed by Alcohol Acyltransferases (AATs).

AAT	*Acyl donor*	*Acyl acceptor*
AtfA	Palmitoyl-CoA	Hexadecanol
Atf1	Acetyl-CoA	Ethanol
Atf2		Isoamyl alcohol
Eht1	Octanoyl-CoA	Ethanol
Eeb1		
Eat1	Acetyl-CoA	
MsAcT	Ethyl acetate	Primary and secondary alcohols
		Primary amines
	Vinyl acetate	Cyanohydrins
		Secondary alkynols

1.6.3.   Polyketide Synthases

Polyketides (PK) are natural products with a wide range of desirable features for the chemical and pharmaceutical industry, including antibiotic, immunosuppressant and anti-cancer properties. This therapeutic diversity resides on the structural variability of these small molecules, which are naturally produced by the AT family of **PK synthases (PKS)**. PKS act in modular systems for stepwise chain extension from a common pool of simple precursors, in which PK chains are built by successive condensation of simple CoA-derived subunits.[234] The structural and functional modularity of these enzyme complexes as well as the diversity in the AT domain makes PKS an engineering target to obtain new biologically active compounds. PKS can be classified in three groups: Type I are multifunctional enzymes that are organized into modules, each one with a different enzymatic activity; Type II are

multienzyme complexes with a single enzymatic activity; and Type III are homodimeric ketosynthases that catalyze condensation reactions.[235]

Modular type I PKS are mostly found in actinomycetes and they are comprised of several catalytic modules, in charge of PK chain elongation, and of different enzymatic domains, each one of them carrying out a step in modification and extension of the PK chain (**Scheme 2**).[236] The order of modules and domains of a complete PKS is as follows (in the order N-terminus to C-terminus): acyltransferase (AT); acyl carrier protein (ACP) with an SH group on the cofactor, a Ser-attached 4'-phosphopantetheine; ketosynthase (KS) with an SH group on a Cys side-chain; ketoreductase (KR); dehydratase (DH); enoylreductase (ER); methyltransferase (MT); PLP-dependent cysteine lyase (SH) and thioesterase (TE). The AT domain has the role of selecting the α-carboxyacyl-CoA building block and transfer it to the moiety of the ACP domain. Then, the KS domain accepts the PK chain from the ACP and catalyzes a Claisen-like condensation with the ACP-bound α-carboxyacyl-CoA extender unit. Optionally, the fragment of the PK can be altered stepwise by additional domains. The KR domain reduces the β-keto group to a β-hydroxy group, the DH domain eliminates H_2O to yield the corresponding α,β-unsaturated alkene and the ER domain reduces the double bond. The TE domain is responsible for the final PK product release, either by attack of an external nucleophile or by cyclization.
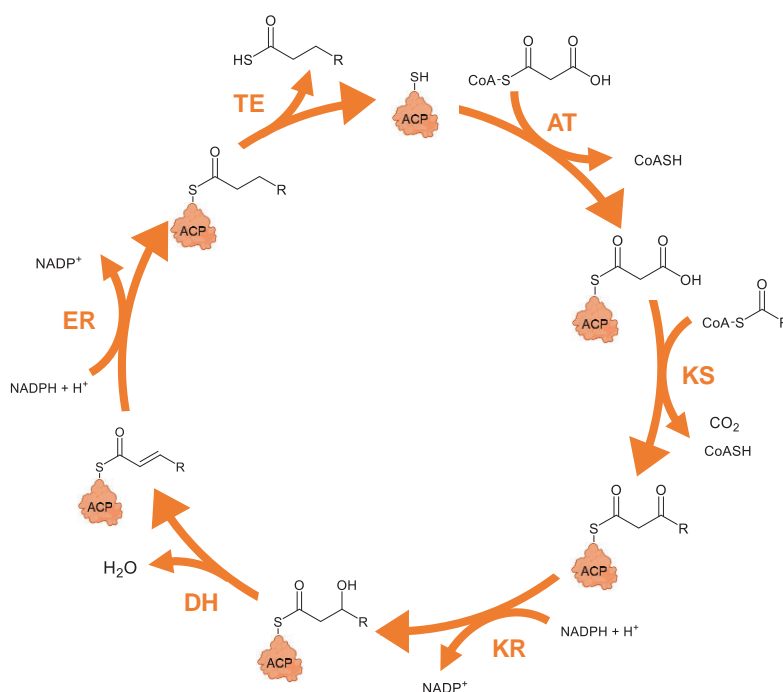
Scheme 2. Polyketide synthases (PKS) coupled mechanism. AT: acyltransferase domain; KS: ketosynthase domain; KR: ketoreductase domain; DH: dehydrogenase domain; ER: enoylreductase domain; TE: thioesterase domain; ACP: acyl carrier protein domain.

AT domains of PKS are specific for several α-carboxyacyl-CoA substrates, being malonyl-CoA and methylmalonyl-CoA the most commonly used in most biosynthetic pathways. CoA-derived extender units are incorporated *via* a ping-pong bi-bi mechanism, in which an acyl-AT intermediate is formed and subsequently attacked by the thiol of the phosphopantetheine group of the ACP.[237] Hence, protein-protein interactions between the AT (along with the KS) and the ACP domains are necessary for all activities in the assembly line. Most of the PKS domains have been shown to accept substrates considerably different from their native ones, which makes them primary targets for engineering efforts.[238,239]

Structural analysis of AT domains and of the rest of the modules make way for PKS domain engineering through different **structural-based approaches** to target acyltransferase substrate specificity (**Figure 12**):

- **AT domain swapping** from an AT homologue to another with different substrate specificity is the most common approach to engineer AT catalytic activity to produce novel PK molecules. Some examples of this strategy are the production of antibiotics from the macrolides family such as desmethyl-erythromycin, desmethyl-6-deoxyerythronolide B or desmethyl-triketide lactone analogues by swapping methylmalonyl-CoA domains in the 6-deoxyerythronolide B synthase (DEBS) for malonyl-CoA specific AT domains from the rapamycin PKS, or the production of 6-desmethyl-6-ethylerythromycin A analogue by swapping AT4 of DEBS for an ethyl-malonyl-CoA specific AT domain from the niddamycin PKS.[240–242]

- Another strategy is the generation of **hybrid AT domains** by substituting specificity-determining regions by a cassette of an analogue domain presenting a different specificity. One example is the creation of hybrid AT domains using methylmalonyl-CoA specific domains from DEBS and the malonyl-CoA specific AT domain from the module 2 of the rapamycin PKS. This research identified a short C-terminal hypervariable region as a possible determinant of AT specificity.[243]

- A less invasive way for AT specificity modification that minimizes perturbations in protein-protein interactions is **site-directed mutagenesis** of important residues. The methylmalonyl-CoA specific YASH motif have been mutated to HAFH, the specific malonyl-CoA motif in AT1, AT4 and AT6 of DEBS, leading to promiscuous AT domains capable of incorporating both extender units.[244,245]

- Despite being considered canonical, *cis*-AT domains found in actinomycetes, myxobacteria and cyanobacteria are missing in some proteobacteria and bacilli, which rather present *trans*-AT domains, another form of the AT domain that have evolved independently. Some PKS with *trans*-AT domains are responsible for the synthesis of disorazoles or kirromycin, amongst others. Hence, AT

specificity could also be engineered by **inactivating the *cis*-AT domain and complementing with a *trans*-AT domain**.[246]
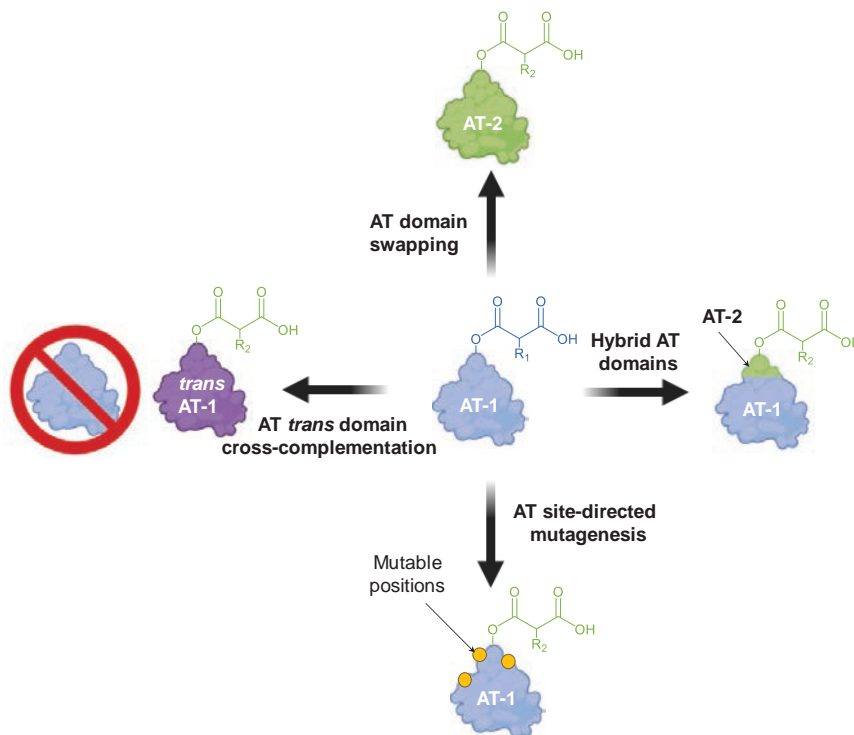
Figure 12. Strategies for acyltransferase (AT) domain engineering of polyketide synthases (PKS) to accept novel substrates.

However, the future of AT domain specificity engineering resides in targeting protein-protein interactions through a combination of computational and experimental techniques. The JDET software package[247] was recently developed to this purpose and consists of an input alignment using different methods as **multiple correspondence analysis** that allows extracting independent sources of variation in the alignment. These variations can be linked to corresponding clusters of proteins in order to identify subfamilies. Another strategy is the **statistical coupling analysis (SCA)**[248], which explicitly accounts for amino acid covariation through a conservation-weighted position correlation matrix that calculates the conservation at individual amino acid positions and their inter-residue covariance. These matrix spectra can be decomposed and analyzed through principal component analysis to identify statistically significant correlations among groups of residues. Coevolution analysis is both useful to identify residues involved in substrate specificity and to detect key positions for the protein-protein interactions that allow a proper assembly.

The molecular basis of AT substrate specificity can be elucidated through **MD simulations**. One example is the molecular docking of (2S)-methylmalonyl-CoA into the active site of a structural model of DEBS AT6 and its subjection to 30 ns MD simulation followed by QM/MM optimization, confirming the role of active-site residues and other relevant

positions involved in the maintaining of a proper chemical environment. This strategy also allowed the identification of a valine in the active site-pocket which structural function was the constriction of the extender unit side chain. The substitution of this valine for an alanine resulted in the acceptance of the non-natural substrate 2-proparglymalonyl-SNAC.[244] MD simulations were key to explain the hydrolytic activity of the pikromycin (Pik) TE because an unproductive accommodation of the epimerized hexaketide in its active site occured.[249] Finally, MD and QM calculations were used to predict substrate scope, site selectivity and stereoselectivity of monooxygenase PikC.[250]

Finally, **directed evolution** of AT domains is another possibility to increase substrate scope of PKSs. With this technique, mutant libraries of the AT LovD were generated through error-prone PCR and saturation mutagenesis. This enzyme naturally transfers α-*S*-methylbutyrate to monacolin J acid (MJA) to yield cholesterol-lowering drug lovastatin, while the directed evolution process engineered an enzyme with improved activity towards α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP) to yield the blockbuster cholesterol-lowering drug simvastatin, delving into this mechanism in next sections of this text.[251,252]

CHAPTER 2:

Background and
State-of-the-Art

2.1. Enzymatic Synthesis of Cholesterol-lowering Drugs

Hypercholesterolemia is a primary risk factor related to coronary artery diseases and atherosclerotic lesions, which are the most-important causes of death in western countries. As endogenous **cholesterol** represents two-thirds of total body cholesterol, targeting *de novo* cholesterol synthesis is an effective way of lowering plasma cholesterol levels. **Statins** are fungal polyketide secondary metabolites that selectively inhibit **hydroxymethyl glutaryl-coenzyme A (HMG-CoA) reductase**, the first enzyme in cholesterol biosynthesis and the rate-limiting step among all the 25 enzymes involved in the cholesterol biosynthetic pathway.
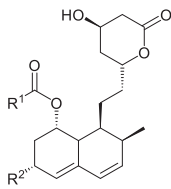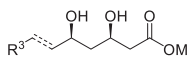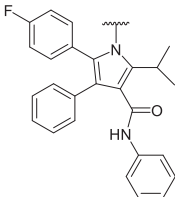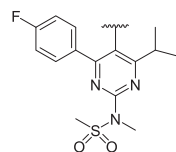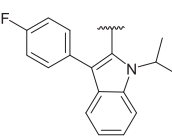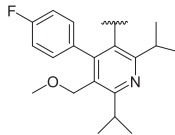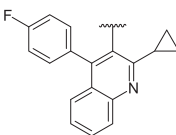
2.1.1. Statins Structure and Mechanism of Action

HMG-CoA reductase contains a "cis-loop" necessary to allocate HMG. All statins share a motif with structural homology to the HMG-CoA intermediate, thus inhibiting HMG-CoA reductase in a competitive way.[253] The inhibition of this enzyme results in HMG-CoA accumulation, which can be metabolized to simpler compounds lacking the sterol ring.

Statins can be divided in two classes based on their structure (**Table 1**):[254]

- **Type I statins** (compactin, pravastatin, lovastatin and simvastatin) are naturally found in fungi and feature a decalin ring and a butyryl group in carbon C8 besides the HMG-like moiety. Their binding to HMG-CoA reductase involves numerous hydrogen bonds similar to those formed with HMG-CoA, as well as interaction of the decalin ring with a helix of the enzyme.

- **Type II statins** (rosuvastatin, atorvastatin, cerivastatin, fluvastatin and pitavastatin) contain a fluorophenyl group and a methylethyl (or cyclopropyl) group that is not present in type I inhibitors, except for the HMG-like moiety. They differ from each other in the central ring structure. These inhibitors exhibit additional binding interactions between their fluorophenyl groups and the HMG-CoA reductase Arg590 residue. Furthermore, rosuvastatin and atorvastatin establish hydrogen bonds with the HMG-CoA Ser565 residue involving a sulfone oxygen atom in rosuvastatin and a carbonyl oxygen atom in atorvastatin.

Table 1. Structure of statins.

Decalin core common to Type I statins:	R^1	R^2	
		H	**Compactin (Mevastatin)**
		OH	**Pravastatin**
		Me	**Lovastatin**
		Me	**Simvastatin**

HMG-CoA-like side-chain common to Type II statins:	R^3

M: H, Na, Ca

Atorvastatin   **Rosuvastatin**   **Fluvastatin**

Cerivastatin   **Pitavastatin**

2.1.2. Biosynthetic and Semi-Synthetic Pathways of Natural Fungal Type I Statins

Lovastatin is a chemical naturally synthesized by *Aspergillus terreus* that has been commercialized as a cholesterol-lowering drug with the name of Mevacor®. The lovastatin gene cluster is composed of 18 genes, six of which are essential for lovastatin biosynthesis: *lovA*, *lovB*, *lovC*, *lovD*, *lovF* and *lovG*[255]. *lovB* and *lovF* encode PK synthases (PKS), whereas *lovA*, *lovC*, *lovD* and *lovG* encode a cytochrome P450 oxygenase, an enoyl reductase (ER), an acyl transferase (AT) and a thioesterase (TE), respectively. LovC activity complements LovB as the ER domain of the PKS is inactivated. These two enzymes catalyze the conversion of nine malonyl-CoA molecules into the dihydromonacoline L precursor in 35 reactions. In the first step, the acetyl group of malonyl-CoA is attached to LovB, which ketosynthase (KS), ketoreductase (KR) and dehydrogenase (DH) domains catalyze subsequent rounds of PK elongation. When the tetraketide is formed (four elongation rounds), the methyl transferase (MT) domain transfers a methyl group from S-adenosyl-L-methionine (SAM) to the growing PK chain. LovC catalyzes enoyl reduction at the tetraketide, pentaketide and heptaketide

stages, while in the hexaketide stage a Diels-Alder cycloaddition occurs to form the fused rings of the decalin system of dihydromonacolin L. After the last elongation, when the nonaketide is formed, the final dihydromonacolin L is then released from LovB by LovG and oxidized to monacolin J by LovA. On the other hand, the lovastatin diketide synthase LovF condensates and reduces two acetyl units and transfers a methyl group from SAM to form the α-methylbutyrate side chain. In the last step, this chain is covalently attached to monacolin J by LovD to yield lovastatin (**Scheme 1**).[256]
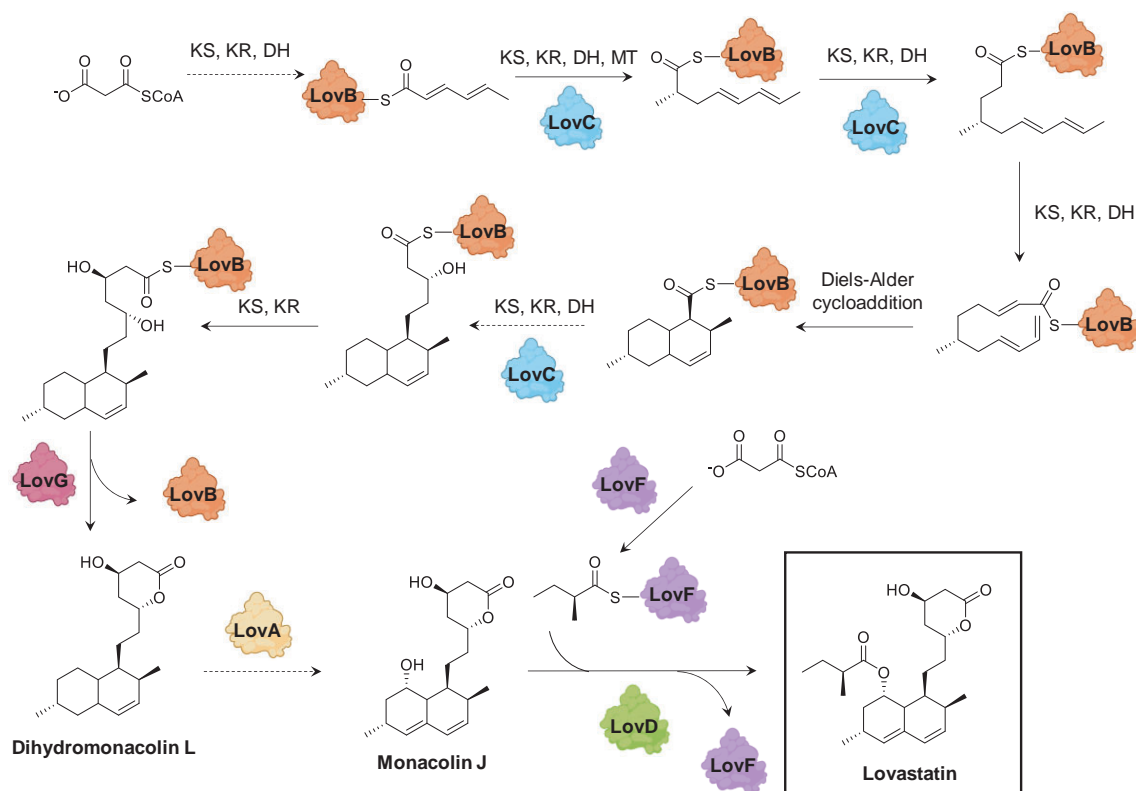
Scheme 1. Biocatalytic synthesis of lovastatin. KS: ketosynthase domain; KR: ketoreductase domain; DH: dehydrogenase domain; MT: methyl transferase domain.

Compactin, also known as **mevastatin**, has a structure very similar to the one of lovastatin, only differing in the lack of the C6 methyl group. The compactin gene cluster contains nine genes, *mlcA-mlcH* and *mlcR*.[255] *mlcA* and *mlcB* encode two PKS analogues to LovB and LovF, respectively. ER MlcG is analogue to LovC as MlcA lacks ER domain. In contrast to LovC, MlcG does not require the transfer of the methyl group of SAM, and the system is able to recognize and accept a desmethyl-dihydromonacolin L intermediate. Finally, MlcC and MlcH are analogues to LovA and LovD. Compactin is the precursor of the more powerful cholesterol-lowering agent **pravastatin**, commercially known as Pravacor®, its 6-β-hydroxy derivative form. This hydroxylation step can be carried out in actinomycetes fermentation processes and its productivity often depends on the production of a cytochrome P450-containing enzyme system[257].

Simvastatin, commercially known as Zocor®, is a semi-synthetic blockbuster analogue of lovastatin which is more effective in treating hypercholesterolemia because of the greater HMG-CoA reductase inhibitory effect due to the substitution of the α-methylbutyrate side chain with α-dimethylbutyrate. This drug can be chemically synthetized through two different routes. Route 1 starts with the hydrolysis of the C8-ester of lovastatin to yield the triol monacolin J, followed by selective silylation of the C13-alcohol, esterification of C8-alcohol with dimethylbutyryl chloride and deprotection of C13-alcohol to finally yield simvastatin.[258] In route 2, lovastatin is treated with *n*-butylamine and *tert*-butylmethylsilyl chloride (TBSCl), the resulting product α-methylated and finally transformed into simvastatin by hydrolysis and lactonization (**Scheme 2**).[259] Both multistep processes are time- and cost-expensive, and turn simvastatin synthesis into a much more laborious procedure than lovastatin production.
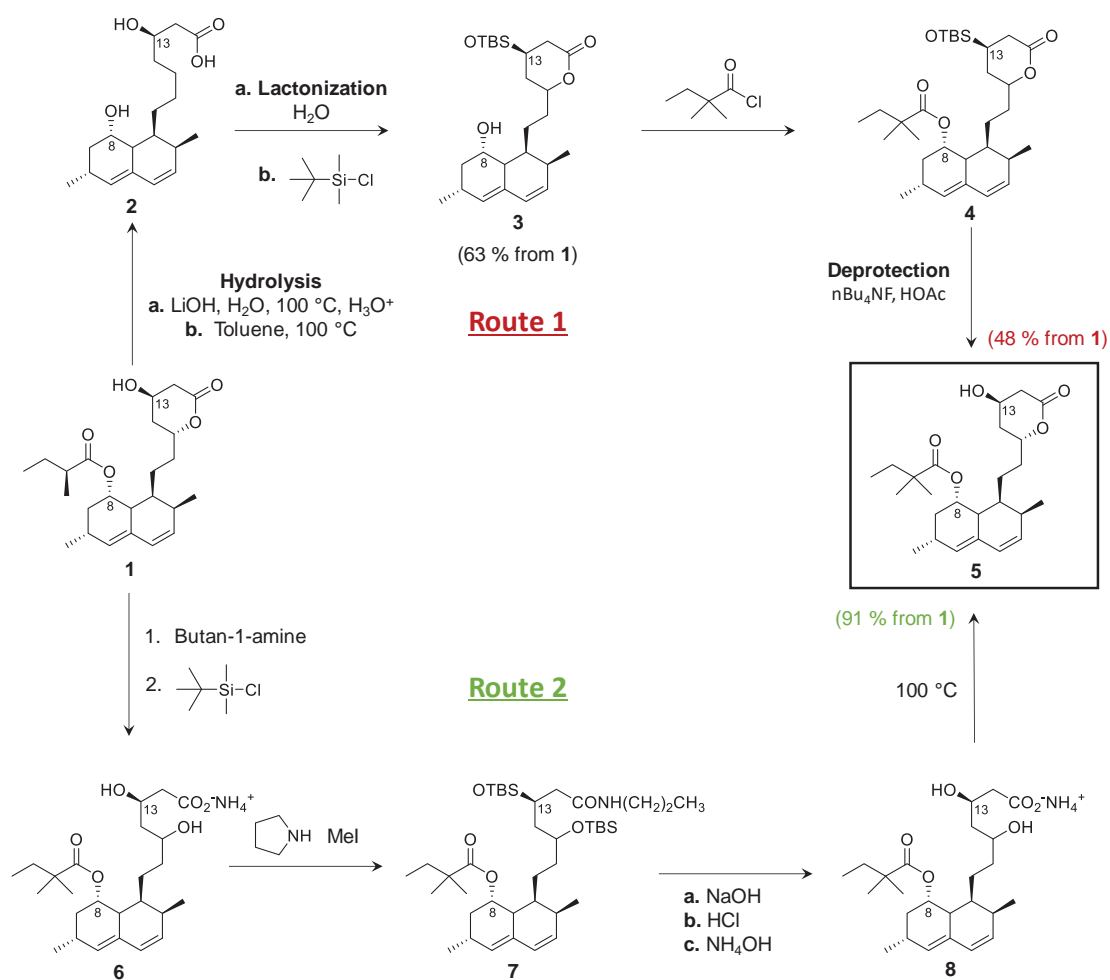
Scheme 2. Chemical transformation of lovastatin to simvastatin.

2.2. Overview of Previous Engineering of Acyltransferase LovD to Improve Simvastatin Synthase Activity

As discussed above, the multi-step chemical synthesis of simvastatin from lovastatin is a harsh process. Enzymatic alternatives concerning lipases and esterases have been explored, but they require a regioselective esterification of the C8-alcohol that requires protection of other alcohol groups often translating into low product yields. Hence, a process to selectively esterify the C8-alcohol of monacolin J acid (MJA) would be highly attractive.

In a first attempt, the Tang group at the University of California Los Angeles (UCLA) studied lovastatin analogues biosynthesis through acyltransferase LovD promiscuity towards different acyl donors (**Table 2**).[260] This was the first *in vitro* trans-acylation assay with LovD and the first-time lovastatin analogues were synthetized without the presence of the acyl-carrier protein LovF. Acyl moiety diversity was increased by substituting the acyl-S-LovF protein partner by synthetic and natural thioesters. In this study, Xie et al. demonstrated a regioselective acylation of monacolin J acid using butyryl-CoA as acyl donor with 87 % conversion. LovD also showed a detrimental hydrolytic capability of lovastatin back to MJA and substrate inhibition at high MJA concentrations. As aforementioned, they assayed the tolerance of LovD toward different acyl donors for many of which the enzyme showed high product yields: commercially available acyl-CoAs, *N*-acetylcysteamine (SNAC) and methyl-thioglycolate (SMTG) thioesters. Moreover, the enzyme also showed to be promiscuous towards acyl acceptors with different decalin cores. Hence, this study demonstrated that LovD can catalyze the synthesis of lovastatin analogues independent from protein-protein interactions with the mega-synthase LovF. The utilization of membrane-permeable thioesters α-dimethylbutyryl-*S*-*N*-acetylcysteamine (DMB-SNAC) and α-dimethylbutyryl-*S*-methylthioglycolate (DMB-SMTG) allowed the efficient and cost-effective synthesis of the blockbuster drug simvastatin.

The poor turnover numbers obtained with DMB-SNAC and DMB-SMTG lead the exploration of other α-dimethylbutyryl donors for the acylation of MJA. In a follow-up study, the same authors identified α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP) as a superior acyl donor, showing a k_{cat} value 30-fold higher than its previous analogues and a K_M value comparable to the one of MJA, hence eliminating substrate inhibition.[261] In the same study, Xie et al. tried to optimize simvastatin synthesis using a batch *in vitro* process with purified LovD, but they were unsuccessful in achieving more than 60 % of MJA conversion due to the competing simvastatin hydrolytic reaction and the readily precipitation of the enzyme. However, they did not observe simvastatin hydrolysis in *E. coli* whole-cell fermentations (probably due to simvastatin extrusion outside the inner membrane) and they were able to achieve around 90% MJA conversion.

Table 2. Acyl thioesters as LovD substrates LovD.

Acyl thioester substrate	Conversion (%)	Acyl thioester substrate	Conversion (%)
	5.1		50
	35		22
	52		10
	87		52
	32		2
	7		33
	89		58
	35		92
	69		17
	99		70

One significant drawback associated to the use of DMB-SMMP in a whole-cell fermentation process was the hydrolysis of the methyl ester group to yield α-dimethylbutyryl-*S*-mercaptopropionic acid (DMB-SMPA). Xie et al. identified *E. coli* carboxyl-esterase BioH as the enzyme responsible for the catalysis of this DMB-SMMP hydrolytic reaction. They demonstrated that using the Δ*bioH* expression strain YT2, the degradation of DMB-SMMP in the whole-cell fermentation process was completely abolished.[262] Another productivity problem associated to LovD was its high insolubility, being completely insoluble when *E. coli* culture was expressed at temperatures over 30 °C. This effect can be due to the presence of nine cysteine residues in the outer surface that can form intra- and inter-molecular disulfide bonds, leading to oligomers and accumulation of insoluble protein aggregates. The strategy adopted to overcome this problem was cysteine substitution by polar residues: the double mutant C40A/C60N exhibited around 50% increase in soluble protein levels without attenuating its catalytic activity.[263]

The most significant achievements in the development of a simvastatin biocatalyst involved directed evolution of LovD, carried out also by Tang group. In a first attempt, Gao et al. performed a directed evolution campaign by generating mutant libraries through saturation mutagenesis and ep-PCR and a screening method that relied on the growth inhibition of *Neurospora crassa* by statins.[251] After seven rounds of evolution, they obtained a variant (G7) with seven amino acid changes (A86V/D12G/A190T/G275S/K26E/H161Y/V334F) that led to a 11-fold increase in simvastatin

synthase productivity compared to that of the wild-type as a result of enhanced catalytic efficiency and thermal stability. Another key achievement of the study was the structural and mechanistic information revealed by the crystallization of LovD variants. This acyltransferase contains the aforementioned α/β hydrolase fold and shows a structure homologous to the one of the cephalosporin esterase EstB from *Burkholderia gladioli*, being the most relevant common features the base Tyr188 and the conserved SXXK patch that contains the active site nucleophile Ser76. The catalysis of this enzyme operates via a ping-pong mechanism that starts with Tyr188 assisting the initial nucleophilic attack of Ser76 hydroxyl group on the α-*S*-methylbutyryl group. Tyr188 also assists activation of C8 hydroxyl group of MJA to attack the acyl-enzyme complex. Lys79 is well-positioned to aid in the activation of both hydroxyl groups by forming hydrogen bonds with Tyr188 and Ser76. The α-*S*-methylbutyryl aliphatic carbons are surrounded by the positively charged side chains of Arg73, Lys79 and Arg173 that might help stabilizing the oxyanion hole (**Figure 1**).
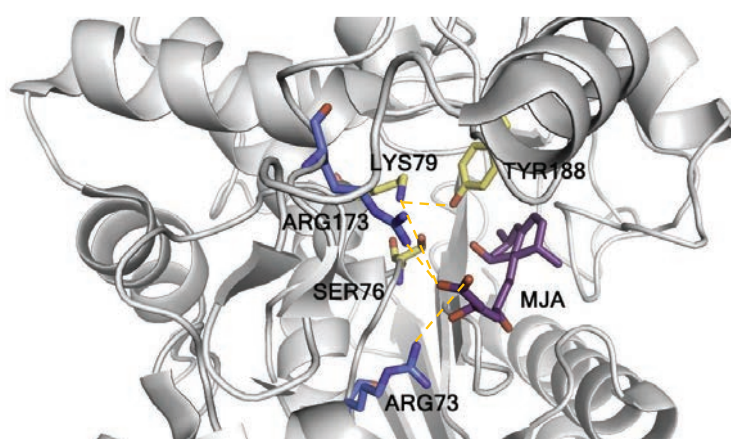
Figure 1. LovD catalytic pocket with monacolin J acid (MJA). Catalytic residues S76, K79 and Y188 are represented in yellow, other stabilizing residues (R73 and R173) are represented in blue and MJA is represented in purple. Hydrogen bonds are represented as dashed yellow lines.

A second directed evolution attempt performed by Codexis Inc. in collaboration with the Tang and Houk groups at UCLA was much more successful in terms of simvastatin synthase productivity. In this study, LovD was subjected to nine rounds of directed evolution using ProSAR-based technologies that gave rise to the super-variant LovD9, with 29 mutations scattered throughout the entire enzyme, which turned to be 1,000-fold more productive than the native enzyme (**Table 3**).[252] Crystallographic structures of the evolved mutants and molecular dynamics (MD) revealed that the active site was gradually more buried, and that the substrate access channel was reduced throughout evolution. These structural changes led to a complete loss of the LovD9 ability to accept the ACP domain of LovF, being the enzyme more prone to accept small surrogates as DMB-SMMP. Despite changes in the binding cleft, the active site remained unaltered in crystal structures. Microsecond MD simulations shed light on the effect of distal mutations in the catalytic triad configuration. In the wild-type enzyme, the catalytic Ser76-Lys79-Tyr188 triad undergoes a rapid transition to an inactive conformation characterized by a very long Lys79-Tyr188

distance. Rounds of evolution progressively maintained the catalytically competent conformation of the triad for longer simulation times. It was also proposed that a group of mutations located in an α-helix close to the active site (N191S, N191G and L192I) could affect the interaction between Asn191 and Tyr188. Another group of mutations in an internal β-sheet (L361M, V370I, A383V and I35L) may increase enzyme stability through improved hydrophobic packing of the core. A third set of late mutations (Q241M, A261V, A261H, N43R, D96R and H404K) could provide thermostability to the protein and reduce its aggregation.

Table 3. Cumulative mutations produced through directed evolution on LovD. Relative productivity was measured as relative turnover numbers in cell lysates referred to the wild-type enzyme. ND: not determined.

Variant (number of mutations)	New mutations	Relative productivity
LovD (0)	None	1
LovD1 (1)	L361M	2
LovD2 (6)	A123P, L174F, A178L, N191S, A247S	12
LovD3 (7)	G275S	36
LovD4 (9)	M157V, S172N	89
LovD5 (15)	A9V, K26E, L192I, R250K, Q297E, A383V	220
LovD6 (20)	N43R, S164G, N191G, Q241M, V370I, H404K	595
LovD7 (23)	S256T, A261V, Q297G, N391S	ND
LovD8 (25)	I4N, R28S, A261H	ND
LovD9 (29)	I35L, D96R, S109C, L335M	1002

Another limitation in simvastatin biosynthesis is the preparation of MJA. Alkaline hydrolysis is still the most common method to convert lovastatin to MJA, but as mentioned before, it is a tedious process. *Penicillium chrysogenum* esterase PcEST can cleave the 2-methylbutyrate side chain of compactin. Due to the structural similarity between compactin and lovastatin, Huang et al. characterized PcEST for the *in vitro* hydrolysis of lovastatin and further introduced it in *A. terreus* to construct a cell factory for the direct production of MJA. With this methodology, most of the biosynthesized lovastatin can be hydrolyzed into MJA by the heterologously expressed lovastatin hydrolase.[264]

Metabolic engineering for simvastatin manufacturing was also used by Bond et al. at the Tang laboratory, through the introduction of six heterologous biosynthetic genes in *S. cerevisiae*.[265] They split the proposed pathway in three different modules: module 1 involved the synthesis of the precursor dihydromonacolin L acid (DMLA) and required the expression of LovB, LovG and LovC; module 2 involved the conversion of DMLA into MJA and required the expression of LovA and its partner cytochrome P450 reductase (CPR); and module 3 required the expression of LovD9 and exogenous addition of DMB-SMMP to afford SVA. Gene knock-out, pH optimization and *in situ* chemical lysis of the cell walls were necessary conditions to finally reach 55 mg L^{-1} SVA and nearly 100 % conversion from MJA. This was the first time SVA was produced using a single heterologous host and that *in situ* lysis was used for production of a semisynthetic natural product.

CHAPTER 3:

Objectives

As summarized in the previous chapter, in the last decade many biotechnological efforts have been directed toward the industrial production of the blockbuster cholesterol-lowering drug simvastatin using a biocatalytic approach. These efforts led to the superior simvastatin synthase LovD9 obtained through directed evolution, which has already been implemented in a whole-cell bioreactor using *S. cerevisiae* as a heterologous host for one-pot simvastatin production. Nevertheless, little is known about the contribution of LovD9 mutations in the proper configuration of the active site to catalysis or other enzyme properties such as thermal stability or protein aggregation. This knowledge could be of great value for further *in silico* design of novel LovD candidates. Moreover, LovD has already proven to accept a small group of acyl donors, but there are still many others that could show even better affinity to the enzyme. Finally, *in vitro* application of an evolved LovD biocatalyst through enzyme immobilization to develop a competent bioreactor remains unexplored. With this background in mind, the main objectives of this work are as follows:

1. Elucidate the contribution of some of the mutations in LovD9 to different catalytic properties of the enzyme through their **rational deconvolution** in three clusters of mutations.

2. Design **new LovD variants** by targeting both the original LovD9 mutated positions and identifying new mutational "hotspots" using David Baker's RosettaDesign coupled with MD simulations and dynamic network analysis.

3. Identify **new LovD acyl donors and acceptors** to produce natural type I statins or unrelated molecules.

4. Immobilize **a promising LovD candidate** in solid porous materials **for continuous *in vitro* production of simvastatin**.

CHAPTER 4:

Deconvoluting
the directed evolution pathway
of engineered acyltransferase LovD

4.1. Introduction

As described in **Chapter 2**, LovD9 catalyzes simvastatin acid (SVA) biosynthesis through a ping-pong acylation mechanism where **MJA acts as a competitive inhibitor** of the acyl-enzyme complex formation.[260] Furthermore, LovD also presents a **side thioesterase activity** (described for the first time in this work) under aqueous conditions that competes with the acyltransferase activity, hydrolyzing the acyl-enzyme complex and yielding 2,2-dimethylbutanoic acid (DMB) as a dead-end pathway. Such thioesterase activity precludes the acylation of MJA thus hampering the production of SVA. This thioesterase activity has not been evaluated before for either the wild-type LovD or its engineered variants. Finally, LovD also presents an **esterase activity** able to hydrolyze SVA back to MJA and the corresponding acid (DMB), diminishing the final yield of the target product (**Scheme 1**).

1: Substrate inhibition
2: Thioesterase activity
3: Acyltransferase activity
4: Esterase activity
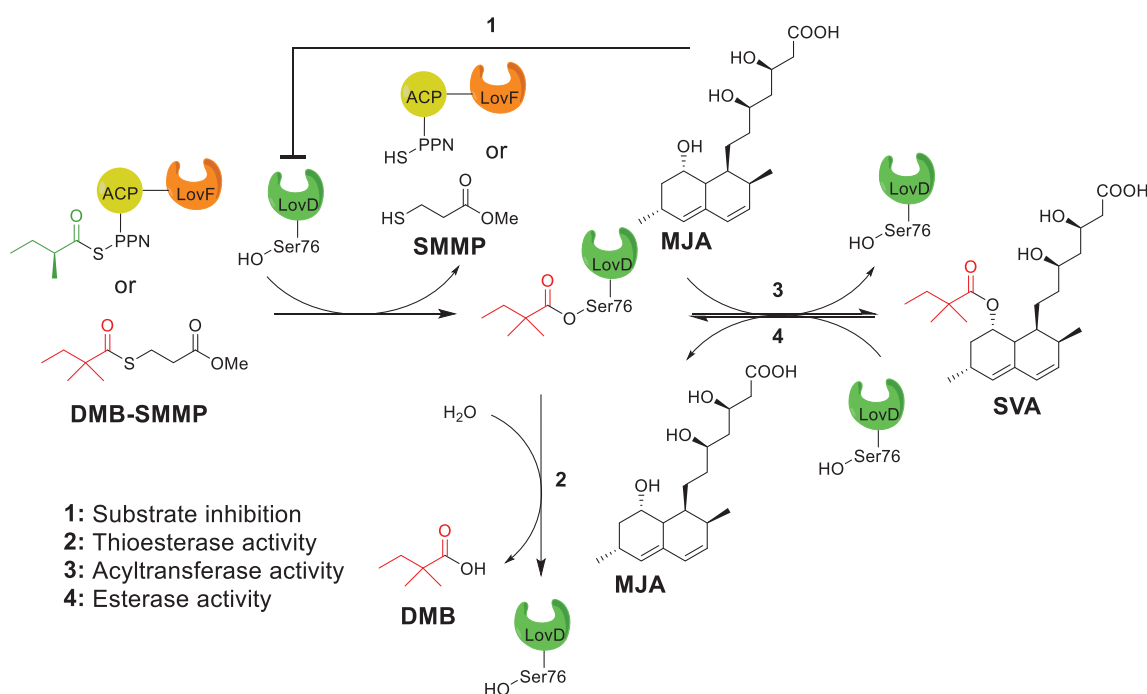
Scheme 1. Catalytic reactions involved in simvastatin acid (SVA) synthesis catalyzed by LovD. DMB-SMMP: α-dimethylbutyryl-*S*-methyl-mercaptopropanoate. SMMP: *S*-methyl-mercaptopropionic acid. MJA: Monacolin J acid. SVA: Simvastatin acid. DMB: 2,2-dimethylbutanoic acid. ACP: acyl carrier protein. PPN: 4'-phosphopantetheine.

This catalytic mechanism resembles the kinetically controlled synthesis of esters and amides catalyzed by esterases and amidases,[266,267] in which there are two main side unwanted reactions affecting product yield, i.e. the hydrolysis of the activated acyl-donor and the hydrolysis of the final product. This mechanistic perspective encouraged us to study how different mutations in LovD9 contribute to relevant features such as substrate inhibition, side hydrolytic activities and thermal stability which define the operational productivity of the engineered biocatalyst. As previously reported,[268] **partial deconvolution**

can be used to access important information regarding additive, cooperative or antagonistic mutational effects. In this work, we have generated three new variants, dividing the 29 mutations of LovD9 into three main clusters: residues located at the substrate entrance channel (LovD-Ch), at buried positions near the active site cavity (LovD-Bu), and a combination of both (LovD-BuCh1,2). We have performed kinetic studies on these three variants and quantified their thioesterase and esterase activities, comparing them to those measured for highly active variants LovD7 and LovD9. These data have provided a sequence landscape where we can weigh the contribution of the specific amino acid clusters to the enzyme improvement during the artificial evolution path.

4.2. Experimental Section

4.2.1. Materials

All LovD genes were synthetized and cloned into expression vector pET28b by GeneScript Gene Synthesis service (Piscataway, NJ, USA). The sequences are provided in **Table 1**. The genotype of the bacteria strains used for molecular biology and expression purposes are described in **Table 2**. Substrates MJA and DMB-SMMP were synthetized and donated by Tang lab (UCLA, USA). Simvastatin hydroxyl acid ammonium salt 98% was purchased from Toronto Research Chemicals (Toronto, Canada). 2,2'-Dithiodipyridine 100% and Amicon Ultra-0.5 centrifugal filter units (10 kDa) were purchased from Sigma-Aldrich (St. Louis, IL, USA). Agarose microbeads with cobalt chelates (AG-Co^{2+}) (50-150 μm diameter) were purchased from Agarose Bead Technologies (Madrid, Spain). Polypropylene (12 x 32 mm, 300 μL volume) vials were purchased from Waters (Milford, Massachusetts, USA). MicroWell 96-well microplates were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA).

Table 1. Protein sequences of LovD variants.

Variant	Sequence (Inserted mutation are highlighted in bold orange)
LovD	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD1 (1)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD2 (6)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSR**F**APP**L**VNDPGAEWIYGA**S**LDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNS**S**DGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD3 (7)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS

	GLSYVFLHPLLREYMAQGHLQSAEKFGIQSR**F**APP**L**VNDPGAEWIYGA**S**LDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNS**S**DGRLRYDDSVYFRADGEECFGGQGVFSS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD4 (9)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQSAEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**S**LDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNS**S**DGRLRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD5 (15)	MGSIIDAA**V**AADPVVLMETAFRKAV**E**SRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQSAEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**SI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNS**S**DG**K**LRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**E**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNIVWQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD6 (20)	MGSIIDAA**V**AADPVVLMETAFRKAV**E**SRQIPGAVIMARDASG**R**LNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQ**G**AEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**GI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**M**THRNS**S**DG**K**LRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**E**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**I**WQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFE**K**AIYAQYQQG
LovD7 (23)	MGSIIDAA**V**AADPVVLMETAFRKAV**E**SRQIPGAVIMARDASG**R**LNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQ**G**AEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**GI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**M**THRNS**S**DG**K**LRYDD**T**VYFR**V**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**G**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**I**WQIDPKAGLCTL**V**FFQLEPW**S**DPVCRDLTRTFE**K**AIYAQYQQG
LovD8 (25)	MGS**N**IDAA**V**AADPVVLMETAFRKAV**E**S**S**QIPGAVIMARDASG**R**LNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQ**G**AEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**GI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**M**THRNS**S**DG**K**LRYDD**T**VYFR**H**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**G**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**I**WQIDPKAGLCTL**V**FFQLEPW**S**DPVCRDLTRTFE**K**AIYAQYQQG
LovD9 (29)	MGS**N**IDAA**V**AADPVVLMETAFRKAV**E**S**S**QIPGAV**L**MARDASG**R**LNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLV**R**LDETVDRLLPDL**C**AMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQ**G**AEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**GI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**M**THRNS**S**DG**K**LRYDD**T**VYFR**H**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**G**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMV**M**RRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**I**WQIDPKAGLCTL**V**FFQLEPW**S**DPVCRDLTRTFE**K**AIYAQYQQG
LovD-Ch (6)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQSAEKFGIQ**N**R**F**APPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**M**THRNSADGRLRYDD**T**VYFR**V**DGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-Bu (8)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**L**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**L**VNDPGAEWIYGA**GI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**I**WQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-BuCh1 (14)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**L**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQSAEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**GI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**M**THRNSADGRLRYDD**T**VYFR**V**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**I**WQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-BuCh2 (14)	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**L**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**V**AQGHLQSAEKFGIQ**N**R**F**APP**L**VNDPGAEWIYGA**GI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**M**THRNSADGRLRYDD**T**VYFR**V**DGEECFGGQGVFS**S**PGSYMKVLH

	SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW
	RRKGS**M**TFGGGPNI**I**WQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG

Table 2. Strains, plasmids and oligonucleotides used in this study.

Plasmids	Relevant genetic characteristics
pET28b *lovD*	*lovD* gene (1251 bp) from *A. terreus* cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD1*	*lovD1* gene (1251 bp) from *A. terreus* cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD2*	*lovD2* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD3*	*lovD3* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD4*	*lovD4* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD5*	*lovD5* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD6*	*lovD6* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD7*	*lovD7* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD8*	*lovD8* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD9*	*lovD9* gene (1251 bp) from directed evolution design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-Ch*	*lovD-Ch* (1251 bp) from rational design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-Bu*	*lovD-Bu* gene (1251 bp) from rational design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BuCh1*	*lovD-BuCh1* gene (1251 bp) from rational design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BuCh2*	pET 28b *lovD-BuCh1* gene as template (QuickChange site directed mutagenesis)
Strains	Relevant phenotype
E. coli DH5α	F⁻ φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r$_K^-$, m$_K^+$) *pho*A *sup*E44 λ⁻ *thi*-1 *gyr*A96 *rel*A1
E.coli BL21	F⁻ *omp*T *hsd*S$_B$ (r$_B^-$, m$_B^-$) *gal dcm ara*B::T7RNAP-*tet*A

Oligonucleotides	Sequences
V261H	GTT(V) → CAT(H) Forward: 5' CCGTGTATTTTCGTCATGATGGCGAGGAATGCTTTGG 3' Reverse: 5' CCAAAGCATTCCTCGCCATCATGACGAAAATACACGG 3'

4.2.2. QuickChange site-directed mutagenesis

The *lovD* gene mutation A261H was introduced with the QuickChange site-directed mutagenesis strategies using the plasmid pET28b *lovD-BuCh1* A261V as template and performed according to the protocol described in **Table 3**. Chemically competent *E. coli* DH5α strain cells were transformed with DNA preparation. The plasmid DNA was isolated with the NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel) and the expected mutations were confirmed by DNA sequencing by STAB vida (Caparica, Portugal).

Table 3. QuickChange site-directed mutagenesis protocol.

Components	V (µL)	Final concentration	PCR cycles		
Water free nuclease	41.5		Step	Temperature (°C)	Time (min)
10X Buffer reaction prof. DNA polymerase (NZY)	5		Initial denaturation	95	5
dNTP mix	1	0.2 mM each		95	1
Forward primer	0.5	0.5 ng µL^{-1}	30 PCR cycles	55	0.5
Reverse primer	0.5	0.5 ng µL^{-1}		72	6.5
Template DNA	1	2 ng µL^{-1}		72	10
DNA polymerase (NZY)	0.5	0.02 U µL-1	Final extension	4	hold

4.2.3. Expression of LovD variants

The plasmids encoding the His-tagged LovD variants were transformed into chemically competent *E. coli* strains DH5α and BL21 through heat shock[269] for plasmid propagation and recombinant protein expression, respectively. Single colonies of *E. coli* containing a plasmid encoding a variant of LovD were inoculated into 3 mL of Lysogeny broth (LB) medium containing 30 µg/mL of kanamycin. Cells were grown overnight at 37 °C with shaking at 250 rpm. The culture was diluted 1:50 into 50 mL of LB medium containing 30 µg/mL of kanamycin and the culture was grown until OD$_{600nm}$ reached 0.6-0.8. At that point, the protein expression was induced by adding isopropyl-D-thiogalactoside (IPTG) to a final concentration of 0.1 mM and the culture was incubated overnight at 21 °C with shaking at 250 rpm. Cells were collected by centrifugation (2057 xg for 30 min at 4 °C) and cell pellet was resuspended with 6 mL of 50 mM sodium phosphate buffer (pH 8.0). Cells were lysed by sonication on ice and cell debris was removed by centrifugation (12857 xg for 30 min).

4.2.4. Purification of LovD variants

All the LovD His-tagged proteins were purified through immobilized metal affinity chromatography (IMAC) using agarose-based resin functionalized with cobalt chelates (AG-Co^{2+}) in bulk. First, the 1 mL of cell lysate was incubated with 100 mg of resin in purification buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 10 mM MgCl$_2$ at pH 8) for 1 h at 4 °C, then the bound enzymes were washed with the same purification buffer and eluted with 200 mM imidazole in 50 mM HEPES 10 mM MgCl$_2$ at pH 8.0. Protein concentrations were qualitatively assessed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (**Figure 1**) and quantitatively determined by Bradford protein assay using bovine serum albumin as a standard.[270]
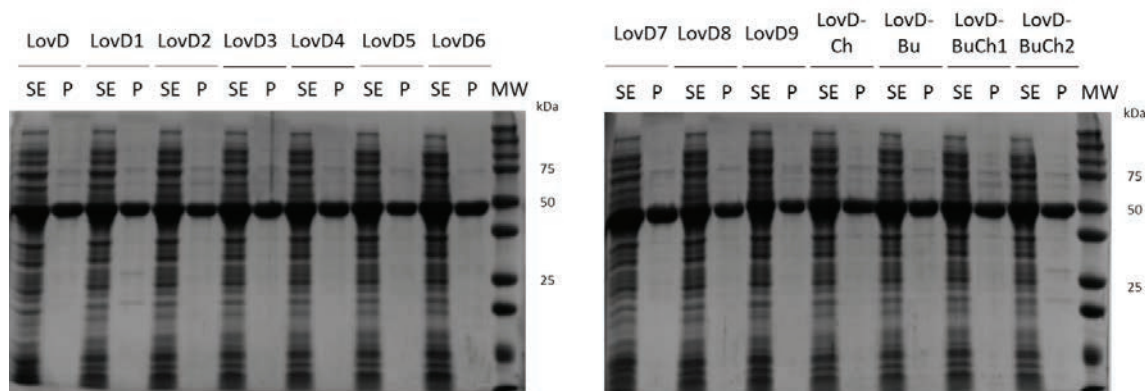
Figure 1. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for protein expression and purification of different LovD variants. SE: Soluble Extract. P: Pure enzyme. MW: Standard Molecular Weight Marker.

4.2.5. Time-course of enzyme catalyzed reactions with LovD variants

Time-course assays of SVA production for different LovD variants, were performed at 5 mM MJA, 2 mM DMB-SMMP in 50 mM HEPES, 10 mM $MgCl_2$ (pH 8.0) and 10% dimethyl sulfoxide (DMSO) to facilitate solubilization of DMB-SMMP. Reactions were triggered by adding 1 µM of each LovD variant, and samples were withdrawn at different time points by passing them through tangential ultrafiltration unit (Amicon Ultra centrifugal filters, 10 kDa). Samples were analyzed by Ultra Performance Liquid Chromatography (UPLC) (Waters 2690) equipped with a Photodiod array (PDA) detector using a ACQUITY UPLC® BEH C18 1.7 µm (2.1 x 50 mm) Waters column coupled to a LCT XE time-of-flight (TOF) mass spectrometry detector with electrospray ionization source (ESI). Analytes were eluted with an isocratic mobile phase composed of 52 % (v/v) of acetonitrile in water (0.1 % (v/v) formic acid) during 15 min at flow rate of 0.3 mL/min (**Figure 2**). The source parameters of the mass spectrometer were: capillary voltage 1000 V, cone voltage 50 V, cone gas 50 L/h, desolvation gas 600 L/h, mass range 100-1000 m/z. Source temperature was set at 120 °C.
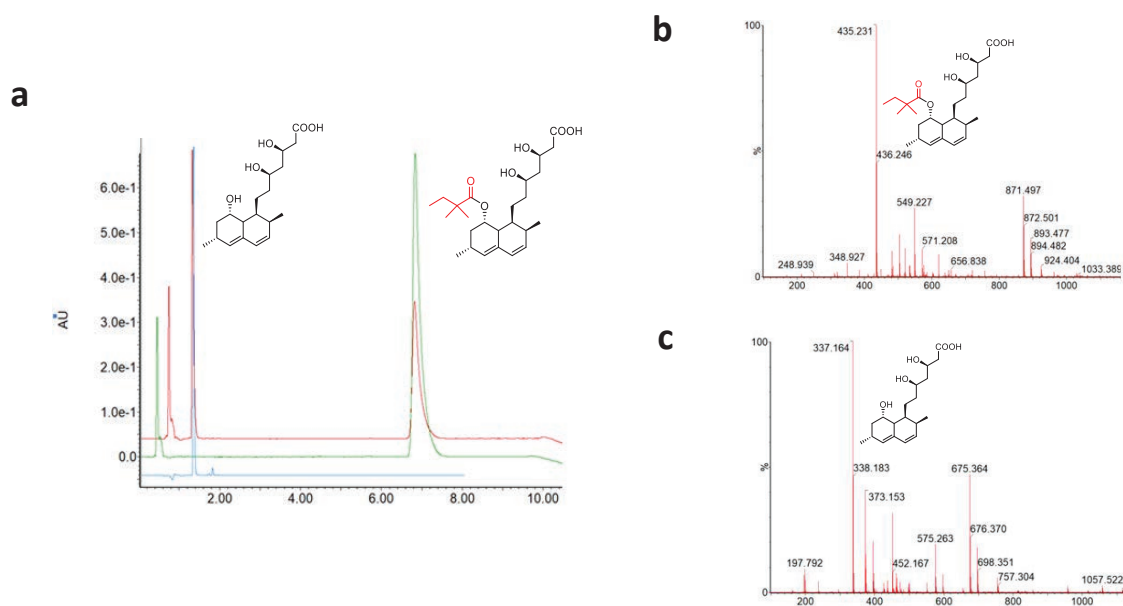
Figure 2. (a) Ultra performance liquid chromatography (UPLC) spectra after 24 h reaction time with 1 µM of enzyme, 3 mM monacolin J acid (MJA) and 2 mM α-dimethylbutyryl-*S*-methyl-mercaptopropionate (DMB-SMMP) (in red), a pure sample of 2 mM simvastatin acid (SVA) (in green) and a pure sample of 3 mM MJA (in blue). The peak at 1.8 min is assigned to MJA and the peak at 7.0 minutes is assigned to SVA. **(b,c)** Mass spectrometry (MS) spectra after 24 h reaction time with 1 µM of enzyme, 3 mM MJA and 2 mM DMB-SMMP (negative-ion mode). The peak of 436.246 in **(b)** corresponds to SVA mass and the peak of 337.164 in **(c)** corresponds to MJA mass.

4.2.6. Kinetic characterization of LovD variants

To determine the kinetic parameters of the different LovD variants, the concentration of MJA was varied from 0 to 10 mM, whereas the DMB-SMMP concentration was fixed at 2 mM. The assays were performed at the same conditions as described in the previous section. Reactions catalyzed by LovD7, LovD9, LovD-BuCh1 and LovD-BuCh2 were triggered with 1 µM of enzyme and stopped after 30 min, 1 h and 2 h, respectively. Reactions catalyzed by LovD-Bu and LovD-Ch were triggered with 2 µM and 4 µM enzyme, respectively, and stopped after 1 h and 2 h for LovD-Bu and after 1 h, 16 h and 18 h for LovD-Ch. Data were plot and fitted to the Michaelis-Menten (**Equation 1**) and substrate inhibition models (**Equation 2**):

$$v = \frac{v_{max}\,[S]}{K_M+[S]} \quad \textbf{(1)}$$

$$v = \frac{v_{max}\,[S]}{K_M+[S]\left(1+\frac{[S]}{K_i}\right)} \quad \textbf{(2)}$$

in which v_{max} is the maximum velocity, K_M is the Michaelis constant and K_i is the inhibition constant.

77

4.2.7. Thioesterase spectrophotometric assay

Reactions to characterize thioesterase activity were performed in 96-well plates for 30 min at 37 °C using 3 µM enzyme, 0 or 3 mM MJA, 2 mM DMB-SMMP and 2 mM 2,2'-dithiodipyridine (2-DTDP) in 50 mM HEPES 10 mM MgCl$_2$ (pH 8.0) 10 % DMSO. The thiol formed after the enzymatically driven thiolysis of DMB-SMMP spontaneously reacts with 2-DTDP, generating 2-thiopyridone that absorbs at 323 nm with an absorption coefficient of 7600 M^{-1} cm^{-1} (**Scheme 2**). [271]
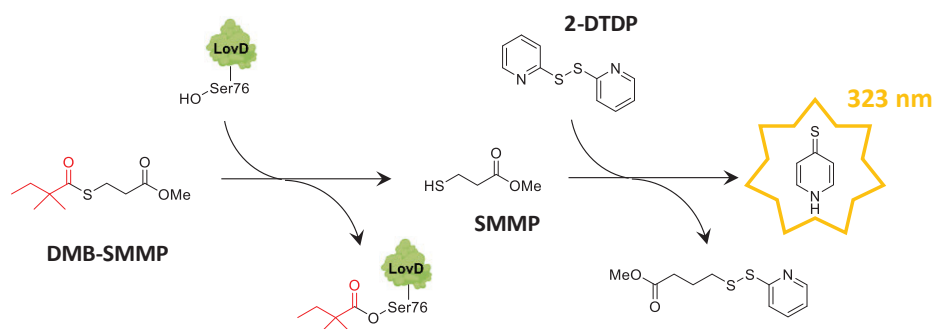
Scheme 2. Spectrophotometric assay for measuring LovD thioesterase activity. DMB-SMMP: α-dimethylbutyryl-*S*-methyl-mercaptopropanoate. SMMP: *S*-methyl-mercaptopropionic acid. 2-DTDP: 2,2'-dithiodipyridine.

4.2.8. Esterase activity characterization

The ability of the LovD variants to hydrolyze SVA yielding MJA was monitored by UPLC as described above. The reactions were triggered by adding 2 µM enzyme to a reaction mixture of 2 mM SVA in 50 mM HEPES, 10 mM MgCl$_2$ (pH 8.0), 10 % DMSO and stopped after 24 h passing through tangential ultrafiltration (Amicon Ultra 0.5 mL centrifugal filters, 10 kDa) and quantified with the same UPLC method described previously.

4.2.9. Thermal shift assay

To determine the melting temperature (T$_m$) of LovD variants, 4 µg of the protein was incubated with a 50X stock solution of SYPRO Orange Dye in a final volume of 25 µL in a 96-qPCR well plate with a ramp from 25 °C to 95 °C in increments of 0.5 °C for 10 s.[272]

4.2.10. Thermal inactivation assays

To determine the thermostability of some LovD variants, 3 µM of each variant was diluted in 50 mM HEPES 10 mM MgCl$_2$ (pH 8.0) and further incubated at 37 °C. The residual activity was measured at different incubation times using the spectrophotometric assay previously described. Data were plot and fitted to a first order enzyme inactivation equation:

$$C_t = C_0\, e^{-k\,t} \quad \text{(3)}$$

where C_0 is the initial concentration and k is the inactivation constant.

4.3. Computational Section

4.3.1. Molecular Dynamics simulations

Starting structures of the simulations were generated from the S5 mutant of simvastatin synthase from *Aspergillus terreus* complex with monacolin J acid (MJA) (PDB 3HLD; resolution: 2.00 Å).[251] Mutations were introduced manually using PyMol.[273] Microsecond Molecular Dynamics (μs-MD) simulations were carried out with AMBER 20 package implemented with *ff14SB*[274] and general Amber (GAFF2)[275] force fields for the proteins and ligands, respectively. Ligand parameters were generated with the antechamber module of AMBER[276], using GAFF2 force field and with partial charges set to fit the electrostatic potential generated with HF/6-31G(d) using the RESP method.[277] The charges were calculated according to the Merz-Singh-Kollman scheme using Gaussian 16.[278] Protein complexes were immersed in a water box with a 10 Å buffer of TIP3P[279] water molecules and neutralized by adding explicit Na^+ or Cl^- counterions. A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules and ions, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 10 kcal/mol were applied to the solute, and the Andersen temperature coupling scheme[280] was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Water molecules were treated with the SHAKE algorithm[281] such that the angle between the hydrogen atoms is kept fixed through the simulations. Long-range electrostatic effects were modelled using the particle mesh Ewald method.[282] An 8 Å cut-off was applied to Lennard-Jones interactions. Each system was equilibrated for 2 ns with a 2 fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for additional 1.0 μs under the same simulation conditions. Tunnels and channels in the protein structures were analyzed using Caver 3 software[283,284] and visualized using PyMol. A total of 100 frames were extracted and analyzed from the whole 1 μs trajectories for each system. The starting point for the tunnel search was set at the catalytic triad (Ser76-Lys79-Tyr188) center of mass. The probe radius and the tunnel clustering threshold were set to 0.7 and 2.0 Å, respectively. Default values were used for the rest of parameters.

4.3.2. Protein structure prediction and scoring

All simulations were carried out with Rosetta 3.[103] First, a full-length model was constructed from the partially resolved crystallographic structure of the *apo* LovD9 mutant of simvastatin synthase from *Aspergillus terreus* (PDB 4LCM; resolution: 3.19 Å). The missing loop between residues 257-265 was transplanted from the crystallographic structure of the *apo* LovD6 mutant of the same enzyme (PDB 4LCL; resolution: 1.80 Å).[252] The missing flexible N-terminus (residues 1-11) was transplanted from the crystallographic structure of the *apo*

S5 mutant of the same enzyme (PDB 3HLC; resolution: 2.00 Å).[251] Fragments were assembled and steric clashes were removed by performing a highly restricted geometry minimization of the protein with AMBER 20 (see **Molecular Dynamics simulations** section), in which only the connecting residues were allowed to move. The resulting model was backbone and sidechain pre-minimized using the Rosetta *minimize_with_cst* application, which uses harmonic distance constraints on all C-alpha atoms within 9 Å in the input structure. Then, mutations corresponding to the directed evolution and cluster variants were introduced into the pre-minimized scaffold with the Rosetta *ddg_monomer* application[252] using the high-resolution protocol which allows a small degree of backbone conformational freedom. The lowest-Rosetta-energy structure resulting after 500 iterations runs for each variant were selected and used as starting geometries for the Rosetta *relax* application.[285,286] By interlacing 15 cycles of sidechain packing with gradient based minimization of torsional degrees of freedom, *relax* searches the local conformational space around the starting structure (large-amplitude conformational changes are not accessible to this method). Up to 2000 decoys were generated for each LovD variant and evaluated using the *ref2015* score function using the *score_jd2* application.[287,288] The root-mean-square deviation (rmsd in Å) of decoys' C-alpha atomic positions with respect to those of the input structures were calculated.

4.4. Results and Discussion

4.4.1. Rational design and functional characterization of new cluster variants of LovD

Depending on the distribution of the 29 mutations across the LovD9 3D structure we arbitrarily classified all mutations in three groups; one group including eight buried mutations near the active site (I35L, A178L, N191G, L192I, G275S, L361M, V370I and A383V), a second group including six mutations at the substrate entrance channel (M157V, S172N, L174F, Q241M, S256T and A261H) and a third group encompassing the remaining mutations located at the protein surface (I4N, A9V, K26E, R28S, N43R, D96R, S109C, A123P, S164G, A247S, R250K, Q297G, L335M, N391S and H404K). During the evolution path, the first four rounds increased the acyltransferase activity 89 times by mostly introducing buried and channel mutations (5 out of 7 total mutations in LovD4), whereas the last five rounds increased the enzyme activity 11 times through enriching the protein mainly with surface mutations (13 out of 22 total mutations from LovD4 to LovD9).[252] According to these previous results, we hypothesized that both buried and substrate entrance channel mutation clusters are the major contributors to the activity enhancement during the *in vitro* evolution.

To demonstrate this hypothesis, we have generated three new variants containing each mutation cluster: buried cluster variant (LovD-Bu), channel cluster variant (LovD-Ch), as well as a combined variant that contains both buried and channel clusters (LovD-BuCh1). LovD-Bu, LovD-Ch and LovD-BuCh1 contain 27%, 21% and 48% of the 29 mutations introduced in LovD9, respectively **(Figure 3.a, Table 1)**. To understand the specific contribution of such clusters to the enhancement of the catalytic efficiency along the evolution path, we extracted the initial rate and the product yields from the reaction time-course analyses **(Figure 3.b-f)** performed with pure LovD1-9, LovD-Bu, LovD-Ch and LovD-

BuCh1 variants. The reaction time-courses provide an accurate picture of the initial synthetic rates (v_0) of the kinetically controlled acylation process and product yield at 24 h. We identified the specific contributions of each cluster to the enzyme productivity by analyzing these two parameters under the same reaction conditions (**Figure 3.b**).
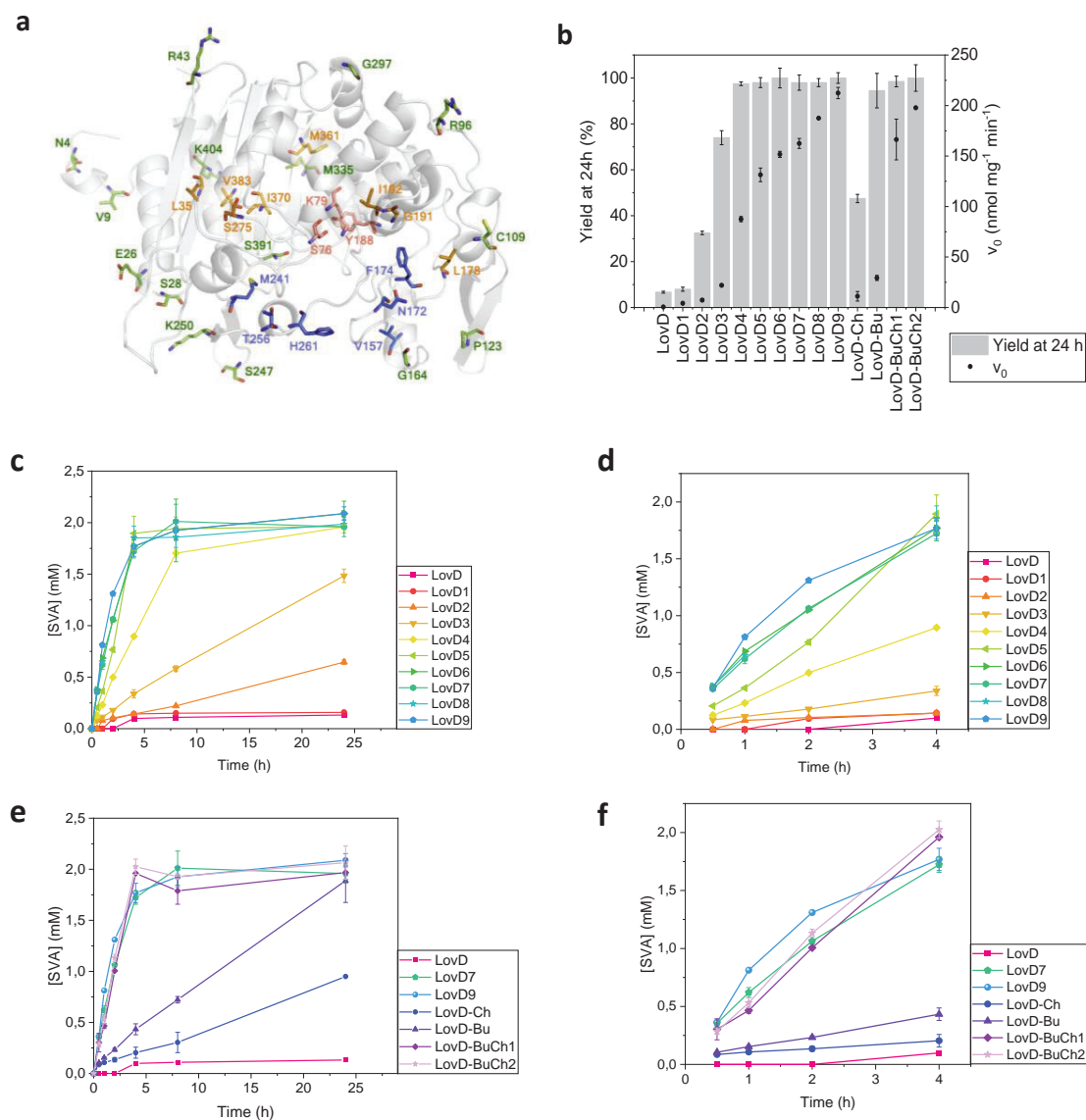
Figure 3. (a) Arbitrarily identified clusters in LovD9. Buried mutations are colored in gold, channel mutations are colored in blue, surface mutations are colored in green and catalytic triad residues are colored in pink. **(b)** Initial velocities (v_0, in circles) and yield after 24 h (in bars) for simvastatin acid (SVA) production catalyzed by LovD evolved variants[252] and the new cluster variants produced in this work. Reaction conditions: 5 mM monacolin J acid (MJA), 2mM α-dimethylbutyryl-*S*-methyl-mercaptopropionate (DMB-SMMP) in 50 mM HEPES, 10 mM MgCl$_2$ (pH 8.0) and 10% dimethyl sulfoxide (DMSO) (see **Experimental Section**). **(c-f)** Time courses of enzyme reactions with LovD directed evolution variants from 0 to 24 h **(c)** and from 0 to 4 h of reaction **(d)** and LovD cluster-designed variants compared with LovD wild-type, LovD7 and LovD9 from 0 to 24 h **(e)** and from 0 to 4 h **(f)**. Initial reaction rate (v_0) for each variant was calculated using reaction rates in **(d)** and **(f)**.

As expected, the v_0 of the reaction increases along the evolution rounds. The largest increase in v_0 occurs within the first rounds: while wild-type LovD exhibits a $v_0 = 0.56$ nmol mg $^{-1}$ min^{-1}, LovD4 shows a $v_0 = 87.5$ nmol mg $^{-1}$ min^{-1}. This means that only 9 mutations were sufficient to enhance v_0 around 150 times. From this intermediate variant to the most evolved one (LovD9), 20 more mutations increased the initial rate just 2.4 times (LovD9, $v_0 = 212$ nmol mg $^{-1}$ min^{-1}). These data follow the same trend described in the original directed evolution study[252]. The key role of those 9 initial mutations in LovD productivity is also supported by the fact that all evolved variants from LovD4 reached 100% simvastatin yield in 24 hours. During the first 4 rounds of mutation, half of the so-called substrate entrance channel and buried mutations were introduced, reflecting the importance of these positions at improving both the formation of the acyl-enzyme complex with the acyl donor (DMB-SMMP) and the subsequent acyl transfer to MJA.

Noteworthy, the LovD-Bu variant (8 mutations) reaches a 95% product yield in 24 h with $v_0 = 29.3$ nmol mg $^{-1}$ min^{-1}, an activity that falls between those of LovD3 (7 mutations) and LovD4 (9 mutations) which share only 4 buried mutations (L361M, A178L, N191S and G275S) with LovD-Bu (Table 1). Therefore, buried mutations clearly contribute to increase the synthetic capacity towards the formation of SVA starting from artificial acyl donors, but these mutations by themselves are not enough to pair the catalytic activity observed for the most evolved variants. On the contrary, the LovD-Ch variant containing the six substrate entrance channel mutations found in LovD9 only reaches an activity similar to that of LovD2, revealing that these mutations contribute to the total LovD9 activity to a lower extent than the buried ones when taken separately.

When we combined buried and substrate entrance channel mutations, the resulting variant (LovD-BuCh1) reaches 100% yield in 4 h with v_0 similar to that measured for LovD7, but with 9 fewer mutations. Merging both engineered clusters gave rise to one variant 7 and 18 times more active than LovD-Bu and LovD-Ch, respectively. This finding evidences a cumulative effect between these two mutation clusters that enhances the enzyme performance towards SVA biosynthesis. Starting from the LovD-BuCh1 variant, we created a new variant with an additional single mutation involving position 261. We hypothesized that this position plays an important role in activity, since it was mutated to Val in the 7th round and to His in the 8th round. To our delight, the simple replacement of A261V by A261H mutation increased v_0 by 20%. This single amino acid substitution achieved roughly the same activity enhancement as the one obtained during the evolution from LovD7 to LovD8 that required 3 new mutations.

Hence, the cumulative clustering of entrance channel and buried mutations including the A261H replacement gave rise to a new variant (LovD-BuCh2) as efficient as LovD8 (with 11 additional mutations mainly located at external positions).

4.4.2. Kinetic characterization of LovD variants

As reported in **Chapter 2**,[252,261] wild-type LovD performs the acyl transfer reaction through a ping pong bi-bi substituted-enzyme mechanism where the acyl acceptor (MJA) acts as a competitive inhibitor. Contextualizing the industrial production of SVA, substrate inhibition jeopardizes the process productivity under the high substrate concentrations often demanded by large-scale manufacturing batch processes. In this work, we provide additional information on this key feature.

Although MJA negligibly inhibits wild-type LovD when using DMB-SMMP as acyl donor, we studied substrate inhibition for two of the most productive LovD engineered variants (LovD7 and LovD9) together with our new cluster variants, complementing the kinetic parameters determined in the original evolution study.[252] To that aim, we performed steady-state kinetics for the synthesis of SVA under fixed DMB-SMMP concentration (2 mM) and varying the concentration of the acyl acceptor (MJA). When we inspected the Michaelis-Menten (MM) curves we observed that wild-type LovD and LovD-Ch did not presented inhibition at high MJA concentrations, while a strong substrate inhibition for LovD-Bu and some of the most evolved variants at MJA concentrations higher than 3 mM was detected **(Figure 4)**. To assess the impact of substrate inhibition on the catalytic efficiency of each variant, we fitted the MM plots to the classical kinetic MM model[24] for LovD and LovD-Ch **(Equation 1, see Experimental Section)** and to the classical kinetic substrate inhibition model[289] for LovD-Bu, LovD-BuCh1, LovD-BuCh2, LovD7 and LovD9 **(Equation 2, see Experimental Section)**.

In **Table 4,** the LovD-Ch variant shows a k_{cat}/K_M of 2.9 M^{-1} s^{-1}, which improves around 3 times the catalytic efficiency of wild-type LovD, mostly due to a 4-times lowering of K_M. LovD-Bu variant resulted in a k_{cat}/K_M 19 times higher than wild-type, but the K_i of 1.83 mM suggests that substrate inhibition emerges when internal mutations near the active site are introduced. When the two mutation clusters are combined into LovD-BuCh1, k_{cat}/K_M increases 10 and 59 times with respect to LovD-Bu and LovD-Ch variants, respectively, and K_i is higher than in LovD-Bu. These kinetic data suggest that entrance channel mutations mitigate the inhibition capability fostered by buried mutations, increasing the overall catalytic efficiency of the enzyme. k_{cat}/K_M and K_i of LovD-BuCh2 are very similar to those measured for LovD7, in agreement with the initial rate values obtained from reaction time-courses. Remarkably, the increase in the catalytic efficiency of LovD-BuCh2 accompanies an increase in K_i, proving so that the V261H mutation benefits catalysis by reducing substrate inhibition thus decreasing K_M. Finally, LovD9 variant shows both the highest k_{cat}/K_M and K_i of all variants herein studied. Therefore, these kinetic data support that the last evolution rounds (from LovD7 and LovD9) minimize MJA inhibition to decrease the Michaelis constant (K_M).
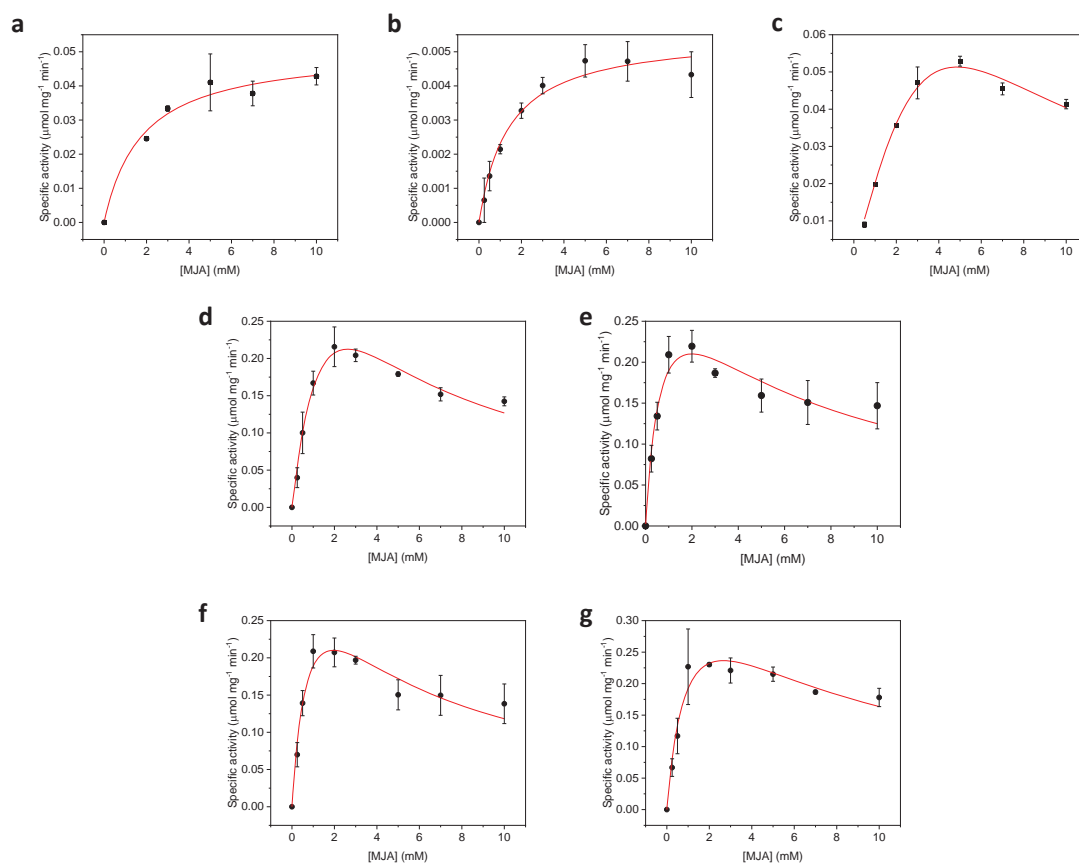
Figure 4. Michaelis-Menten (MM) plots for monacolin J acid (MJA). The experimental data (black dots) were fitted (red line) with the proposed MM and substrate inhibition models. 2 mM of α-dimethylbutyryl-*S*-methyl-mercaptopropionate (DMB-SMMP) was incubated with 0.25-10 mM of MJA and 1 µM of enzyme, for different variants: LovD **(a)** LovD-Ch **(b)**, LovD-Bu **(c)**, LovD-BuCh1 **(d)**, LovD-BuCh2 **(e)**, LovD7 **(f)** and LovD9 **(g)**. Each experimental data is the mean value of three independent replicates.

Table 4. Kinetic parameters of LovD variants varying the concentration of monacolin J acid (MJA) (0.25; 0.5; 1; 2; 3; 5; 7 and 10 mM) at 2 mM α-dimethylbutyryl-S-methyl-mercaptopropanoate (DMB-SMMP). k_{cat} is the turnover number, K_M is the Michalis constant and K_i is the inhibition constant. k_{cat}/K_M is the catalytic efficiency (specificity constant).

Mutant	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1}s^{-1})	K_i (mM)
LovD	0.004 (± 0.002)	4.2 (± 1.2)	1.0 (± 0.3)	-
LovD-Ch	0.004 (± 0.001)	1.4 (± 0.7)	2.9 (± 1.1)	-
LovD-Bu	0.24 (± 0.15)	14.6 (± 1.6)	16.4 (± 6)	1.8 (± 0.3)
LovD-BuCh1	0.54 (± 0.15)	3.2 (± 1.3)	168.8 (± 57)	2.7 (± 1.3)
LovD-BuCh2	0.35 (± 0.07)	1.0 (± 0.27)	350.0 (± 173)	4.1 (± 1.2)
LovD7	0.29 (± 0.01)	0.9 (± 0.1)	322.2 (± 13)	5.7 (± 1.2)
LovD9	0.29 (± 0.01)	0.6 (± 0.1)	483.3 (± 103)	8.4 (± 1.5)

4.4.3. Evaluation of thioesterase activity as a competing reaction for the acyltransferase capacity of LovD variants

As DMB-SMMP is a surrogate substrate for the acyl transfer reaction, the catalytic efficiency of the native ACP-dependent LovD to accept the acyl group from such an artificial substrate is very low. The evolution path towards an efficient simvastatin synthase must necessarily improve the DMB-SMMP binding affinity to efficiently set up the acyl-enzyme complex that will be further attacked by MJA to ultimately yield SVA. Nevertheless, water competes with MJA for attacking the acyl-enzyme complex, driving the enzyme back to its initial state and releasing 2,2-dimethylbutanoic acid (DMB) as an undesired by-product (**Scheme 1**). This side reaction is fatal to attain high SVA yields because it consumes one of the starting materials, derailing the process towards unwanted products. According to this, LovD acts as a promiscuous enzyme with two competitive catalytic capabilities: thioesterase (unproductive hydrolysis of DMB-SMMP) and acyltransferase (productive acyl transfer from DMB-SMMP to MJA). To assess which catalytic capability is preferred by each LovD variant, we developed a colorimetric method based on the detection of methyl 3-mercaptopropanoate released during the first reaction step with and without the acyl acceptor (MJA) in the reaction media. Measurements in the absence of MJA reflect the thioesterase activity (TE), where only water attacks the acyl-enzyme complex, while those performed in the presence of MJA report a competition between the thioesterase and the acyltransferase (AcT) activities, thus a competition between water and MJA to attack the acyl-enzyme complex. A similar methodology has been recently exploited for the identification of promiscuous acyltransferase activity in esterases.[290] The (AcT:TE) ratio measures which activity dominates over the other: low ratios indicate that hydrolysis of the acyl-enzyme complex is faster than acyl-transfer to MJA, while higher ratios prove the prevalence of acyl transfer reaction.

Figure 5.a shows that from evolution round 2 to 6, the thioesterase activity scales with the SVA productivity, suggesting that the acyltransferase activity of LovD evolved in parallel to its thioesterase activity. The LovD scaffold likely evolved to better allocate DMB-SMMP as an acyl donor until the sixth evolution round. This adaptation to artificial evolutionary pressure is particularly apparent in round 4, where the substrate entrance channel mutations M157V and S172N are introduced in LovD4 variant. These two mutations seem to underpin the productive binding of DMB-SMMP into the active site to form the acyl-enzyme complex. Once enough affinity towards DMB-SMMP was achieved, further evolution rounds slightly decreased the thioesterase activity but still increased further the SVA synthetic rate. Remarkably, **Figure 5.b** shows an inflexion point in the AcT:TE activity ratio after LovD4 where the evolution path apparently starts to gradually reduce the unproductive hydrolysis of the acyl-enzyme complex, while kept increasing the capacity of that complex to transfer the acyl group to MJA. We identify another inflexion point after the 7th evolution round, where the AcT:TE activity ratio reaches its maximum value that declines steadily afterwards. It is remarkable that the most active variant (LovD9) shows an AcT:TE activity ratio only slightly higher than the wild-type.
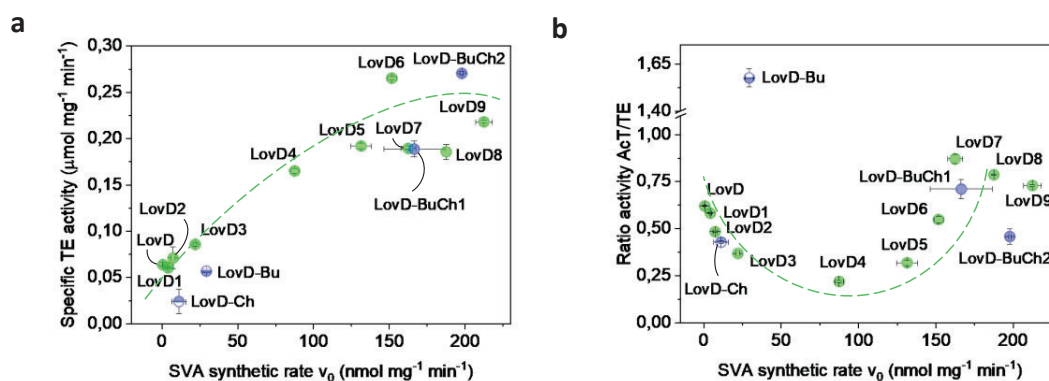
Figure 5. (a) Specific thioesterase activity *versus* simvastatin acid (SVA) synthetic rate (v_0), calculated through reaction time-courses (see **Figure 3.d,f**). **(b)** Acyltransferase/thioesterase activity ratio (AcT:TE) *versus* SVA synthetic rate (v_0). Green circles represent the native and engineered variants obtained through the laboratory evolution experiment (LovD and LovD1-9). Half-filled blue circles represent the separated cluster variants (LovD-Bu and LovD-Ch). Full-filled blue circles represent the variants with the two clusters (LovD-BuCh1-2).

The deconvoluted channel variant LovD-Ch exhibited even lower thioesterase activity and AcT:TE ratio than wild-type LovD. On the contrary, the LovD-Bu variant was as good thioesterase as the native enzyme, but it was the only variant herein studied whose acyltransferase activity significantly outperforms the thioesterase one (AcT:TE = 1.6), so the enzyme prefers MJA as a nucleophile rather than water. These data clearly indicate that both buried and channel residues by themselves do not play a role in the formation of the acyl-enzyme complex, but buried mutations strongly contribute to improving the acyltransferase capability of LovD variants. These insights are supported by the fact that the LovD-Bu variant exhibits 4 times higher SVA synthetic rate than the LovD-Ch one. When both buried and channel clusters were merged, LovD-BuCh1 outperforms the thioesterase activity of the variants harboring each cluster separately, exhibiting an AcT:TE ratio roughly 2 times lower than LovD-Bu. Hence, we suggest that the similar SVA synthetic rates (v_0) measured for both deconvoluted LovD-BuCh1 and evolved LovD7 rely on the combination of these channel and buried clusters that additively enhance the kinetics of both first (acyl-enzyme complex formation) and second (acyl-transfer) steps. Comparing LovD-BuCh1 and LovD-BuCh2, the latter favors both thioester hydrolysis and SVA synthesis. Moreover, the AcT:TE ratio of LovD-BuCh2 was lower than the one observed for LovD-BuCh1. These results suggest that a His residue in position 261 contributes to form the acyl-enzyme complex using the DMB-SMMP as an acyl donor.

Altogether, these results support that laboratory evolution experiments can sequentially optimize several features (acyl donor intake, acyl-enzyme complex stability, acyl transfer rate and product hydrolysis) of the complex reaction mechanisms. In this particular case, the *in vitro* evolution of LovD reshapes the acyl transferase active site to adapt an unnatural acyl group (2,2-dimethylbutiryl) donated by an abiological acyl donor (DMB-SMMP), forming an acyl-enzyme complex that is preferentially attacked by the enzyme natural acyl acceptor (MJA).

4.4.4. Evaluation of detrimental esterase activity catalyzed by LovD variants

In a batch enzymatic synthesis, product hydrolysis arising from the reversible nature of this reaction limits SVA productivity, which obligates to add excesses of MJA to shift the equilibrium towards the synthetic direction and thus maximize SVA titers. As SVA hydrolysis has already been reported when using wild-type LovD[261], we characterized this unwanted reversible activity for all the variants harvested in the directed evolution campaign as well as for the original cluster variants herein developed. **Figure 6** shows the correlation between the SVA hydrolytic and synthetic activities for all the mutants.

The SVA hydrolysis evolution profile follows a similar trend to that one observed for the AcT:TE ratio (**Figure 5.b**). **Figure 6** shows how the hydrolytic activity increases until the 4th evolution round as LovD4 hydrolyzed 7 times more SVA than the wild-type enzyme. After the 5th round, the SVA hydrolysis diminished until round 7, and then slightly increased again during rounds 8 and 9. The higher hydrolytic capacity of LovD9 is likely compensated by its lower substrate inhibition profile, which might be one of the main factors contributing to its superior SVA productivity.
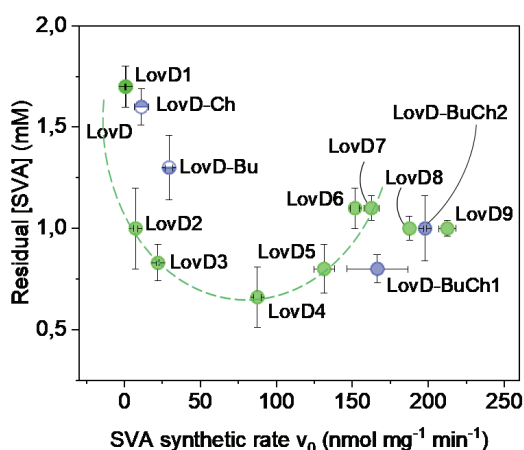
Figure 6. Correlation between simvastatin acid (SVA) hydrolysis and SVA synthetic rate (v_0) of laboratory evolved and deconvoluted cluster variants. SVA hydrolysis is determined as the residual SVA concentration after 24 h of reaction, being 2 mM the initial SVA concentration. v_0 is determined through the time-course reactions shown in **Figure 3.d,f**. Green circles represent the native and engineered variants obtained through laboratory evolution (LovD and LovD1-9). Half-filled blue circles represent the separated cluster variants (LovD-Bu and LovD-Ch). Full-filled blue circles represent the variants with the two clusters (LovD-BuCh1-2).

Along with the improvement of the acyl-enzyme complex formation supported by the enhancement of the thioesterase activity, the first 4 evolution rounds also make LovD variants able to more efficiently hydrolyze SVA, reinforcing the important role of water accessibility to the active site. On the contrary, the second half of the evolution pathway diminishes the hydrolytic reaction in favor of the synthetic one. When the SVA hydrolytic

activity was assessed for both LovD-Bu and LovD-Ch, both deconvoluted variants presented higher hydrolase activity than the wild-type enzyme but lower than any of the other evolved mutants beyond LovD1. As expected, the combination of buried and channel mutations in LovD-BuCh1-2 variants increased both synthetic and hydrolytic SVA activities. Once again, the evolution path re-shaped the substrate channel and the binding pocket to productively accommodate both acyl donor and acceptor to synthesize SVA. However, that molecular re-shaping also facilitates the binding of the acetylated product in the absence of the acyl acceptor, explaining the enhancement of the hydrolytic activity during the first part of the evolution path.

4.4.5. Assessing the role of cluster mutations in active site integrity and acyl acceptor binding through MD simulations

In our previous report,[252] we rationalized the gain in catalytic activity throughout laboratory evolution of different LovD variants, by analyzing the structural integrity of the active site (i.e., the Ser76-Lys79-Tyr188 catalytic triad) using MD simulations. We proposed that wild-type and the early evolved variants present a catalytic Tyr188 able to experience wider motions, as reflected by conformational changes, that disrupted the catalytic triad thus hampering the catalytic activity in the absence of the acyl carrier protein partner LovF. This feature was elusive to x-ray crystallography, which showed a high degree of similarity of the catalytic triad among different LovD variants despite their very different activities. Herein, we further analyze the ability of different variants in their acyl-enzyme complex forms to transfer the α-dimethylbutyryl group to the acyl acceptor MJA. To this aim, we generated models for the selected evolved and cluster Ser76-acylated mutants with MJA bound to the active site as in available crystallographic structures (**Figure 7.a**).[251] These models were further subjected to microsecond MD simulations under physiological conditions (see **Computational Section**).

By monitoring the distance between the reactive fragments of acylated Ser76 and MJA (carbonyl and hydroxyl groups, respectively) throughout the simulations, we uncovered the ability of the intramolecular interactions optimized by the engineering process to keep the substrate in a productive configuration. For wild-type LovD, such distance fluctuates between 2.7-6.2 Å for the first 600 ns and finally MJA abandons the initially modeled binding mode. On the contrary, such distance is maintained at 3.3±0.2 Å along the whole simulation of the very efficient variant LovD6, reflecting the superior ability of highly evolved mutants to bind MJA in a productive orientation (**Figure 7.b**). Also, and reproducing the trend observed in our previous report,[252] the catalytic Tyr188 of the native enzyme undergoes a conformational transition at the early stages of the simulation (*ca.* 130 ns) which obliterates its necessary catalytic contacts with both the acylated donor and acceptor (**Figure 7.c**). Noteworthy, we observed a similar trend in the MD simulation of the cluster variant featuring only the channel mutations (LovD-Ch). On the contrary, cluster variants LovD-Bu and LovD-BuCh1 comprising mutations in buried regions of the protein, maintained the starting, productive MJA binding mode and catalytic contacts (**Figure 7.d,e**). Therefore, MD studies

support the notion that the improvement of the acyl transfer step to synthesize SVA mainly relies on buried mutations rather on channel ones.
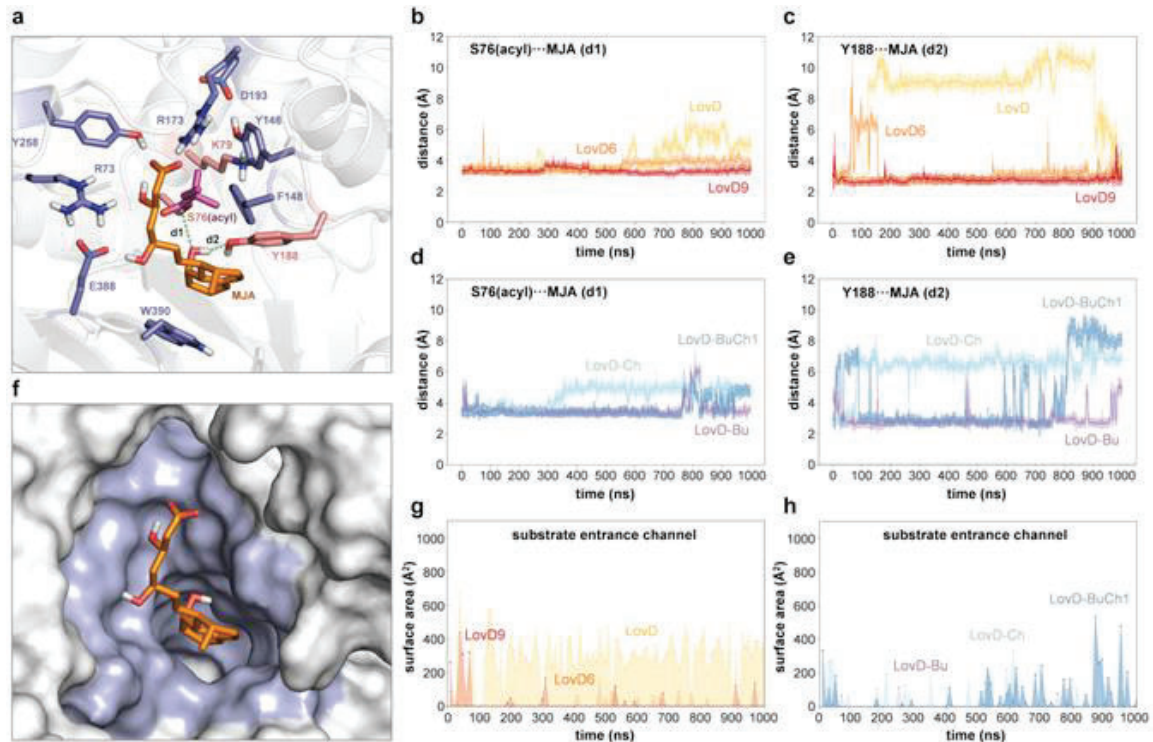
Figure 7. **Molecular dynamics simulations of LovD variants** (**a**) Active site residues (in blue), catalytic triad (in pink; acyl group in purple) and substrate (in orange) for modeled acyl-enzyme complexes of LovD bound to monacolin J acid (MJA). (**b,d**) Distance between the reactive carbonyl carbon of acylated Ser76 and hydroxyl group of MJA and (**c,e**) between the reactive hydroxyl groups of MJA and Tyr188 along the molecular dynamics (MD) simulations. (**f**) Surface representation of the residues constituting the substrate entrance channel (in blue) of LovD, together with bound MJA (in orange sticks). (**g,h**) Computed surface area of the main tunnel cluster calculated along the MD simulations.

Furthermore, molecular dynamics on the *apo* protein unveiled another structural feature elusive to crystallography, namely the size and depth of the substrate entrance channel to the active site (**Figure 7.f-h**, **Figure 8**). The ACP-dependent wild-type LovD variant showed a much wider channel that remains open most of the simulation time, being prone to accommodate the 4'-phosphopantetheine (PPN) acylating group present in its natural acylating partner; LovF.[252] However, highly evolved mutants showed a much narrower channel which barely opened during the simulations (17% and 20% of the simulation time for LovD6 and LovD9, respectively). This trend of narrowing the substrate entrance channel is observed also for the cluster variants, particularly for LovD-Bu, supporting how the laboratory evolution drastically alters the structure and dynamics of a protein scaffold.[143] Besides eliminating its need for LovF-ACP to render the catalytic site organization through mutations, substantially narrowing the substrate entrance channel likely mitigates exposure of both the acyl-enzyme complex and bound SVA to water, thus reducing detrimental

hydrolytic activities and favoring product formation as supported by the experimental data shown above (**Figures 5** and **6**).
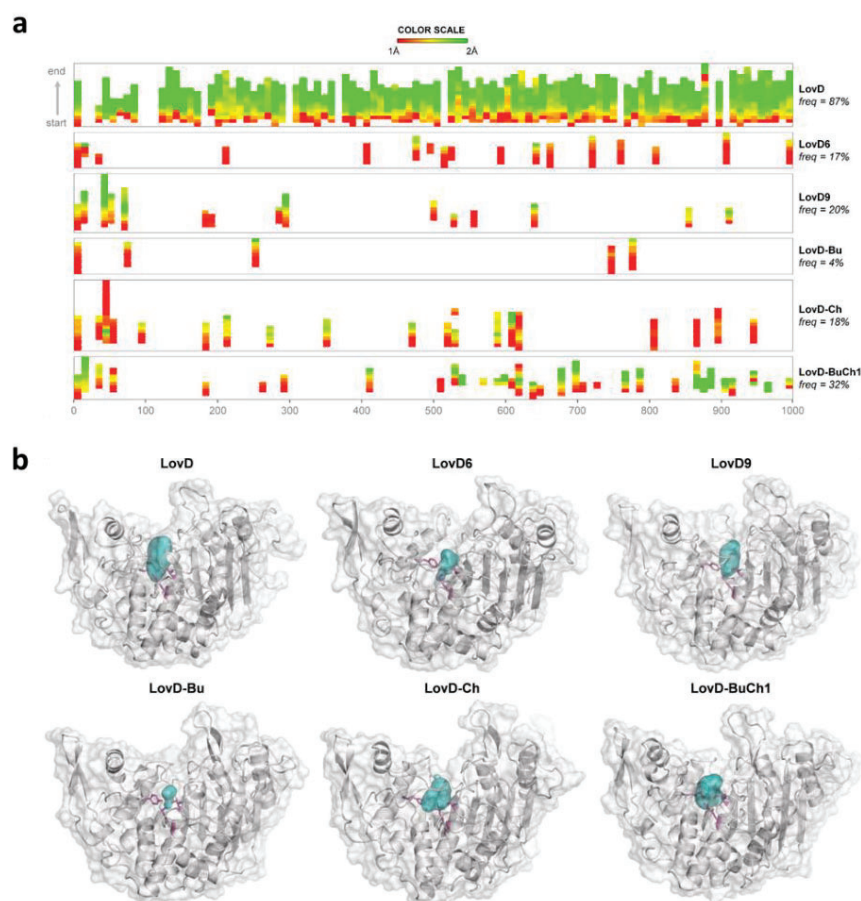
Figure 8. **(a)** Profile heat map of the main tunnel cluster calculated for selected directed evolution and cluster LovD variants from molecular dynamics (MD) simulations. The first bar represents the average of the tunnels calculated throughout each trajectory. The height of each bar represents the tunnel length (i.e., number of segments constituting the tunnel). The color scale represents the width (i.e., bottleneck radii) of each tunnel segment from narrow (red) to wide (green). For wild-type, tunnels are found with a high frequency (87%) and all calculated tunnels become wider as the distance from the starting point increases. For highly evolved LovD6 and LovD9 and cluster LovD-Bu, LovD-Ch and LovD-BuCh1 variants, tunnels are found with a much lower frequency (8-20%) and the calculated tunnels are shorter and narrower. For Lov-BuCh1, found tunnels become wider as the simulation progresses. **(b)** Surface representation (in teal) of the main tunnel cluster calculated selected directed evolution and cluster LovD variants from molecular dynamics (MD) simulations. The catalytic triad (Ser76-Lys79-Tyr188) is shown as magenta sticks. Protein backbone is shown as a faded grey cartoon.

4.4.6. Thermal stability of LovD variants

Thermal stability of enzymes is a hallmark that guarantees the robustness of one biocatalyst for its biotechnological applications. Normally, highly thermostable enzymes operate for longer times under drastic conditions, reducing the enzyme replacement costs of the process. Reaction time-courses (**Figure 3.b-f**) demonstrate that mutations at the protein surface have a less significant impact in enhancing the catalytic properties of LovD for the synthesis of SVA. Despite of being significantly far from the substrate entrance channel and the active site, those external mutations might play a role in the overall stability of the enzyme as seen for other engineered biocatalysts.[43,291]

Figure 9.a represents the melting temperature (T_m) of all the directed evolution and cluster designed variants of LovD, revealing how the evolution process improves the thermodynamic stability of the enzymes against temperature. A maximum T_m (52 ºC) was achieved after the 6th evolution round. All the deconvoluted variants with the channel, buried and the combination of both mutations showed lower T_m than LovD6-9 variants. These data confirm the relevance of surface mutations, mostly introduced in the second half of the evolution process.

Figure 9.b shows the thermal inactivation first-order kinetics at 37 °C for wild-type LovD, an intermediate (LovD4) two of the most evolved variants (LovD7 and LovD9) and the four deconvoluted variants herein developed. These inactivation courses were fitted to **Equation 3** to determine their inactivation constants and half-lives (**Table 5**). The complete inactivation of LovD occurred after 30 min. Consequently, the native enzyme was 40 times less stable to thermal stress than the most evolved LovD9 variant. The 9 mutations introduced in LovD4 significantly stabilized the protein scaffold and the most evolved LovD7 and LovD9 seemed to have reached the same stability limit. Cluster variants LovD-BuCh1 and LovD-BuCh2 exhibited respectively a half-life of around 3.3 and 1.3 times shorter than the most evolved variants. Therefore, the trend observed in the kinetic thermal deactivation (**Figure 9.b**) agrees with the data extracted from the thermodynamic resistance to unfolding, which is maximum in LovD6 (**Figure 9.a**). Hence, we postulate that the external positions introduced along the evolution pathway have an important role in increasing enzyme stability. From these data, it is apparent that the increasingly harsher conditions employed during laboratory evolution imposed a selective pressure towards stability, which in turn contributes to the higher productivity of the most evolved variants.
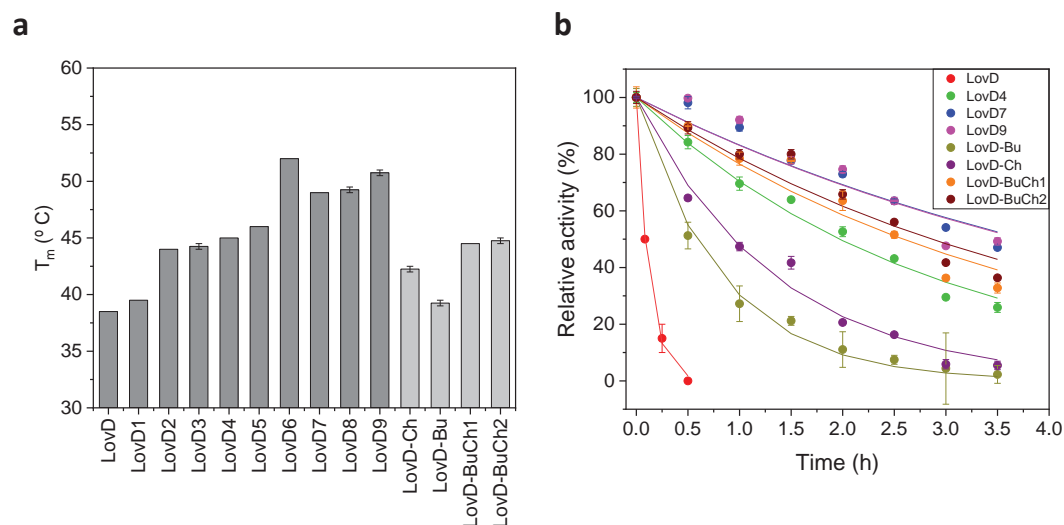
Figure 9. (a) Thermal shift of LovD variants from 25 to 95 °C. **(b)** Thermal inactivation profiles for selected LovD variants measured at increasing incubation times at 37 °C. Residual activity (%) at different times means the enzyme activity at time = t divided by the enzyme activity at t = 0 (before thermal incubation) and expressed as a percentage.

Table 5. Kinetic inactivation constant (k) and half-life ($t_{1/2}$) measured for different LovD variants.

Enzyme	k	$t_{1/2}$ (h)
LovD	8.082 (± 0.898)	0.09 (± 0.010)
LovD4	0.352 (± 0.012)	1.97 (± 0.002)
LovD7	0.184 (± 0.053)	3.77 (± 0.003)
LovD9	0.185 (± 0.005)	3.74 (± 0.11)
LovD-Ch	0.743 (± 0.006)	0.93 (± 0.008)
LovD-Bu	1.195 (± 0.004)	0.58 (± 0.002)
LovD-BuCh1	0.659 (± 0.013)	1.05 (± 0.020)
LovD-BuCh2	0.242 (± 0.007)	2.86 (± 0.08)

This thermostability trend could be qualitatively reproduced using Rosetta protein structure prediction software[103] (see **Experimental Section** and **Figure 10**). **Figure 10.a** shows calculated Rosetta scores (i.e., total system energy, TSE) for selected directed evolution and cluster LovD variants. The lowest (i.e., most negative) scores related to that calculated for wild-type LovD (ΔTSE) provide a rough estimation of the relative thermostability of each enzyme[292] (**Figure 10.b**). Of note, the Pearson correlation coefficient between the calculated ΔTSE and experimental T_m values is remarkably good for directed evolution variants LovD to LovD9 (ρ = -0.93; strong negative correlation) but decreases when the cluster mutants are included (ρ = -0.77; weak negative correlation). Together with these energy-related values, the gradually decreasing root-mean- square deviations (rmsd) with respect to LovD9's

crystallographic structure, suggest that laboratory evolution gradually optimized thermostability while sculpting the protein backbone for improved simvastatin productivity.
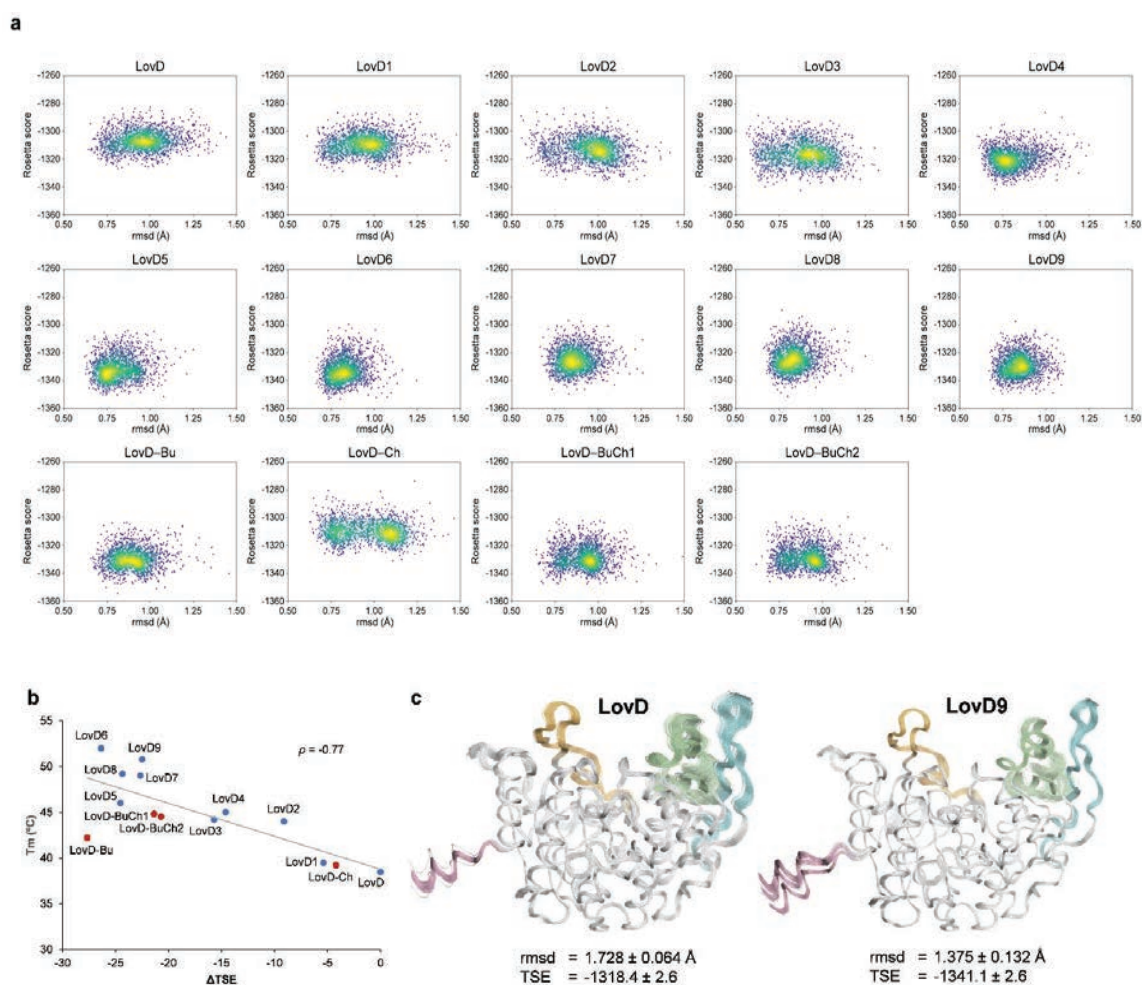
Figure 10. (a) Distribution of Rosetta scores (i.e. total system energy, TSE) calculated for model full-atom structures of all directed evolution and cluster LovD variants (2000 decoys for each mutant, blue dots). The root-mean-square deviation (rmsd) with respect to the Rosetta-minimized crystallographic structure of *apo* LovD9 allowed identifying two clusters of structures for some variants (LovD2, LovD3, LovD-Ch and LovD-BuCh1-2), likely reflecting different suboptimal conformational states of flexible loops in intermediate and cluster variants.[252] **(b)** Relative total system energies (ΔTSE) calculated from Rosetta scores for directed evolution (blue) and cluster (red) LovD variants, versus their experimental melting temperatures (T_m). The Pearson correlation coefficient (ρ) is calculated using all data points. **(c)** Ribbon representation of the 200 lowest-energy decoys calculated for wild-type LovD (left) and LovD9 (right). Highly flexible loops, most of them being part of the substrate entrance channel, are colored in pink (residues 1-11), cyan, (residues 104-128), green (residues 147-179) and orange (residues 243-264).

Remarkably, modeled structures showed a high flexibility at the loops located near the substrate entrance channel; these loops become more rigid at the late stages of evolution, particularly in LovD9 (**Figure 10.c**). This observation, together with the narrower tunnel

calculated for evolved and cluster variants with respect to LovD (**Figure 7.g** and **Figure 8**), suggest that evolution concurrently increased catalytic activity and thermostabilized the protein backbone, at least in part, by partially locking flexible loops in a productive conformation.[252] Such evolutionary strategy (or better said, outcome) likely mimics the effect of protein-protein interactions occurring between native, wild-type LovD and its acylating partner protein LovF.[252]

4.5. Conclusions

In this study, we have elucidated the intricate factors governing the reactivity of several variants of acyltranferase LovD engineered by directed evolution. Through deconvoluting the most evolved variant LovD9 in two mutational clusters that involve residues at the entrance channel and binding site of the substrates, we shed light on how the evolution path gradually improves the complex functionality of this acyl transferase. The functional analysis of the directed evolution pathway unveiled how the mutations inserted during the first rounds of the directed evolution of LovD[252] enhanced the productive intake of the acyl donor surrogate (DMB-SMMP), generating an acyl-enzyme intermediate more efficiently.

Dissecting the 29 mutations of LovD9 into three clusters, we found that mutations at buried positions largely favor the acyl transfer vs. acyl-enzyme complex hydrolysis, while mutations at the substrate entrance channel mitigate the inevitable product hydrolysis. Remarkably, the combination of these two clusters had a cumulative effect as supported by the high productivity that LovD-BuCh1-2 exhibit towards the synthesis of SVA. Furthermore, the different functional properties exhibited by LovD-BuCh1 and 2 suggest that the polarity of the solvent-exposed position 261 at the substrate entrance channel enhances enzyme acylation. The SVA productivity and thermostability observed for LovD-BuCh2 were comparable to the most evolved variants but with only half of the mutations. These insights demonstrate the dominant but cumulative role of buried and substrate entrance channel alterations in the enhancement of both enzyme activity and thermostability. Therefore, the mutational cluster deconvolution emerges as a valid approach to comprehend the specific contribution of mutations inserted during the laboratory evolution paths as well as to find new minimalist variant that reduce the number of mutations without worsening the functional properties of the most evolved variants. We foresee that this molecular exercise supported by experimentally and computational data has built the basis for future designs of new variants involving alternative mutational pathways.

CHAPTER 5:

Computational methodologies for *in silico* design and screening of new LovD variants

5.1. Introduction

De novo **design of enzymes** that can promote non-natural chemical reactions can be achieved developing computational methods based on minimal quantum mechanical (QM) models (theozymes) and protein scaffold databases, such as Kemp-eliminases,[137–139] retro-aldolases[140,141] and Diels-alderases.[142,143] Nevertheless, all these *de novo* designs present much lower activity that enzymes catalyzing natural reactions.

Laboratory-directed evolution (LDE) makes use of high-throughput screening of large libraries to explore a broad range of protein functions.[43] Due to their high selectivity, enzymes use to exhibit low levels of activity for side reactions involving non-natural substrates, but this promiscuity gives room to LDE. While natural evolution occurs in a very large time scale (millions of years), LDE operates in a much more accelerated process. A clear successful example of LDE is the conversion of halohydrin dehalogenase (HHDH) to accept non-natural nucleophile CN^-.[90,293] However, the vast number of variants generated makes LDE dependent of high-throughput screening techniques, which are time consuming and cost expensive. Moreover, LDE does not give a structural explanation of how mutations are correlated with activity, making a key challenge the identification of individual variants that display the desired improvements out of the large set of random changes.

Rational or semi-rational enzyme engineering normally focus on mutating positions located near the active site to force a change in activity. By contrast, LDE introduces many mutations in remote positions, suggesting whether an allosteric effect of those positions in catalysis or a stabilizing effect in the protein backbone. Rational design and LDE are often combined to overcome both limitations, but the success is still moderate to develop a standardized procedure to engineer any enzyme.

In the last years, tremendous advances have been experienced in some experimental techniques that provide a detailed protein structural insight, such as nuclear magnetic resonance (NMR),[294–296] X-ray crystallography[297–299] or cryo-electron microscopy (Cryo-EM),[300–302] being very valuable to better understand the **Structure-Activity Relationship (SAR)** of proteins. Nevertheless, one of the most relevant features of proteins in solution is their dynamism, given by their structural changes in a conformational landscape of transition structures with different energies, and affected by their interaction with water molecules and ions.

Molecular dynamics (MD) of proteins are a beneficial tool for understanding the relation between the protein structure dynamics and its function. By that means, improving the accuracy of predictions can lead to the obtaining of novel designs.[303] The main advantage in the use of MD is the capability of providing atomic details of given conformational structures and remark the differences between native and engineered states. With continuing advances in computing (computer speed, advanced processing software, etc.), the accuracy and simulation times of MD is exponentially improving in the recent years, turning it into a more attractive technique for protein modelling. The combination of all

experimental (LDE) and computational approaches (QM theozymes and MD) are proven to be a winning strategy to reconstruct SAR and obtain reliable predictions to propose new enzyme designs.[134]

Enzyme design software **RosettaDesign** identifies low energy amino acid sequences for targeted protein structures.[103,304] This software uses Monte Carlo optimization to search for amino acid sequences that pack well, bury hydrophobic residues in the core and satisfy hydrogen bonding of polar residues. A higher sequence variability is observed in the surface of the protein, due to fewer packing constraints. RosettaDesign has been used for protein stabilization[305], enhancement of protein binding affinity[42] and generation of proteins with novel structures.[306]

Even in the absence of large conformational shifts, variability introduced with mutagenesis in a particular position can affect the dynamics of other protein regions through allosteric effects. Computational efforts have been directed to construct optimal pathways of correlated motions between allosteric sites and the catalytic pocket of enzymes. **Weighted Implementation of Suboptimal Paths (WISP)** is also capable of identify slightly longer additional suboptimal pathways. WISP was applied for the first time to study the glutamine amidotransferase HisF-HisF.[307] Paths are constructed as a network between nodes, which correspond to amino acid positions. **Betweenness centrality** is a way of detecting the amount of influence a node has over the network. It is often used to find nodes that serve as a bridge from one part of the network to another.[308]

On the other hand, when structural information is not available, **multiple protein sequence alignment (MPSA)** provides direct information to identify potential mutation sites. A clear example is the activity enhancement of endo-1,4-β-glucanase of *Reticulitermes speratus* (RsEG) via sequence comparison with other cellulases from different sources.[309] MPSA are usually required for the identification of functional-related residues, specificity determining-positions, homology modelling and protein function prediction. **ClustalW** is a classical MPSA method widely exploited to align highly homologous sequences.[310] Its application for the construction of 100 homologues of prolyl endopeptidase from *Sphingomonas capsulate* (ScPEP) afforded the identification of 30 potentially beneficial mutations.[311]

As already mentioned in **Chapters 2** and **4**, LDE has been successfully applied to the acyltransferase LovD towards the synthesis of the anti-cholesterol drug simvastatin acid (SVA).[251,252] Crystallographic studies of this LDE process were unable to distinguish between active and inactive variants, maintaining the backbone and side chain orientations of the catalytic triad in LovD, LovD6 and LovD9 in a conformation very close to the optimum positions predicted through QM optimizations. On the contrary, a computational MD analysis of the LovD variants revealed the existence of catalytic and non-catalytic conformations of the active site.[251] In the simulations of the wild-type enzyme, the apparently catalytic conformation of the Ser76-Lys79-Tyr188 triad in which all the three residues are arranged at less than 3 Å of distance from each other, underwent into an

inactive conformation characterized by a much longer Lys79-Tyr188 (9 Å) distance produced by a conformational change in Tyr188. The first round of evolution (LovD1) carrying the single mutation (L361M) restores Tyr188 backbone in its catalytic conformation for about 300 ns before undergoing to the non-productive configuration. The introduction of subsequent rounds of LDE increased the simulation time in which Ser76-Lys79-Tyr188 remained in a catalytically competent configuration, converging to its maximum from LovD6 on.

Stablishing the Lys79-Tyr188 distance as a successful criterion to discern between active and inactive variants, the main objective of this study was the development of a computational method (**Scheme 1**) useful for the identification of protein sequence alterations that can generate highly active engineered variants of the acyltransferase LovD for SVA synthesis. We proposed three different *in silico* enzyme design protocols:

1- **RosettaDesign/MD in predefined mutation sites:** we generated 60 variants in the same positions found by LDE with RosettaDesign software and subjected all the variants to iterative **RosettaDesign**/MD simulations with the catalytic atom map constrained in the catalytically competent arrangement. After that, unrestrained 1 μs MD simulations were performed to obtain a fully relaxed LovD variant. A total of 10 LovD candidates that maintained the catalytic conformation during most of the MD trajectory were selected to be tested *in vitro*. From them, only 8 were expressed as soluble proteins, and none of the designs showed activity for SVA synthesis.

2- **Identification of hotspots through allosteric dynamic analysis:** we assessed the identification of hotspots of mutations different from the LDE positions carrying out an allosteric communication analysis through MD simulations. This analysis was performed with WISP and consisted in the generation of a correlation matrix from a 200 ns MD trajectory in which we identified the residues with a higher variance of the correlation motion comparing the calculated paths in the catalytic and non-catalytic conformation of the Ser76-Lys79-Tyr188 triad. In this manner, an allosteric effect of these positions to the active site is assumed. Two hotspots located in external loops close to the channel by which the substrate access the active site were identified. Then, sequence search of homologues completed the protocol to select mutable residues in the identified positions.

3- **Identification of hotspots through betweenness centrality analysis and homology modelling:** Hotspots search was also achieved through WISP betweenness centrality analysis, which correlates the motion of each residue of the protein with the rest of the residues. Then, sequence search of homologues completed the protocol to select mutable residues in the identified positions. In this case, positions scattered along all the protein backbone were identified as mutable.

From all the identified positions with strategies **2** and **3**, we selected 9 single mutants (N43R, D119S, A123P, S164G, S172N, Y279M, F363L, F363W and F384V), 2 double mutants

(D119S/A123P and Y279M/F363W), 4 triple mutants (N43T/D119S/A123P, Y279M/Y327F/F363W, Y279M/Y327F/F363Y and Y279M/Y327F/F363L) and 4 multiple mutants (Table 1) to test their expression and activity *in vitro*. Out of the 19 tested variants, 7 of them showed a significant increase in SVA synthetic activity when compared to the wild-type.
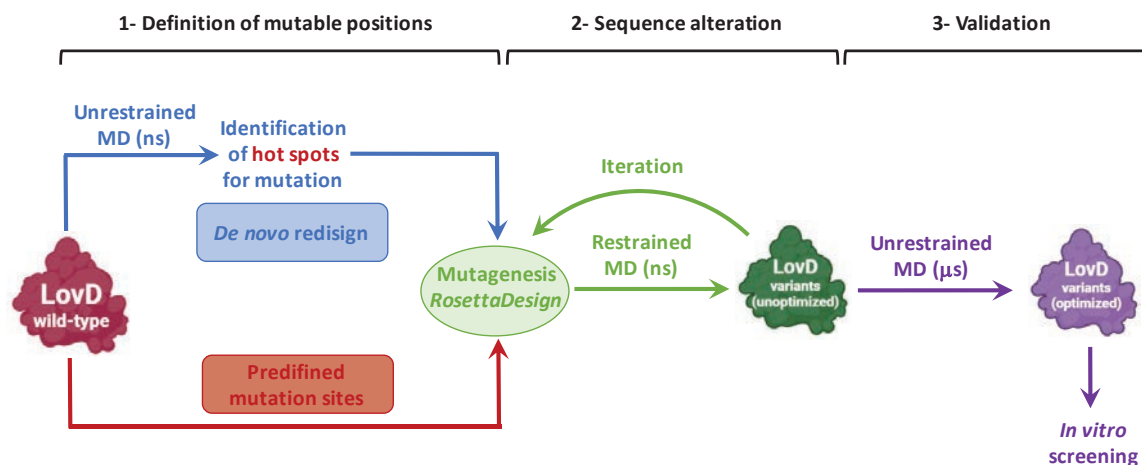
Scheme 1. Workflow for the iterative RosettaDesign/Molecular dynamics computational design and evaluation protocol.

5.2. Computational section

5.2.1. Molecular dynamics simulations

MD simulations were performed at the same conditions that in the Computational Section of Chapter 4, with the exception of the run time of production, which was 200 ns.

5.2.2. RosettaDesign combined with MD

A protocol combining MD simulations and RosettaDesign was employed to LDE predefined mutation sites to proposed alternative mutations in the same regions. First, a short MD (2 ns) with the catalytic atom map constrained in the catalytically competent arrangement is run for the wild-type initial structure, then a first design is generated by Rosetta (allowing to mutate to the 20 possible amino acids). The structure presenting the lowest Rosetta score among 5 possible poses is considered the starting point for the next iteration of the protocol. Ten independent replicas of this protocol were run so that 10 different designs were obtained. Finally, 1 μs unrestrained MD simulation is carried out.

5.2.3. Homology analysis

PSI-BLAST (Position-Specific Iterated BLAST) program[312] was used to search for the most similar protein sequences to LovD from the non-redundant protein sequences database. Then, sequences were exported as FASTA files and filtered to just keep the transesterase variants. These variants were later on aligned with the help of Clustal program.[313] Finally, Jalview software[314] was used to calculate the amino acid composition in percentage for each position.

5.2.4. Allosteric communication analysis through MD simulations

The Weighted Implementation of Suboptimal Paths (WISP)[307] was used for the analysis of dynamical networks. First, a correlation matrix (C_{ij}) is generated from 2.000 snapshots extracted every 1.0 ns from a converged 200 ns MD trajectory, by calculating the correlation motion among node-node pairs with **Equation 1**. In our model, nodes are defined by the whole-residue center of mass, and two nodes are considered to be in contact if the mean distance between them along the MD simulation is 6 Å or less. The length of the edges connecting these nodes quantifies the degree of dynamic communication between pairs of connected nodes as defined in **Equation 2**. This pathway length is inversely proportional to the correlation motion between nodes, meaning that shorter w_{ij} values indicate tightly correlated or anticorrelated nodes, whereas larger values indicate less correlated nodes.

$$C_{ij} = \frac{\langle \Delta \vec{r_i}(t) \cdot \Delta \vec{r_j}(t) \rangle}{\sqrt{\langle \Delta \vec{r_i}(t)^2 \rangle \langle \Delta \vec{r_j}(t)^2 \rangle}} \qquad (1)$$

$$w_{ij} = -\log(|C_{ij}|) \qquad (2)$$

Then, Dijkstra's algorithm is used to generate all force-node paths, finding the shortest (i.e. optimal) path. To identify not only the optimal but also suboptimal pathways, WISP employs a bidirectional search. Suboptimal pathways are defined as those closest in length to the optimal one, but not including it. The available code rapidly calculates both optimal and suboptimal communication pathways between two user-specified residues of a protein. A total number of 100 pathways were calculated between the residues of the catalytic triad (Ser76, Lys79, Tyr188) and all the other residues of the protein. These paths were recalculated for two different forms of the wild-type system, the former when the catalytic triad residues forming the catalytic triad are organized in the catalytic conformation and the latter when these residues have abandoned this conformation. The average distances of all calculated paths in the *catalytic* form were compared to those in the *non-catalytic* form in order to locate which residues exhibit the highest variance in their communication with the active site due to the conformational change of the catalytic triad.

5.2.5. Betweenness centrality analysis

The betweenness centrality was employed as a metric to identify possible hotspot positions to be mutated. This parameter quantifies the number of times a node acts as a bridge along all the shortest paths between all distinct pairs of nodes. Betweenness centrality was calculated for all residues.

5.3. Experimental section

5.3.1. Materials

All the materials are already described in **Materials** section of **Chapter 4**. LovD and LovD-Mut1-12 genes were synthetised and cloned into expression vector pET28b by GenScript Gene Synthesis service (Piscataway, NJ, USA).

Table 1. Protein sequences of LovD variants.

Variant	Sequence (Inserted mutation are highlighted in bold red)
LovD	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-29M1	MGS**D**IDAA**R**AADPVVLMETAFRKAV**L**S**G**QIPGAV**C**MARDASGNL**T**YTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLV**T**LDETVDRLLPDL**H**AMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**L**AQGHLQ**N**AEKFGIQ**DR**TAPP**M**VNDPGAEWIYGA**QI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**R**THRNS**S**DG**S**LRYDD**R**VYFR**H**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**D**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGEN**K** RRKGSLTFGGGPNIVWQIDPKAGLCTL**L**FFQLEPWNDPVCRDLTRTFE**E**AIYAQYQQG
LovD-29M2	MGS**S**IDAA**K**AADPVVLMETAFRKAV**A**S**G**QIPGAV**Q**MARDASGNL**T**YTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLV**H**LDETVDRLLPDL**K**AMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**L**AQGHLQ**Y**AEKFGIQ**WR**DAPP**M**VNDPGAEWIYGA**AS**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**S**THRNS**DDG**S**LRYDD**E**VYFR**R**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**K**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGEN**A** RRKGS**M**TFGGGPNI**W**WQIDPKAGLCTL**S**FFQLEPW**G**DPVCRDLTRTFE**R**AIYAQYQQG
LovD-29M3	MGS**W**IDAA**N**AADPVVLMETAFRKAV**W**S**G**QIPGAVIMARDASGNL**S**YTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLV**H**LDETVDRLLPDL**H**AMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**K**AQGHLQ**D**AEKFGIQ**NR**TAPP**L**VNDPGAEWIYGA**QM**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**D**THRNS**E**DG**S**LRYDDSVYFR**Y**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**K**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGEN**Y** RRKGS**M**TFGGGPNIVWQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFE**L**AIYAQYQQG
LovD-29M4	MGS**W**IDAA**Y**AADPVVLMETAFRKAVKS**G**QIPGAVIMARDASGNL**T**YTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLV**H**LDETVDRLLPDL**M**AMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREY**R**AQGHLQ**W**AEKFGIQ**DR**TAPP**L**VNDPGAEWIYGA**QI**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRAD**D**THRNS**DDG**S**LRYDDSVYFR**H**DGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQP**D**TVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGEN**R** RRKGSLTFGGGPNIVWQIDPKAGLCTL**V**FFQLEPW**G**DPVCRDLTRTFE**E**AIYAQYQQG
LovD-BU1	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**E**VNDPGAEWIYGAN**H**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG

LovD-BU2	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**L**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**M**VNDPGAEWIYGA**GM**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**W**WQIDPKAGLCTL**S**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-BU3	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**V**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**F**VNDPGAEWIYGA**AM**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFS**N**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNI**W**WQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-BU4	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**F**VNDPGAEWIYGAN**H**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**M**TFGGGPNI**W**WQIDPKAGLCTL**S**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-BU5	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**M**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**Q**VNDPGAEWIYGA**AM**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGS**R**TFGGGPNI**W**WQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-BU6	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**L**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**E**VNDPGAEWIYGAN**M**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-BU7	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAV**L**MARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**E**VNDPGAEWIYGAN**M**DWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFS**S**PGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNI**F**WQIDPKAGLCTL**V**FFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-N43R	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASG**R**LNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-D119S	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFD**S**AGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-A123P	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD- D119S/ A123P	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFD**S**AGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD- N43T/	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASG**T**LNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFD**S**AGN**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE

D119S/ A123P	NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-S164G	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQ**G**AEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-S172N	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQ**N**RLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-Y279M	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGS**M**MKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-F363L	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**L**GGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-F363W	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**W**GGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-Y279M/ F363W	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGS**M**MKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**W**GGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-Y279M/ Y327F/ F363W	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGS**M**MKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHIN**F**GGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**W**GGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-Y279M/ Y327F/ F363L	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGS**M**MKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHIN**F**GGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**L**GGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-Y279M/ Y327F/ F363Y	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGS**M**MKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHIN**F**GGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**Y**GGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-F384V	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLA**V**FQLEPWNDPVCRDLTRTFEHAIYAQYQQG

LovD-MM1	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTI**A**ALQC**I**ERGLVDLDETVDRLLPDLSAMPVLEGFDDAGN**P**RLR**D**RRGKITLRHLLTH**S**S GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPP**L**VNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**L**GGGPNIVWQIDPKAGLCTLAFFQLEPW**G**DPVCRDLTRTFEHAIYAQYQQG
LovD-MM2	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASG**T**LNYTRCFGARTVRRDENNQLPPLQVDTP CRLASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFD**SNK**N**P**RLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQS**Q**EKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NIC**K**PLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLTFGGGPNIVWQIDPKAGLCTLAFFQLEPWNDPVCRDLTRTFEHAIYAQYQQG
LovD-MM3	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP C**L**LASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGGQGV**V**SGPGS**M**KVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHINYGGPMPMVLRRSFGLGGIIALEDLDGENW **Q**RKGSLT**L**GGGPNIVWQIDPKAGLCTLA**V**FQLEP**P**NDPVCRDLTRTFEHAIYAQYQQG
LovD-MM4	MGSIIDAAAAADPVVLMETAFRKAVKSRQIPGAVIMARDASGNLNYTRCFGARTVRRDENNQLPPLQVDTP C**W**LASATKLLTTIMALQCMERGLVDLDETVDRLLPDLSAMPVLEGFDDAGNARLRERRGKITLRHLLTHTS GLSYVFLHPLLREYMAQGHLQSAEKFGIQSRLAPPAVNDPGAEWIYGANLDWAGKLVERATGLDLEQYLQE NICAPLGITDMTFKLQQRPDMLARRADQTHRNSADGRLRYDDSVYFRADGEECFGG**A**GVFSGPGSYMKVLH SLLKRDGLLLQPQTVDLMFQPALEPRLEEQMNQHMDASPHIN**L**GGPMPMVLRRSFGLGGIIALEDLDGENW RRKGSLT**L**GGGPNIVWQIDPKAGLCTLA**V**FQLEP**P**NDPVCRDLTRTFEHAIYAQYQQG

Table 2. Strains, plasmids and oligonucleotides used in this study.

Plasmids	Relevant genetic characteristics
pET28b *lovD-29M1*	*lovD-29M1* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-29M2*	*lovD-29M2* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-29M3*	*lovD-29M3* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-29M4*	*lovD-29M4* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BU1*	*lovD-BU1* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BU2*	*lovD-BU2* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BU3*	*lovD-BU3* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BU4*	*lovD-BU4* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BU5*	*lovD-BU5* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-BU6*	*lovD-BU6* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-MM1*	*lovD-MM1* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-MM2*	*lovD-MM2* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-MM3*	*lovD-MM3* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
pET28b *lovD-MM4*	*lovD-MM4* gene (1251 bp) from *in silico* design cloned in pET28b (*NdeI* and *XhoI*)
Strains	**Relevant phenotype**
E. coli DH5α	F⁻ φ80*lacZ*ΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ⁻ *thi*-1 *gyr*A96 *rel*A1
E.coli BL21	F⁻ *omp*T *hsd*S_B (r_B⁻, m_B⁻) *gal dcm ara*B::T7RNAP-*tet*A

Oligonucleotides	Sequences
N43T	AAC(N) → ACC(T) Forward: 5' CGTGACGCGAGCGGCACCCTGAACTACACCCGTTG 3' Reverse: 5' CAACGGGTGTAGTTCAGGGTGCCGCTCGCGTCACG 3'
N43R	AAC(N) → CGC(R) Forward: 5' CGTGACGCGAGCGGCCGCCTGAACTACACCCGTTG 3' Reverse: 5' CAACGGGTGTAGTTCAGGCGGCCGCTCGCGTCACG 3'
D119S	GAT(D) → AGT(S) Forward: 5' GTGCTGGAGGGGTTTTGATAGTGCGGGTAACGCGCGTCTG 3' Reverse: 5' CAGACGCGCGTTACCCGCACTATCAAAACCCTCCAGCAC 3'
A123P	GCG(A) → CCG(P)

	Forward: 5' GATGATGCGGGTAACCCGCGTCTGCGTGAACG 3'
	Reverse: 5' CGTTCACGCAGACGCGGGTTACCCGCATCATC 3'
	GAT(D) → AGT(S); GCG(A) → CCG(P)
D119S+A123P	Forward: 5' GAGGGTTTTGATAGTGCGGGTAACCCGCGTCTGCGTGAACG 3'
	Reverse: 5' CGTTCACGCAGACGCGGGTTACCCGCACTATCAAAACCCTC 3'
	ACG(S) → GGC(G)
S164G	Forward: 5' AGGGCCACCTGCAAGGCGCGGAAAAGTTTG 3'
	Reverse: 5' CAAACTTTTCCGCGCCTTGCAGGTGGCCCT 3'
	AGC(S) → AAC(N)
S172N	Forward: 5' GGAAAAGTTTGGTATTCAGAACCGTCTGGCTCCGCCGGCGGTG 3'
	Reverse: 5' CACCGCCGGCGGAGCCAGACGGTTCTGAATACCAAACTTTTCC 3'
	TAC(Y) → ATG(M)
Y279M	Forward: 5' GTGTTTAGCGGCCCGGGTAGCATGATGAAGGTTCTGCACAGCCTGCTG 3'
	Reverse: 5' CAGCAGGCTGTGCAGAACCTTCATCATGCTACCCGGGCCGCTAAACAC 3'
	TAT(Y) → TTT(F)
Y327F	Forward: 5' CATGGACGCGAGCCCGCACATCAACTTTGGTGGCCCGATGCCG 3'
	Reverse: 5' CGGCATCGGGCCACCAAAGTTGATGTGCGGGCTCGCGTCCATG 3'
	TTC(F) → TGG(W)
F363W	Forward: 5' GTAAGGGTAGCCTGACCTGGGGTGGCGGTCCGAACATC 3'
	Reverse: 5' GATGTTCGGACCGCCACCCCAGGTCAGGCTACCCTTAC 3'
	TTC(F) → TTA(L)
F363L	Forward: 5' GGCGTCGTAAGGGTAGCCTGACCTTAGGTGGCGGTCCGAACATCG 3'
	Reverse: 5' CGATGTTCGGACCGCCACCTAAGGTCAGGCTACCCTTACGACGCC 3'
	TTC(F) → TAT(Y)
F363Y	Forward: 5' GGCGTCGTAAGGGTAGCCTGACCTATGGTGGCGGTCCGAACATCG 3'
	Reverse: 5' CGATGTTCGGACCGCCACCATAGGTCAGGCTACCCTTACGACGCC 3'
	TTC(F) → GTC(V)
F384V	Forward: 5' GCGGGCCTGTGCACCCTGGCGGTCTTTCAACTGGAGCCGTGGAACG 3'
	Reverse: 5' CGTTCCACGGCTCCAGTTGAAAGACCGCCAGGGTGCACAGGCCCGC 3'

5.3.2. QuickChange site directed mutagenesis

QuickChange mutagesis was performed as described in the **Experimental Section** of **Chapter 4**.

5.3.3. Expression and purification of LovD variants

Expression and purification of LovD variants was performed as described in the **Experimental Section** of **Chapter 4**.

Figure 1. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for protein expression and purification of different LovD variants. SE: Soluble Extract. P: Pure enzyme. MW: Standard Molecular Weight Marker.

5.3.4. Thioesterase spectrophotometric assay

Thioesterase spectrophotometric assay was performed as described in the **Experimental Section** of **Chapter 4**.

5.3.5. SVA synthase and esterase activities characterization

SVA synthase and esterase activities were determined as described in the **Experimental Section** of **Chapter 4**.

5.3.6. Analysis of intrinsic protein fluorescence

70 μg of free enzyme eluted from AG-Co^{2+} with 50 mM HEPES and 250 mM imidazole were placed in a 96-well dark plate before and after incubation for 24 h at room temperature. The plate was irradiated at 280 nm with excitation and fluorescence emission spectra was recorded between 300 and 500 nm; bandwidths of 5 nm were used for both excitation and emission.

5.3.7. Thermal shift assay

Thermal shift of LovD variants was measured as described in the **Experimental Section** of **Chapter 4**.

5.4. Results and Discussion

5.4.1. Generation and screening of LovD variants through RosettaDesign/MD in predefined mutation sites

The first computational protocol tested in this study consisted on the generation of new LovD variants by selecting predefined mutation sites that corresponded to the same positions discovered previously by directed evolution.[252] We set up a protocol in RosettaDesign (see **Computational section**) to introduce amino acid variability and create two different lineages of new LovD variants: the first one displayed new combinations of mutations in the 29 positions of LovD9 (LovD-29Mx), and the second one exhibited mutations in the eight buried positions (LovD-BUx, nearest to the active site) of the directed evolution study (see **Chapter 4**).[252] We generated 30 variants for each LovD lineage (**Table 3**). In that way, we intended on one hand to explore the effect of a thorough modification of the enzyme backbone through the modification of already known key positions, and on the other hand to assess the impact of mutations close to the catalytic triad without altering the stability of the solvent-exposed regions.

As mentioned before, the average distance between the catalytic residues is a key parameter to improve the catalytic efficiency of the enzyme, and the major structural motion that delineates catalytic configuration is the conformational shift of Tyr188 sidechain from the catalytic pocket (catalytic) to that natively occupied by Tyr327 (non-catalytic) (**Figure 2.a,b**).[252] We therefore calculated the average distances between all the catalytic residues along μs MD simulations and used them as a selection parameter to identify potentially enhanced LovD variants (**Table 4**). The selection criteria established that the three catalytic

distances (S76-K79, S76-Y188 and K79-Y188) should display an average distance under 4 Å to consider the variant as a candidate for *in vitro* expression and activity screening.

Table 3. Mutated positions (in bold red) of the LovD variants generated with RosettaDesign.

Position	4	9	26	28	35	43	96	109	123	157	164	172	174	178	191	192	241	247	250	256	261	275	297	355	361	370	383	391	404
LovD WT	I	A	K	R	I	N	D	S	A	M	S	S	L	A	N	L	Q	A	R	S	A	G	Q	W	L	V	A	N	H
1	W	N	W	G	I	S	H	R	A	K	D	N	T	Q	Q	M	Q	E	S	S	S	S	K	R	L	V	L	N	D
2	D	R	L	G	C	T	T	H	A	L	N	D	T	M	Q	I	R	S	S	R	H	S	D	K	L	V	L	N	E
3	D	R	L	G	V	S	H	A	A	R	W	D	T	M	Q	M	T	S	S	G	F	S	S	R	M	V	M	N	S
4	D	R	K	G	L	S	R	S	A	K	W	N	T	M	Q	M	T	S	S	G	N	S	K	R	M	T	V	N	Q
5	D	R	W	G	I	T	H	S	A	L	W	D	T	H	Q	M	S	D	S	G	N	S	R	T	M	T	V	N	L
6	D	R	L	G	I	T	Q	A	A	Y	E	D	T	M	Q	A	R	D	S	N	N	S	T	L	M	V	V	D	E
7	S	T	E	G	Q	Q	L	G	A	R	E	C	Y	M	E	A	N	K	S	R	Y	S	K	Q	M	A	S	G	R
8	D	R	E	G	Q	T	R	A	G	L	R	W	S	M	E	A	S	T	T	W	N	S	D	R	M	H	S	G	E
9	D	L	K	G	Q	T	R	A	A	E	N	W	H	M	E	A	N	S	S	Y	A	S	K	L	M	Y	S	G	E
10	S	K	A	G	Q	T	H	K	A	L	Y	W	D	M	A	S	S	D	S	E	R	S	K	A	M	W	S	G	R
11	W	N	W	G	I	S	H	H	A	K	D	N	T	L	Q	M	D	E	S	S	Y	S	K	Y	M	V	V	N	L
12	S	T	L	G	I	T	Q	S	A	K	T	D	T	L	Q	M	N	S	S	D	H	S	T	R	M	V	L	N	E
13	D	K	K	G	L	T	R	A	A	R	W	Q	R	L	Q	M	S	T	S	D	C	S	D	R	M	V	V	N	E
14	D	R	W	G	L	T	R	A	A	R	W	K	L	L	Q	M	S	S	S	D	R	S	D	R	M	V	A	N	Q
15	D	K	E	G	L	S	R	G	A	R	W	D	T	L	Q	Q	R	D	S	D	R	S	R	R	M	N	N	N	Q
16	D	S	E	G	I	S	Q	G	A	R	W	D	T	L	T	R	Q	S	S	S	R	S	K	Q	M	Q	L	N	L
17	E	K	A	G	V	T	H	G	A	R	W	D	R	L	N	R	L	S	S	H	Y	S	R	R	M	V	L	N	E
18	E	R	E	G	I	T	R	G	A	R	W	D	T	Q	P	A	N	S	E	T	R	S	R	R	M	Y	L	G	E
19	D	K	A	G	V	Q	G	G	A	R	T	D	R	Q	P	A	Q	T	S	Y	Y	S	R	K	M	Y	L	N	D
20	E	K	E	G	V	T	H	G	A	R	R	K	D	L	P	A	S	D	S	R	K	S	K	G	R	Y	L	H	L
21	W	L	W	N	V	S	H	H	A	K	D	N	T	L	Q	M	Q	S	S	D	S	S	K	H	L	V	L	N	W
22	W	T	L	G	V	T	L	M	A	R	W	D	T	L	Q	M	D	S	S	T	H	S	T	K	L	V	M	G	D
23	W	Y	K	G	I	T	H	M	P	R	W	D	T	L	Q	I	D	D	S	S	H	S	D	R	L	V	V	G	E
24	D	L	L	G	I	T	R	S	P	Y	W	D	R	L	Q	M	D	D	S	S	Y	S	R	K	M	A	V	G	E
25	D	H	W	G	I	T	H	S	P	Y	W	D	R	F	Q	D	T	N	S	S	Y	S	K	R	L	T	V	G	D
26	D	K	L	G	L	S	H	A	P	R	W	L	R	F	Q	T	V	T	S	N	Y	S	R	G	M	Y	V	G	L
27	W	T	L	G	L	T	H	G	A	R	W	D	R	L	S	M	V	S	S	A	T	S	D	G	M	W	V	G	N
28	W	R	L	G	T	T	V	G	Q	R	W	D	T	Q	H	H	T	S	S	Q	L	S	D	G	M	Y	L	G	R
29	S	R	E	G	C	S	R	G	P	R	W	D	T	Q	H	H	T	T	S	N	E	S	E	G	M	Y	L	G	W
30	S	K	L	G	V	T	Q	A	P	W	W	D	R	Q	N	A	I	T	S	N	E	S	D	G	M	Y	L	G	E
31	I	A	K	R	I	N	D	S	A	M	S	S	L	E	N	H	Q	A	R	S	A	S	Q	W	L	V	V	N	H
32	I	A	K	R	L	N	D	S	A	M	S	S	L	M	N	M	Q	A	R	S	A	S	Q	W	L	R	V	N	H
33	I	A	K	R	L	N	D	S	A	M	S	S	L	M	N	M	Q	A	R	S	A	S	Q	W	M	Y	A	N	H
34	I	A	K	R	L	N	D	S	A	M	S	S	L	M	A	M	Q	A	R	S	A	S	Q	W	M	W	S	N	H
35	I	A	K	R	L	N	D	S	A	M	S	S	L	M	G	M	Q	A	R	S	A	S	Q	W	M	W	S	N	H
36	I	A	K	R	L	N	D	S	A	M	S	S	L	M	A	A	Q	A	R	S	A	S	Q	W	W	W	V	N	H
37	I	A	K	R	L	N	D	S	A	M	S	S	L	M	A	A	Q	A	R	S	A	A	Q	W	W	W	V	N	H
38	I	A	K	R	L	N	D	S	A	M	S	S	L	F	A	M	Q	A	R	S	A	N	Q	W	L	W	A	N	H
39	I	A	K	R	V	N	D	S	A	M	S	S	L	F	A	M	Q	A	R	S	A	N	Q	W	L	W	A	N	H
40	I	A	K	R	I	N	D	S	A	M	S	S	L	F	N	H	Q	A	R	S	A	S	Q	W	M	W	S	N	H
41	I	A	K	R	I	N	D	S	A	M	S	S	L	E	N	H	Q	A	R	S	A	S	Q	W	L	V	V	N	H
42	I	A	K	R	L	N	D	S	A	M	S	S	L	Q	N	M	Q	A	R	S	A	S	Q	W	M	V	V	N	H
43	I	A	K	R	L	N	D	S	A	M	S	S	L	Q	A	M	Q	A	R	S	A	S	Q	W	L	V	V	N	H
44	I	A	K	R	I	N	D	S	A	M	S	S	L	Q	A	M	Q	A	R	S	A	S	Q	W	M	V	V	N	H
45	I	A	K	R	I	N	D	S	A	M	S	S	L	Q	A	M	Q	A	R	S	A	S	Q	W	M	F	V	N	H
46	I	A	K	R	L	N	D	S	A	M	S	S	L	M	A	M	Q	A	R	S	A	S	Q	W	L	F	V	N	H
47	I	A	K	R	L	N	D	S	A	M	S	S	L	M	A	N	Q	A	R	S	A	S	Q	W	R	F	V	N	H
48	I	A	K	R	L	N	D	S	A	M	S	S	L	M	A	M	Q	A	R	S	A	S	Q	W	R	W	V	N	H
49	I	A	K	R	M	N	D	S	A	M	S	S	L	Q	A	M	Q	A	R	S	A	S	Q	W	R	W	V	N	H
50	I	A	K	R	M	N	D	S	A	M	S	S	L	Q	A	M	Q	A	R	S	A	S	Q	W	R	W	V	N	H
51	I	A	K	R	L	N	D	S	A	M	S	S	L	E	N	H	Q	A	R	S	A	S	Q	W	L	V	V	N	H
52	I	A	K	R	L	N	D	S	A	M	S	S	L	E	N	M	Q	A	R	S	A	S	Q	W	L	V	V	N	H
53	I	A	K	R	L	N	D	S	A	M	S	S	L	E	N	D	Q	A	R	S	A	S	Q	W	L	H	V	N	H
54	I	A	K	R	L	N	D	S	A	M	S	S	L	E	N	M	Q	A	R	S	A	S	Q	W	L	F	V	N	H
55	I	A	K	R	M	N	D	S	A	M	S	S	L	E	A	M	Q	A	R	S	A	S	Q	W	L	F	V	N	H
56	I	A	K	R	I	N	D	S	A	M	S	S	L	E	Q	R	Q	A	R	S	A	S	Q	W	L	F	C	N	H
57	I	A	K	R	V	N	D	S	A	M	S	S	L	E	Q	R	Q	A	R	S	A	S	Q	W	L	W	V	N	H
58	I	A	K	R	V	N	D	S	A	M	S	S	L	E	E	R	Q	A	R	S	A	S	Q	W	L	W	V	N	H
59	I	A	K	R	V	N	D	S	A	M	S	S	L	E	A	R	Q	A	R	S	A	T	Q	W	L	W	V	N	H
60	I	A	K	R	V	N	D	S	A	M	S	S	L	E	A	T	Q	A	R	S	A	S	Q	W	L	W	V	N	H

Table 4. Average catalytic distances of unrestrained 1 μs MD simulations of different LovD variants.

LovD variant	S76-K79 (Å)	S76-Y188 (Å)	K79-Y188 (Å)	Selection
1	3.42 (±1.01)	4.04 (±1.10)	3.37 (±0.45)	NO
2 (LovD-29M1)	3.10 (±0.58)	3.34 (±0.48)	3.48 (±0.51)	YES
3	4.02 (±1.17)	7.20 (±1.09)	7.40 (±0.96)	NO
4	5.94 (±1.00)	5.44 (±1.08)	5.57 (±1.11)	NO
5	3.90 (±1.10)	7.51 (±1.25)	7.32 (±1.03)	NO
6	3.31 (±0.85)	7.11 (±1.02)	7.73 (±0.95)	NO
7	9.07 (±2.13)	6.70 (±2.40)	7.95 (±1.96)	NO
8	7.32 (±1.11)	7.50 (±1.50)	7.37 (±1.05)	NO
9	5.95 (±0.74)	5.05 (±0.87)	5.04 (±0.51)	NO
10 (LovD-29M2)	5.58 (±1.47)	4.09 (±0.73)	4.30 (±0.68)	YES
11 (LovD-29M3)	3.12 (±0.59)	3.52 (±0.70)	3.33 (±0.66)	YES
12	4.02 (±1.11)	7.98 (±3.33)	7.40 (±3.11)	NO
13	4.69 (±1.18)	7.87 (±1.10)	7.56 (±0.96)	NO
14	3.44 (±0.66)	7.21 (±1.31)	7.29 (±1.19)	NO
15	5.39 (±0.78)	8.04 (±1.32)	7.68 (±0.82)	NO
16	4.41 (±1.08)	6.46 (±2.47)	5.46 (±2.51)	NO
17	5.95 (±1.04)	7.03 (±1.29)	7.24 (±1.93)	NO
18	5.85 (±0.96)	6.40 (±1.06)	6.38 (±0.58)	NO
19	6.15 (±1.01)	5.97 (±0.94)	6.34 (±0.64)	NO
20	4.64 (±1.69)	4.18 (±1.01)	4.25 (±0.67)	NO
21	4.16 (±1.22)	7.01 (±1.89)	6.82 (±1.71)	NO
22	4.75 (±1.07)	5.87 (±1.47)	4.54 (±1.32)	NO
23 (LovD-29M4)	3.07 (±0.58)	3.21 (±0.36)	3.37 (±0.38)	YES
24	4.39 (±1.20)	7.60 (±1.29)	7.37 (±1.16)	NO
25	5.07 (±0.78)	9.07 (±1.05)	7.30 (±1.07)	NO
26	4.88 (±0.79)	7.66 (±1.62)	5.28 (±1.28)	NO
27	4.78 (±1.92)	5.60 (±2.43)	3.43 (±0.58)	NO
28	5.35 (±1.56)	9.22 (±2.30)	8.11 (±2.18)	NO
29	5.86 (±1.33)	7.50 (±1.75)	4.93 (±1.99)	NO
30	5.35 (±0.93)	6.05 (±1.15)	4.50 (±1.22)	NO
31 (LovD-BU1)	3.02 (±0.51)	3.32 (±0.46)	3.39 (±0.42)	YES
32	4.99 (±1.24)	6.42 (±1.71)	3.73 (±0.92)	NO
33	4.95 (±0.76)	4.32 (±1.05)	5.20 (±0.94)	NO
34	5.84 (±0.90)	5.92 (±0.86)	5.77 (±0.55)	NO
35 (LovD-BU2)	2.96 (±0.28)	3.28 (±0.45)	3.38 (±0.41)	YES
36	5.68 (±0.86)	7.36 (±1.51)	6.12 (±0.59)	NO
37	4.11 (±1.26)	4.42 (±1.50)	4.25 (±1.27)	NO
38	4.41 (±1.05)	9.36 (±2.11)	7.93 (±1.46)	NO
39 (LovD-BU3)	2.99 (±0.63)	3.48 (±0.76)	3.37 (±0.32)	YES

40 (LovD-BU4)	2.93 (±0.28)	3.42 (±0.41)	3.43 (±0.36)	YES
41	3.02 (±0.51)	3.32 (±0.46)	3.39 (±0.42)	NO
42	5.87 (±0.75)	6.13 (±0.92)	6.22 (±1.15)	NO
43	3.53 (±1.04)	4.57 (±1.86)	4.46 (±1.65)	NO
44	8.23 (±2.11)	8.56 (±1.26)	8.74 (±1.37)	NO
45	4.51 (±1.55)	4.67 (±1.56)	4.59 (±1.32)	NO
46	7.05 (±1.32)	7.01 (±1.47)	7.11 (±1.25)	NO
47	3.31 (±0.92)	4.53 (±1.37)	3.94 (±1.08)	NO
48	3.44 (±1.10)	4.14 (±1.18)	3.76 (±0.77)	NO
49	4.64 (±1.43)	3.88 (±0.78)	4.09 (±0.71)	NO
50	6.87 (±1.25)	6.32 (±1.55)	6.54 (±1.15)	NO
51	3.05 (±0.43)	5.62 (±2.53)	5.75 (±2.77)	NO
52 (LovD-BU5)	2.96 (±0.27)	3.27 (±0.32)	3.31 (±0.30)	YES
53	5.70 (±1.59)	8.79 (±1.73)	7.24 (±1.40)	NO
54 (LovD-BU6)	3.37 (±1.01)	3.65 (±0.77)	3.36 (±0.51)	YES
55	5.09 (±1.13)	7.57 (±1.50)	7.03 (±1.43)	NO
56	4.24 (±1.07)	7.63 (±0.97)	6.93 (±1.75)	NO
57	4.73 (±1.07)	7.79 (±3.75)	7.18 (±3.22)	NO
58	5.32 (±1.20)	4.29 (±0.82)	4.22 (±0.60)	NO
59	3.73 (±1.15)	6.74 (±1.88)	6.84 (±2.22)	NO
60	5.79 (±0.68)	8.44 (±0.96)	7.55 (±0.94)	NO

As shown in **Table 4** and **Figure 2.c,d**, 10 different variants (4 from the 29M line and 6 from the BU line) fulfilled the computational selection conditions and were chosen for experimental validation. Out of the 10 variants, 8 were expressed in a soluble form and 2 turned to form inclusion bodies (LovD-29M2 and LovD-BU3) (**Figure 1**). **Figure 2.e** shows thioesterase (TE) activity of each variant, which determines the capability of the engineered enzyme to accept the acyl thioester surrogate α-dimethylbutyryl-*S*-methyl mercaptopropanoate (DMB-SMMP) (see **Chapter 4**). TE activity is reduced to 50% or less for LovD-29M designs when compared to the wild-type. On the contrary, LovD-BU designs increased TE catalytic activity between 3-fold for the less active variant (LovD-BU5) and 7-fold for the most active variant (LovD-BU4) when compared to the wild-type. This enhancement of DMB-SMMP conversion suggested that when targeting these buried positions, the RosettaDesign/MD protocol was successful to reshape a catalytic pocket that optimizes the first step of the reaction, facilitating the acyl-enzyme complex formation in agreement with the results presented in Chapter 4. However, when the entire reaction was screened through SVA analytic detection, none of the LovD-29M nor LovD-BU designs displayed SVA synthetic activity (**Figure 2.f**). The reason for this could be that the simultaneous introduction of a large number of mutations by RosettaDesign could compromise the structural integrity of the protein. This is particularly relevant for the densely mutated LovD-29M variants. LovD-BU designs likely stabilized the active site to better bind DMB-SMMP, although perhaps at the cost of affecting protein stability and/or productive binding of the second substrate, monacolin J acid (MJA), hampering SVA synthesis.
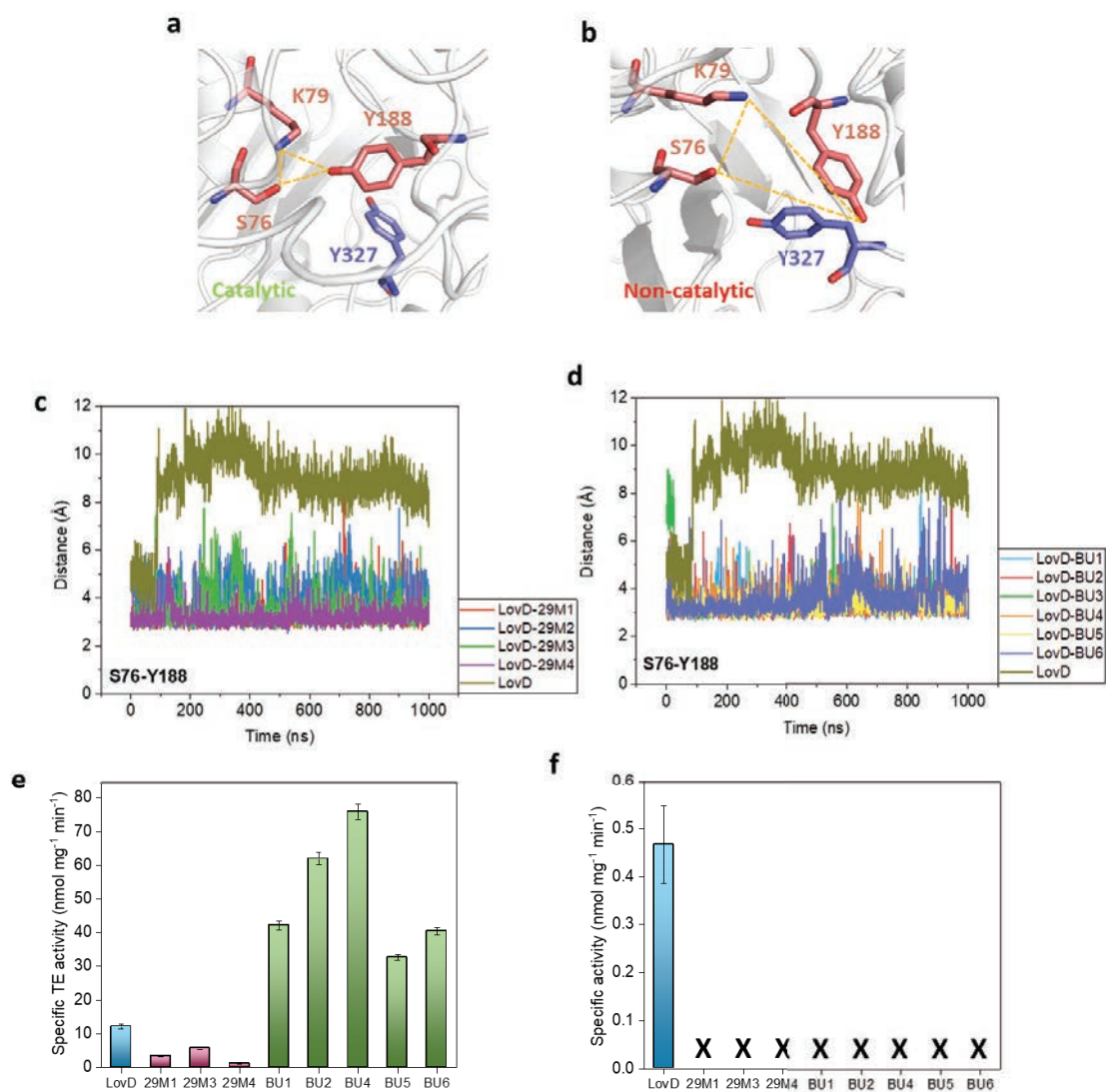
Figure 2. Catalytic **(a)** and non-catalytic **(b)** configuration of the active site. 1 μs unrestricted Molecular dynamics (MD) simulations of LovD wild-type, LovD-29Mx variants **(c)** and LovD-BUx variants **(d)**. **(e)** Thioesterase (TE) specific activity of LovD (in blue), LovD-29Mx variants (in red) and LovD-BUx variants (in green). **(f)** Specific SVA synthase activity of LovD (in blue), LovD-29Mx and LovD-BUx variants. X represents zero SVA productivity.

112

5.4.2. Allosteric network analysis for the identification of new mutational hotspots in LovD

Allosteric effects often occur in proteins between remote regions and the catalytic site. Normally, the catalytic site arrangement is indirectly modified through the binding of a molecule (activator or inhibitor) in an allosteric site, which propagates a series of dynamic motions that finally affect catalytic residues. This dynamic communication between residues can also be modified via mutation of key amino acids in allosteric positions. Significant backbone deformations are not required for an allosteric effect, which can also take place through subtle conformational changes. Within the Weighted Implementation of Suboptimal Paths (WISP) framework, the shortest (i.e. more intense) path maps determine the allosteric relevance of relevant positions: an optimal pathway is the shortest distance between two nodes (residues) along weighted edges.

In our allosteric dynamic network analysis of LovD, we used WISP on 200 ns MD simulations to identify new hotspots of mutations. These hotspots were identified by the construction of a shortest path map for two forms of the wild-type protein, corresponding to the catalytic and the non-catalytic configuration of the active site previously described. In this way, the nodes with a high variance between catalytic and non-catalytic shortest maps are presumably more correlated with the configuration of the catalytic triad. As a result, the protein structure was color-coded to indicate the difference in path lengths from the catalytic and the non-catalytic forms (**Figure 3.a**) and two hotspot regions in external loops near the channel entrance of the substrate were identified: residues 42-43 (loop-1) and 117-123 (loop-2). Amongst these positions, we selected 43 in loop-1 and 119 and 123 in loop-2, as two of them (43 and 123) coincide with those selected by directed evolution and the third one (119) is the most exposed residue of loop-2.

Starting from these positions as targets for mutation, RosettaDesign predicted the following substitutions: N43T, D119S and A123P. Of note, A123P is a mutation introduced at the second round of directed evolution. On the other hand, position 43 is mutated from asparagine to arginine in the sixth round of directed evolution, while RosettaDesign predicted threonine as the best option for amino acid substitution. We therefore selected variants N43R and A123P to study their effect as single mutations in the directed evolution process, and D119S to test the prediction of the combined WISP/RosettaDesign protocol. Moreover, the synergy between those predicted mutations was assessed through the evaluation of the double D119S/A123P (for a local contribution) and triple N43T/D119S/A123P (for the synergic effect study of distant positions) variants. To complete the study, other single mutants derived from directed evolution and close to loop-2 S164G and S172N, were considered. These mutants were tested in order to study the relevance of changes in the environment of this predicted loop in enzyme activity.
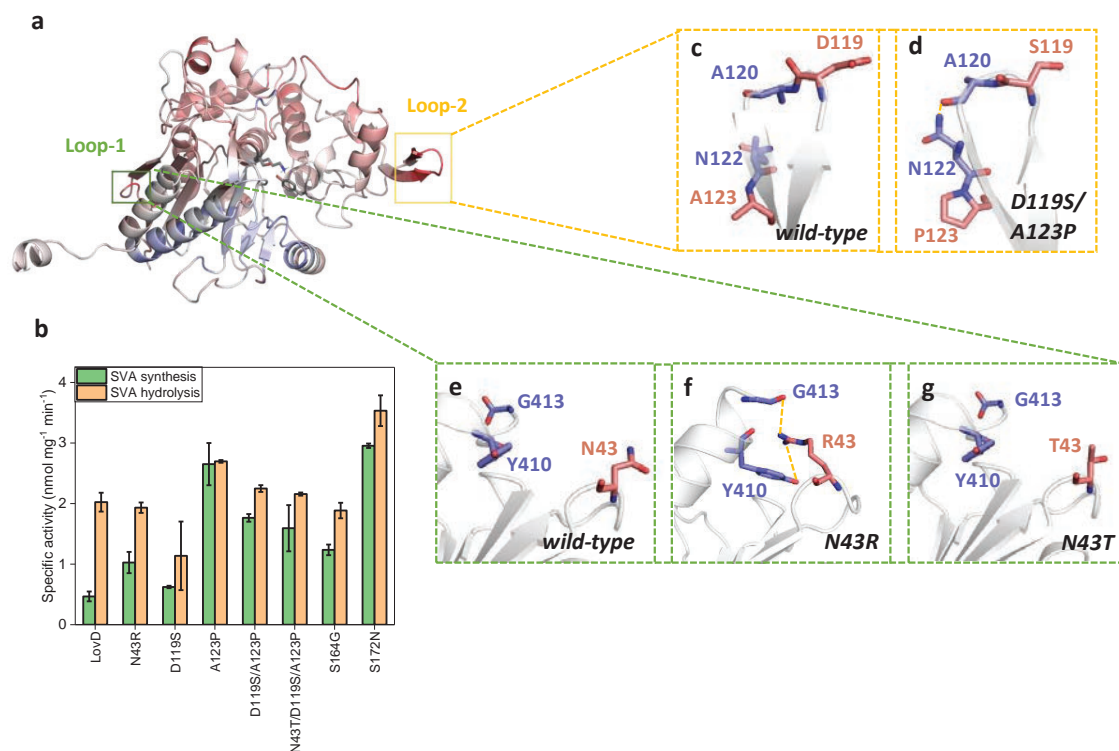
Figure 3. (a) LovD structure colored based on the difference in path lengths between the catalytic and non-catalytic conformations of the active site, calculated with WISP. All the calculated paths start from catalytic Tyr188. Red and blue gradients indicate a decrease and increase in the path lengths, respectively. (b) Specific SVA synthase and hydrolase activities for LovD wild-type and all the mutants designed trough WISP/RosettaDesign. Structure of loop-2 with the native residues in 119 and 123 (c) and the proposed mutations (d) highlighted as sticks. Configuration of loop-1 with native N43 (e), the directed evolution mutation N43R (f) and the variant proposed by RosettaDesign N43T (g) derived from the calculated models. The targeted positions are colored in red and other relevant residues surrounding those positions are colored in blue.

As showed in **Figure 3.b**, mutation N43R improves 2-fold wild-type SVA synthase activity while maintaining SVA hydrolytic activity at the same level. The change of asparagine (**Figure 3.e**) into arginine (**Figure 3.f**) in this position could form two new hydrogen bonds with Y410 and G413 (C-terminal residues) to potentially stabilize loop-1 and the C-terminus α-helix, as observed in the generated model. The designed mutation in loop-2 (D119S) did not provide a significant improvement in SVA synthesis, but decreased SVA hydrolysis around 2-fold when compared to the wild-type. A123P turned to be the most successful predicted substitution, with a 5-fold increase in SVA synthase activity and a much slighter increase in SVA hydrolase activity. The change of an alanine (**Figure 3.c**) by a proline (**Figure 3.d**) might stabilize loop-2 through the shift of N122 and the formation of a hydrogen bond with A120 as observed in the generated model. The combined variants D119S/A123P and N43T/D119S/A123P experienced a detrimental effect when compared with the A123P single mutant for both synthase and hydrolase activities, but were still superior to the synthetic capability of the wild-type (around 3.5-fold). Threonine in position 43 precludes the formation of a salt bridge with C-terminal G413, unlike the arginine in that same position

installed by directed evolution (N43R) (**Figure 3.g**). Finally, S164G and S172N exhibited a 1.5- and 6-fold improvement in SVA synthesis, respectively. Regarding SVA hydrolysis, S164G maintained wild-type level, while S172N displayed a 1.7-fold increase. Mutagenesis conformational stabilization of both loops through the creation of new hydrogen bonds could trigger a signaling network that finally attains the active site dynamics and alters the enzyme specific activity.

5.4.3. Prediction of mutational hotspots based on WISP betweenness centrality and homologue sequences analysis

In the WISP framework, betweenness centrality (BC) quantifies the frequency a given node is present in all the optimal and suboptimal calculated paths between all distinct pairs of nodes. For the top 20 residues with the highest BC values derived from the WISP analysis of LovD in its catalytic conformation (**Figure 4.a**), a filter was applied based on the percentage of conservation of these positions in the different transesterase homologues to LovD. Those positions with a conservation ratio of 80% or more were discarded as a potential mutation site. For the "mutable" residues with less than 80% conservation, three different criteria were used to generate new mutants:

1. Mutation sites selected from all mutable residues from the top 20 of the BC ranking: 21, 73, 186, 270, 273, 279, 327, 356, 363, 384 and 390.

2. Mutation sites selected from the first 3 mutable residues from the top 20 BC ranking: 279, 327 and 363.

3. Mutation sites selected from mutable residues from the top 20 of the BC ranking in which LovD has an amino acid different to that of the consensus sequence: 73, 270, 327, 363, 384 and 390.

Following design criterium 2, RosettaDesign mutations Y279M, F363W an F363L were frequently predicted. When position 327 was introduced as a designable position, Y279M was maintained as the preferred mutation, and Y327F F363Y were preferably predicted. To assess the differences between substitutions in position 363, we generated variants F363W and F363L. Moreover, to analyze the combined effect of the three residues with the highest BC score, we also set Y279M/Y327F/F363W, Y279M/Y327F/F363L and Y279M/Y327F/F363Y as candidates. Finally, the predicted impact of Y327F in the variance of 363 substitution was explored through the generation of the double mutant Y279M/F363W, and its comparison with Y279M/Y327F/F363W. Among the rest of the positions located by BC analysis, the mutation with a highest RosettaDesign score was F384V, so we also tested this single mutant. Following design criterium 3, the combination of RosettaDesign with BC study also revealed other possibly beneficial amino acid changes, such as R73L, R73W, Q270A, F273V, R356Q and W390P.

Finally, multiple sequence alignment followed by RosettaDesign suggested a final group of mutations: M85A, M90I, A123P, E127D, T141S, A178L and N391G. The evaluation of these relevant mutations was assessed through the production of four LovD multiple mutants (LovD-MM1-4) (**Table 1**).
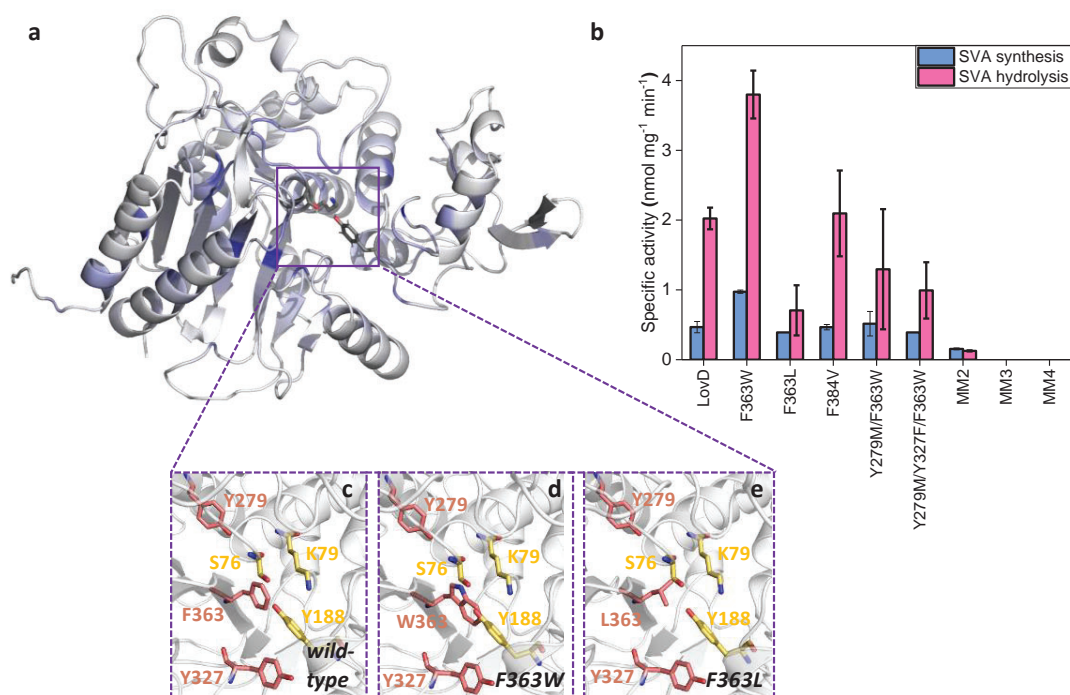

Figure 4. (a) LovD structure colored based on the WISP betweenness centrality values. Blue colored residues are those with higher centrality values. **(b)** Specific SVA synthase and hydrolase activities for LovD wild-type and all the mutants selected with centrality and multiple sequence alignment methods combined with RosettaDesign. Structure of the catalytic pocket with amino acid changes F **(c)**, W **(d)** and L **(e)** in position 363 derived from the calculated models.

From all the expressed variants, Y279M/Y327F/F363L and Y279M/Y327F/F363L were expressed as inclusion bodies in an insoluble form, while LovD-MM1 was underexpressed and the rest of the mutants were correctly overexpressed in soluble forms (**Figure 1**). As shown in **Figure 4.b**, F363W showed a 2-fold increase in both SVA synthetic and hydrolytic capabilities when compared to the wild-type. Among all the tested candidates, F363W showed the highest SVA hydrolase activity. Position 363 in LovD overlaps with position 344 in the lovastatin hydrolase from *Penicillium chrysogenum* (PcEST). The native tryptophan at this position in PcEST is involved in lovastatin hydrolysis through hydrogen bonding with the catalytic triad.[315] This similarity between protein structures and activities could justify the elevated SVA hydrolysis enhancement when F363 in LovD is mutated to a tryptophan. The rest of the variants did not experience a relevant improvement in SVA productivity, although F363L considerably reduced hydrolytic activity. Our hypothesis is that a leucine in position 363 precludes binding stacking interactions with the decalin core of either MJA or SVA, thus hampering both substrate acylation and product hydrolysis. Y279M/F363W is slightly more

active than Y279M/Y327F/F363W, which could justify the hypothesized detrimental activity of the combination of phenylalanine in position 327 with tryptophan in position 363. However, when the tryptophan in this position is substituted for the other two predicted amino acids (leucine and tyrosine) the protein presented solubility problems during expression. Finally, multiple mutants displayed much lower activities compared to the wild-type. Only LovD-MM2 showed residual synthase and hydrolase productivities, while LovD-MM3 and LovD-MM4 turned to be completely inactive.

5.4.4. Correlation of designed LovD variants SVA synthase activity with structural stability effects

Mutagenesis in specific sites can trigger large increases in enzyme activity, but also deleterious conformational changes altering protein structural integrity. These changes in protein structure can be measured through biophysical techniques. Thermal shift assay measures changes in denaturation temperature and hence stability of a protein through the non-specific binding of a fluorescent dye to hydrophobic surfaces of the protein (Thermofluor assay).[272] Protein unfolding can therefore be monitored when these hydrophobic regions are exposed to the solvent through an increase in fluorescence. Another powerful indicator of protein structure based on the same principle is the intrinsic fluorescence of its aromatic residues. When conformational changes occur, amino acids that were previously buried in the hydrophobic core of the protein are exposed to the solvent and their intrinsic fluorescence is shifted to longer emission wavelengths.

Thermofluor assay was carried out to determine the melting temperature (T_m, the midpoint value of the stability curve) of each LovD variant (**Figure 5.a**). Wild-type LovD exhibited a T_m of 39 °C. Most of the designs obtained with the first (RosettaDesign/MD) and third (WISP betweenness centrality/multiple sequence alignment/RosettaDesign) computational protocols displayed T_m values lower than the wild-type, which could be interpreted as a lower structural resistance to temperature changes and thereby an activity loss. On the contrary, all mutants obtained with the second computational protocol (WISP/RosettaDesign) presented T_m values similar or superior to that of LovD. Since this group of variants also showed the best activities, a correspondence between specific activity and T_m can be stablished, as observed for the directed evolution variants described in **Chapter 4**.

As protein precipitation was observed during the 24 hour reactions involving incubation at room temperature for many of the variants, intrinsic fluorescence intensity was measured before and after incubation (**Figure 5.b**) in order to assess how a long-term incubation under mild conditions could affect protein structural integrity. In agreement with the Thermofluor assay results, many mutants obtained with the first (RosettaDesign/MD) and third (WISP betweenness centrality/multiple sequence alignment/RosettaDesign) computational protocols experienced a considerable change in their intrinsic fluorescence after incubation, which indicates partial or total unfolding of the structure that could lead to protein aggregation. Other variants designed through the third method presented

fluorescence changes similar to the wild-type, so these mutations did not produce severe structural changes leading to protein precipitation. Finally, most of the mutants obtained with the second method (WISP/RosettaDesign) better preserved their protein fluorescence, suggesting that engineering flexible loops contributed to protein stabilization, which also had a beneficial effect on enzyme activities.
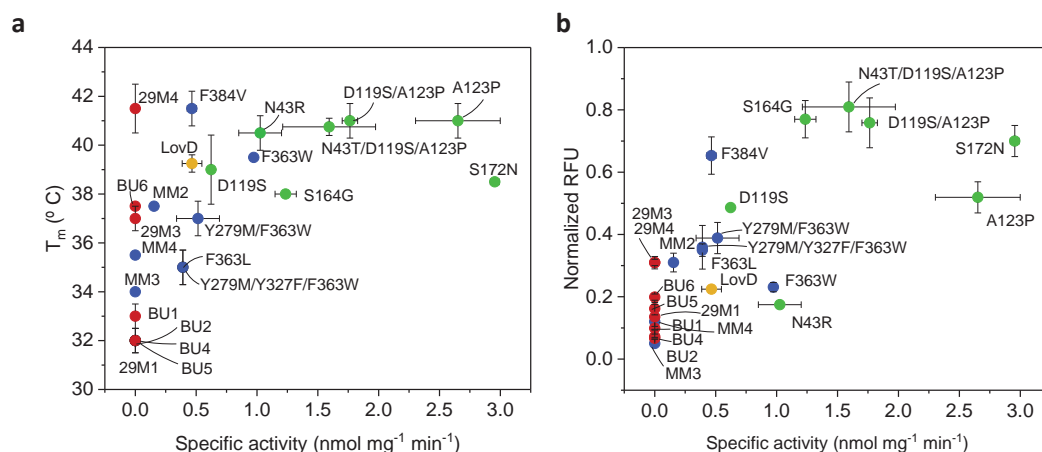

Figure 5. a. SVA specific activity vs T_m in computationally designed LovD variants. **b.** SVA specific activity vs normalized relative fluorescence units after 24-h incubation at room temperature (RFU; RFU=1 corresponds to the fluorescence of each LovD variant before 24-hour incubation at room temperature and 250 rpm), calculated using the equation $\lambda_{max}(t_{24h}) / \lambda_{max}(t_0)$. Mutants designed with the RosettaDesign/MD, WISP/RosettaDesign and WISP betweenness centrality/multiple sequence alignment/RosettaDesign methods are colored in red, green and blue, respectively; wild-type LovD is colored in yellow.

5.5.   Conclusions

Three different computational protein design methods based on the software RosettaDesign have been developed. Through the experimental evaluation of these variants, the effects of specific changes in certain positions in LovD on protein stability and their impact on enzyme activity, have been analyzed.

Mutagenesis on buried positions previously identified through directed evolution produced variants with an improved ability to process the thioester surrogate DMB-SMMP. However, such mutations also reduced protein stability and likely hampered MJA binding, yielding completely inactive variants for SVA synthesis. Mutations in the 29 directed evolution positions using unrestricted alphabets also failed to produce structurally stable LovD variants with improved enzyme activity.

Allosteric network analyses identified two mutational hotspots in flexible loops distal to the active site. Of note, some of these positions were already targeted by directed evolution. Changes in these loops and adjacent regions produced variants (N43R, A123P, S164G and S172N) with an enhanced activity and stability when compared to the wild-type.

Finally, WISP betweenness centrality combined with multiple sequence alignment revealed positions affecting the dynamics of distant residues. This strategy identified position 363 and its relevance in the hydrolytic capability of LovD.

Computational protein design is still a great challenge difficult to fulfill through one single approach, and with this study we have demonstrated that with the combination of different approaches we can identify relevant positions which modification could alter catalytic and structural properties of the enzyme.

CHAPTER 6:

Expanding LovD substrate scope for the biosynthesis of simvastatin analogues

6.1. Introduction

Simvastatin is a semisynthetic analogue of the fungal polyketide lovastatin, produced by *Aspergillus terreus*. Both compounds entail a big pharmaceutical value as they are inhibitors of the hydroxymethylglutaryl coenzyme A (HMG-CoA) reductase,[253] and are commercialized with the names of Mevacor/Altoprev/Altocor and Zocor/FloLipid, respectively. As detailed in **Chapter 2**, simvastatin can be obtained through the hydrolysis of lovastatin lactone to yield the triol monacolin J, followed by the selective silylation of the aliphatic chain alcohol, esterification of the decalin alcohol with 2,2-dimethylbutyryl chloride and final deprotection.[258] However, this chemical process is time- and cost-expensive, so more straightforward alternative pathways are desirable to obtain the blockbuster drug simvastatin.

In a previous research, Tang group explored the **substrate scope** of acyltransferase LovD, involved in the acylation of monacolin J acid (MJA) to yield lovastatin.[260] Synthetic surrogates, such as acyl-CoA, *N*-acetylcysteamine (SNAC) and methyl thioglycolate (SMTG) thioesters were used as acyl donors substituting the *S*-acyl-LovF native protein donor. All three types of thioester surrogates achieved high conversions (up to 92%), demonstrating that protein-protein interactions between LovD and LovF are not required for acyl transfer. However, these acyl donors displayed low k_{cat} values, and consequently MJA acted as a competitive inhibitor for the reaction. On the other hand, this study also unveiled the capability of LovD to accept other decalin cores different from MJA.

In a latter study, Tang group identified α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP) as a superior acyl donor that showed a 30-fold higher k_{cat} value than its previous analogues and a K_M value comparable to that of MJA[261] simvastatin synthesis. Then, as described in previous chapters, LovD was subjected to directed evolution to improve the enzyme catalytic turnover rate toward DMB-SMMP.[251,252] The optimized variant LovD9, with 29 mutations scattered throughout the protein structure, displayed a 120-fold higher k_{cat} than the wild-type. However, in **Chapter 4** we demonstrated that directed evolution not only enhanced the synthetic capability of LovD, but also reduced the inhibitory effect of MJA and the hydrolytic activity on SVA.

Other activated acyl donors that can be accepted by transesterases and acyltransferases have been described in literature. By far the most used acyl donors on a small scale are enol esters, such as **vinyl derivatives**.[316] After acyl transfer, the leaving enol tautomerizes to its keto form thus making the reaction irreversible.[317] Vinyl esters are commercially available and unexpensive. Moreover, they are often liquid at the reaction temperatures and used in large excess and even under solvent-free conditions. These surrogates are highly reactive, so they are often used as acylating agents in non-enzymatic reactions as well. In biocatalysis, they have been used in the lipase-catalyzed acylation of alcohols and the synthesis of terpenyl esters through transesterification,[318] and also in acyltransferases-catalyzed kinetic resolution of amines, cyanohydrins and alkynols,[231] amongst others.

On the other hand, *p*-nitrophenyl esters are also normally cheaper than classical acyl donors such as acyl-CoA derivatives, and they can be easily generated from thousands of commercially available carboxylic acids through a very efficient procedure using *p*-nitrophenyl chloroformate.[319,320] *p*-Nitrophenyl esters have been used as acyl donors for the lipase-catalyzed synthesis of anticancer drug temsirolimus from rapamycin[321] and the acyltransferase-catalyzed synthesis of ethyl acetate.[227] In this study we have selected both vinyl and *p*-nitrophenyl esters as potential substrates for the acyl-transfer reaction catalyzed by LovD.

The acyl acceptors scope might also be extended to other substrates with a similar structure to the decalin core of MJA. **Phenol** is an aromatic organic alcohol that is essential for production of disinfectant and surgical antiseptics manufacturing.[322] Moreover, phenol derivatives can also be utilized as starting materials to make plastics, explosives such as picric acid and drugs such as aspirin. **1-Naphthol** is structurally more similar of the decalin structure of MJA. This compound is the precursor of a variety of insecticides and pharmaceuticals, as well as for the antidepressant sertraline[323] and the antiprotozoal atovaquone.[324] Naphthol esters are interesting compounds that can be useful intermediates for the synthesis of organic compounds in general, and particularly in the field of the dyes and the perfumes. **(+)-Menthol** is a terpene alcohol obtained by thymol hydrogenation that is included in many products, such as decongestants, flavoring in foods, aftershave products, smoking tobacco, liqueurs, oral hygiene products or perfumery, amongst others.[325–327] (+)-Menthol presents cooling properties that are even more potent in some of its derivatives, such as the esters.[328] Esters of these cyclic alcohols could present a broad range of applications in chemical industry, thus being appropriate candidates to test LovD acyl acceptor substrate scope.

Hence, the goals of this study are:

1) broadening **the scope of acyl donors and acceptors** reactions catalyzed by LovD9, the most active evolved variant of acyltransferase LovD,

2) to measure the **kinetic parameters** for some of these reactions

3) to propose feasible **mechanisms for these transformations** through quantum mechanical (QM) calculations.

6.2. Experimental section

6.2.1. Materials

Common materials used in previous chapters are already described in the **Experimental section** of **Chapter 4**. Vinyl acetate (VA), vinyl propionate (VP), vinyl butyrate (VBU), vinyl benzoate (VBZ), ethyl 2-methyl butyrate (EMB), *p*-nitrophenyl butyrate (*p*NPB), N-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC), phenol (PH), (+)-menthol

(MT), 1-naphthol (NT) and yeast alcohol dehydrogenase (yADH) lyophilized powder were purchased from Sigma-Aldrich (St.Louis, IL, USA). Substrates *p*-nitrophenyl 2-methyl butyrate (*p*NPMB) and *p*-nitrophenyl 2,2-dimethyl butyrate (*p*NPDMB) were synthetized in our labs.

6.2.2. Expression and purification of LovD9

Expression and purification of LovD9 was performed as described in the **Experimental Section** of **Chapter 4**.

6.2.3. Synthesis of acyl donors

A 10 mL solution of *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC; 843 mg, 440 mM) was prepared in dichloromethane (DCM). Acyl donor (408 mg for (*S*)-2-methyl butyrate and 464 mg for 2,2-dimethyl butyrate) and HCl were slowly added at room temperature to the solution until a final concentration of 400 mM for the acyl donor was reached. The mixture was stirred for 10 minutes at room temperature and 250 rpm to obtain the corresponding *O*-acylisourea. A 10 mL solution of *p*-nitrophenol (*p*NP; 556 mg, 400 mM) was prepared in DCM. The *p*-nitrophenol solution was then added to the *O*-acylisourea solution at room temperature. 4-dimethylaminopyridine (DMAP) was added as a catalyst (24 mg) to a final concentration of 10 mM. The mixture was stirred overnight at 25 °C and 250 rpm. Reaction progress was monitored through thin-layer chromatography (TLC) under UV light (254 nm). Reaction by-products and unreacted *p*NP were removed by washing with distilled water (3 x 20 mL) and extracting the aqueous phase with hexanes (2 x 30 mL). This process was repeated until the aqueous phase was completely transparent and the yellow color from *p*NP had disappeared. By-products removal was monitored through TLC. Finally, the organic phase was evaporated at reduced pressure using a rotary evaporator. *p*-nitrophenyl (*S*)-2-methylbutyrate (*p*NPMB) and *p*-nitrophenyl 2,2-dimethylbutyrate (*p*NPDMB) were obtained (49% and 45% yield, respectively).

6.2.4. Characterization of acyl transfer reaction with different acyl donors and acceptors through UPLC/MS and GC/MS 24-hour reactions

24-Hour-reactions were carried out with 1 µM LovD9 in 50 mM HEPES, 10 mM MgCl$_2$ (pH 8) and 10 % DMSO, with different concentrations of the acyl donor (1, 10 and 100 mM) and MJA (1 and 3 mM). Reactions with non-MJA substrates (phenol, menthol and naphthol) were carried out with different concentrations of the acyl acceptor (1 and 10 mM) and the acyl donor (1 and 100 mM). Acyl donors tested in this study were ethyl 2-methyl butyrate (EMB), vinyl acetate (VA), vinyl propionate (VP), vinyl butyrate (VBU), vinyl benzoate (VBZ), *p*-nitrophenyl butyrate (*p*NPB), *p*-nitrophenyl (*S*)-2-methyl butyrate (*p*NPMB), *p*-nitrophenyl 2,2-dimethylbutyrate (*p*NPDMB) and α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP). Reaction samples were collected after 24 hours by passing them through a tangential ultrafiltration unit (Amicon Ultra centrifugal filters, 10 kDa). Reactions performed with MJA were analyzed by Ultra Performance Liquid Chromatography/Mass Spectrometry (UPLC/MS; see **Chapter 4**). Reactions performed with phenol, menthol and naphthol were

extracted with of ethyl acetate, dried with $MgSO_4$ and analyzed through gas chromatography using an Agilent Technologies GC System 7820A coupled to an inert XL Mass Selective Detector (MSD) with triple-axis detector 5975C.

6.2.5. Kinetic characterization of combined LovD9 thioesterase and acyltransferase activity with different acyl donors

Three different enzymatic assays were used to measure the kinetic parameters of LovD9-catalyzed ester conversion (combined esterase and transferase activities). For vinyl esters surrogates, acetaldehyde is generated as a by-product. The kinetics of the reaction were indirectly derived by measuring acetaldehyde conversion to ethanol by mixing 0.5 µM LovD9 with 1 µM yeast alcohol dehydrogenase (yADH) and 0.25 mM NADH.[329] NADH disappearance was monitored by UV-VIS spectroscopy at λ = 340 nm (**Figure 1.a**). Reactions were carried out with 1 mM MJA and different concentrations of VA (50-500 mM), VP (5-400 mM) or VBU (10-400 mM). For *p*NP esters surrogates, the reaction was directly monitored by detecting pNP release by UV-VIS spectroscopy at λ = at 405 nm. (**Figure 1.b**). Reactions were carried out with 1 µM LovD9, 1 mM MJA and different concentrations of *p*NPB, *p*NPMB (0.125-5 mM) or *p*NPDMB (1-7 mM). For DMB-SMMP, reactions were monitored by detecting SMMP though reaction with 4 mM 2-dithiodipyridine (2-DTDP) and subsequent 2-thiopyridone release by UV-VIS spectroscopy at λ = 323 nm (**Figure 1.c**).[271] Reactions were carried out with 1 µM LovD9, 1 mM MJA and different concentrations of DMB-SMMP (0.0625-3 mM). All the measurements were performed after 30 min reactions at 30 °C in 96-well plates in a BioTek Epoch2 spectrophotometer and kinetic data were fitted to the Michaelis-Menten model.
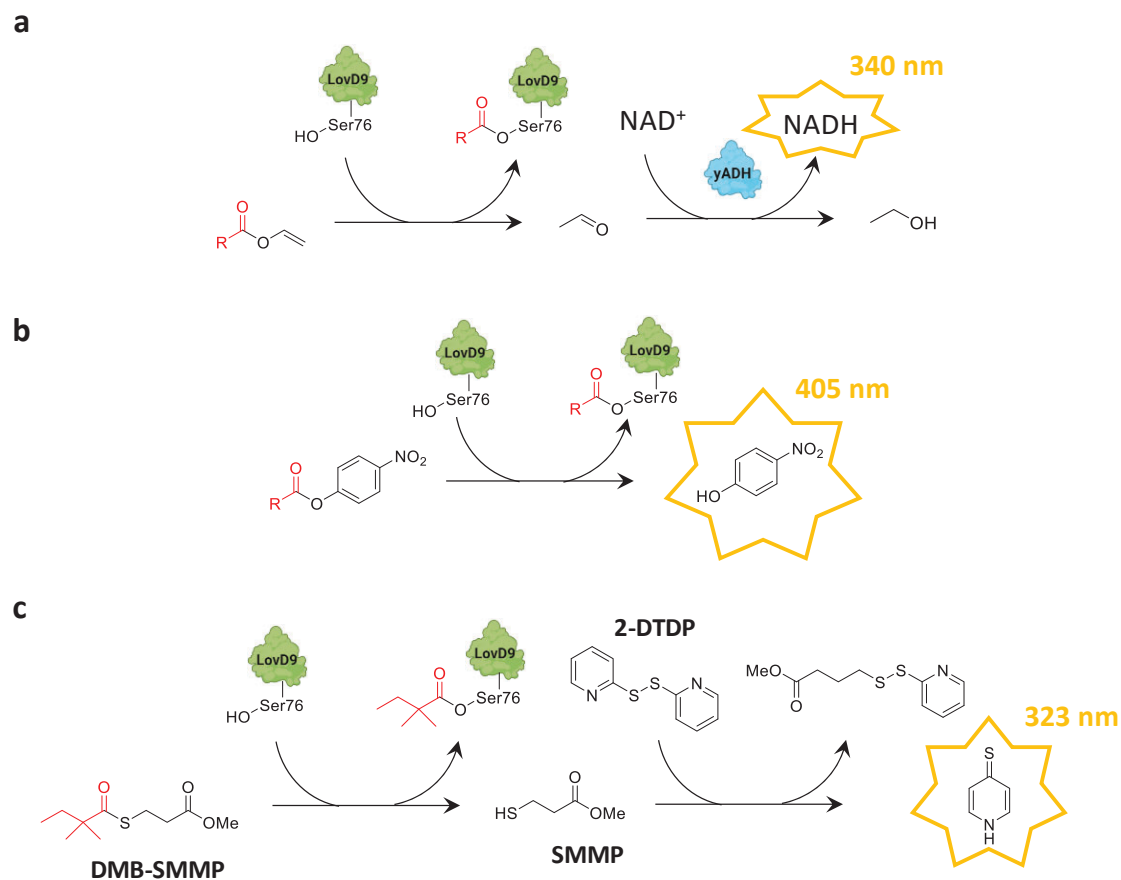
Figure 1. a. Enzymatic assay used for monitoring LovD9-catalyzed reaction of vinyl esters through coupling with yeast alcohol dehydrogenase (yADH) for the detection of NADH at λ = 340 nm. **b.** Enzymatic assay used for monitoring LovD9-catalyzed reaction of *p*-nitrophenyl esters through *p*-nitrophenol detection at λ = 405 nm. **c.** Enzymatic assay used for monitoring LovD9-catalyzed reaction of α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP) through 2-dithiodipyridine (2-DTDP) thiolysis and 2-thiopyridone detection at λ = 323 nm.

6.3. Computational details

6.3.1. Quantum mechanical calculations

Full geometry optimizations and transition structure (TS) searches were carried out with Gaussian 16[278] using the M06-2X hybrid functional[330] and 6-31+G(d,p) basis set with ultrafine integration grids. Bulk solvent effects in water were considered implicitly through the IEF-PCM polarizable continuum model.[331] The possibility of different conformations was taken into account for all structures. All stationary points were characterized by a frequency analysis performed at the same level used in the geometry optimizations from which thermal corrections were obtained at 298.15 K. The quasiharmonic approximation reported by Truhlar *et al.* was used to replace the harmonic oscillator approximation for the calculation of the vibrational contribution to enthalpy and entropy.[332] Scaled frequencies were not

considered. Mass-weighted intrinsic reaction coordinate (IRC) calculations were carried out by using the Gonzalez and Schlegel scheme[333] in order to ensure that the TSs indeed connected the appropriate reactants and products. Gibbs free energies (ΔG) were used for the discussion on the relative stabilities of the considered structures. Free energies calculated using the gas phase standard state concentration (1 atm = 1/24.5 M) were converted to reproduce the standard state concentration in solution (1 M) by adding or subtracting 1.89 kcal mol^{-1} for bimolecular additions and decompositions, respectively. The lowest energy conformer for each calculated stationary point was considered in the discussion; all the computed structures can be obtained from authors upon request. Cartesian coordinates, electronic energies, entropies, enthalpies, Gibbs free energies, and lowest frequencies of the calculated structures are summarized in **Table 3**.

6.4. Results and Discussion

6.4.1. LovD9 substrate scope towards different acyl donors

LovD has already proven to display broad substrate specificity towards acyl substrates,[260] being DMB-SMMP (**4e**) the most convenient surrogate for the synthesis of the cholesterol-lowering drug SVA[261] reported so far. Attempting to expand the synthetic scope to commercially available or synthetically accessible acyl donors, such as ethyl (S)-2-methyl butyrate (**1d**), vinyl acetate (**2a**), vinyl propionate (**2b**), vinyl butyrate (**2c**), vinyl benzoate (**2f**), p-nitrophenyl butyrate (**3c**), p-nitrophenyl 2-methylbutyrate (**3d**) and p-nitrophenyl 2,2-dimethylbutyrate (**3e**), the highly active LovD9 variant[252] was used to acylate MJA (**5**) assuming that the directed evolution process did not alter the high promiscuity of LovD (**Figure 2.a**). In a first screening attempt, we assayed three ratios of acyl acceptor/acyl donor (1:1, 1:10 and 1:100).

Figure 2.b shows that **1e** was the only acyl donor not accepted by the enzyme at any of the tested acyl acceptor/donor ratios. This was probably due to its already reported low reactivity.[232] **2a** is not reactive when used in a 1:1 ratio, but achieved 15% and 85% acylation yields at 1:10 and 1:100 ratios, respectively. UPLC/MS chromatograms (**Figure 3**) revealed that the acylation of **5** when **2a** acted as acyl donor occurred not only at the C8 hydroxyl group, but also at the hydroxyl groups of the aliphatic chain although to a minor extent. Interestingly, when more carbons were introduced in the acyl group, the acylation was more regioselective toward the C8 hydroxyl position, likely due to increased steric hindrance. Moreover, the acylation yield increased to 34% and 37 % for molar acceptor/donor ratio 1:1 **5/2b** and **5/2c**, respectively, and more than 80% for 1:10 and 1:100 **5/2b** and **5/2c**. It is quite remarkable that the substitution of an alkyl chain by a phenyl group as the acyl substituent (**2f**) also afforded similar yields: 26% at 1:1 and more than 90% at 1:10 and 1:100 ratios. In contrast, p-nitrophenyl derivatives such as **3c** and **3d** provided much lower yields (around 20 % at the three ratios), but the introduction of a second methyl group at the alpha carbon of the aliphatic chain (**3e**) increased acylation yield: 35% and 95% at 1:10 and 1:100 ratios. As expected, DMB-SMMP afforded the highest yields: 44% and 90% at 1:1 and 1:100 ratios. As

reported previously,[261] complete conversion to the statin derivatives is not achieved given the enzyme's ability to hydrolyze acylation products back to MJA.

Figure 2.c shows that 90% of *p*-nitrophenyl esters **3c** and **3d** (10 mM) were converted to *p*NP, while conversion with **3e** was only 26%. However, the statin yield was slightly higher with the latter, suggesting a much higher unproductive hydrolysis rate with **3c** and **3d**.
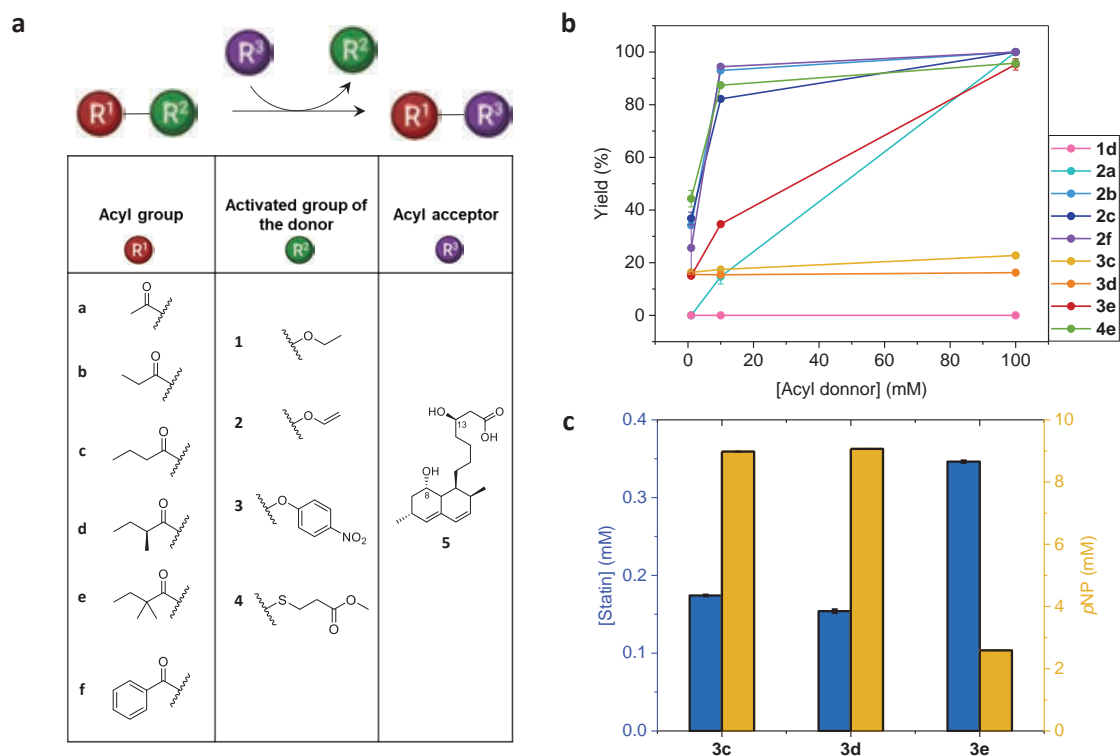
Figure 2. **(a)** Table with different substrates used in this study, decomposed in the acyl group (R[1]), the activated group of the acyl donor (R[2]) and the acyl acceptor (R[3]). Substrates are named as follows: (*S*)-ethyl 2-methyl butyrate (**1d**), vinyl acetate (**2a**), vinyl propionate (**2b**), vinyl butyrate (**2c**), vinyl benzoate (**2f**), *p*-nitrophenyl butyrate (**3c**), *p*-nitrophenyl (*S*)-2-methyl butyrate (**3d**), *p*-nitrophenyl 2,2-dimethyl butyrate (**3e**), α-dimethylbutyryl-S-methyl-3-mercaptopropionate (**4e**) and monacolin J acid (**5**). **(b)** Acylation yields after 24-hour reactions with 3 µM LovD9, 1 mM of **5** and different concentrations (1, 10 and 100 mM) of different acyl donors. **(c)** Concentration of *p*-nitrophenol (*p*NP) and statin derivative after 24-hour reactions with 1 µM LovD9, 1 mM of **5** and 10 mM of *p*NP derivative (**3c**, **3d** or **3e**).

130
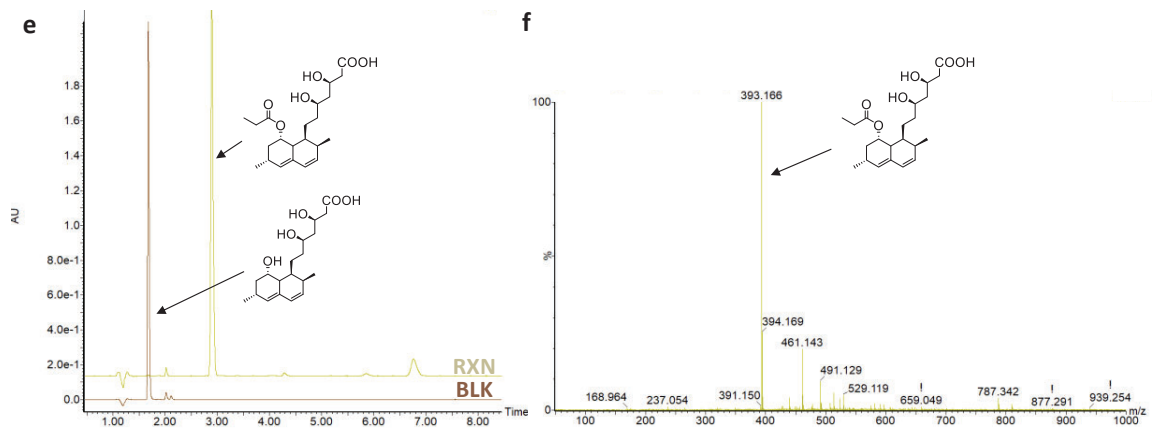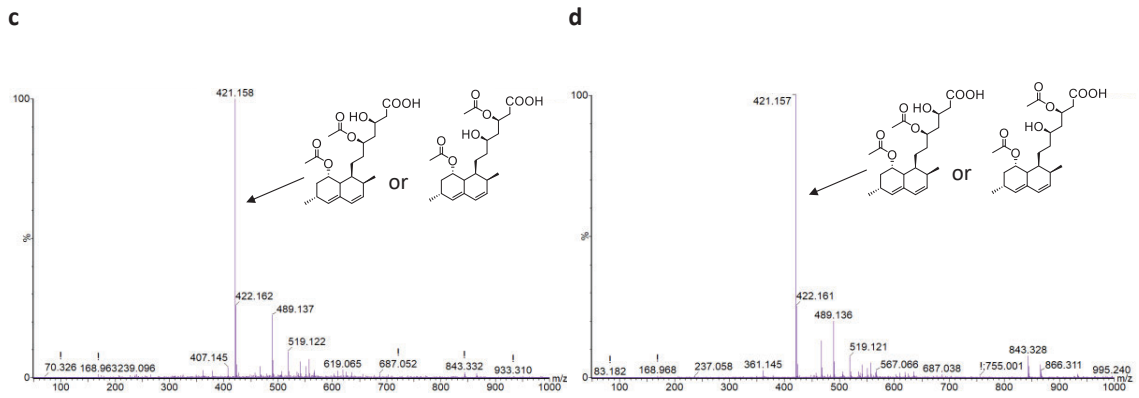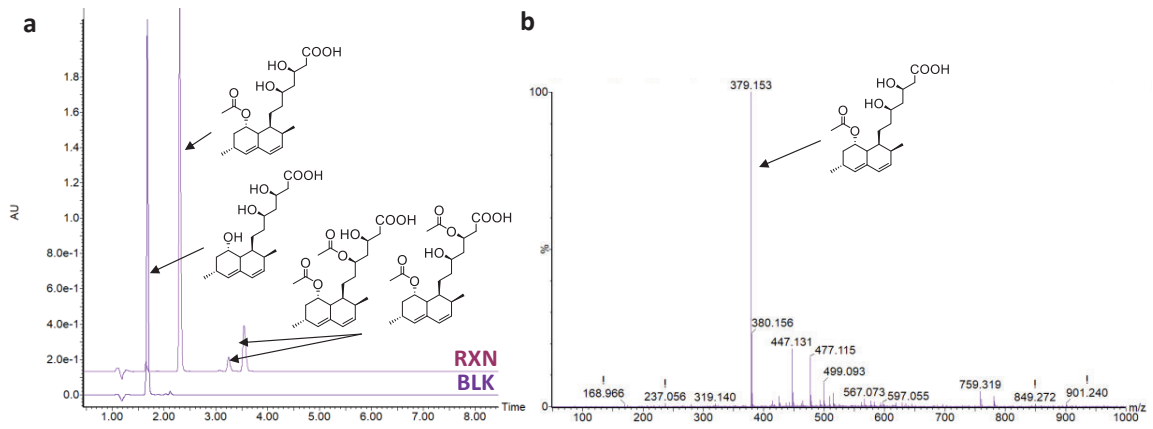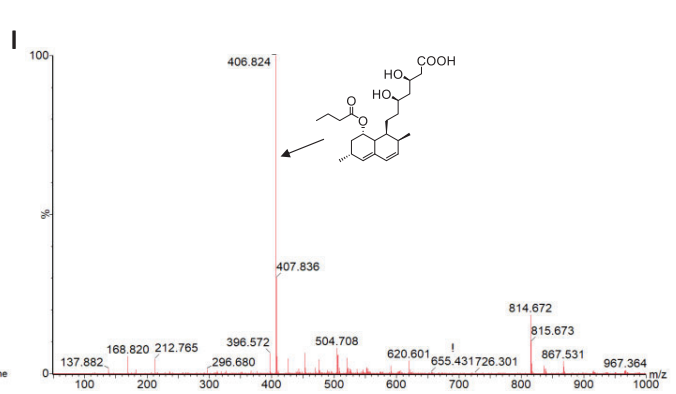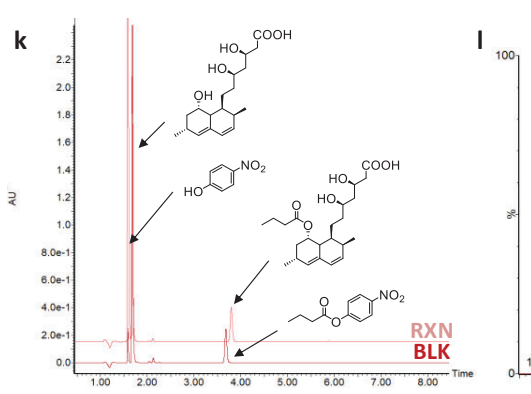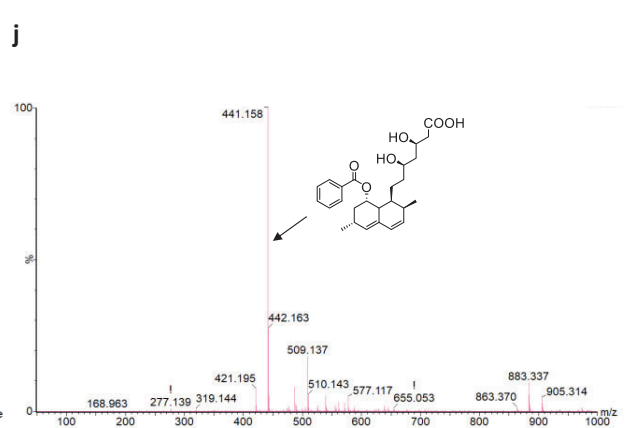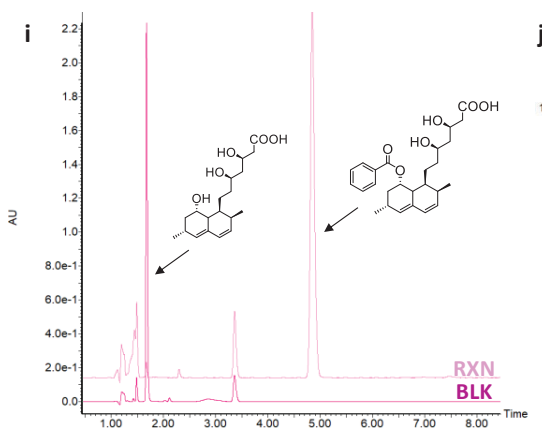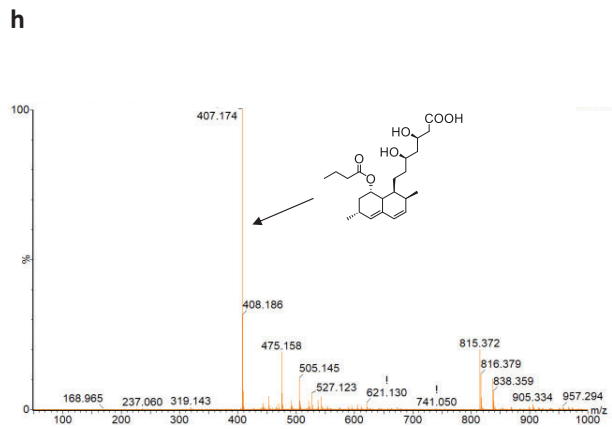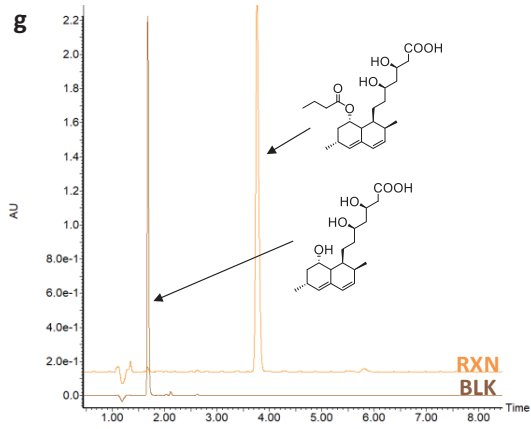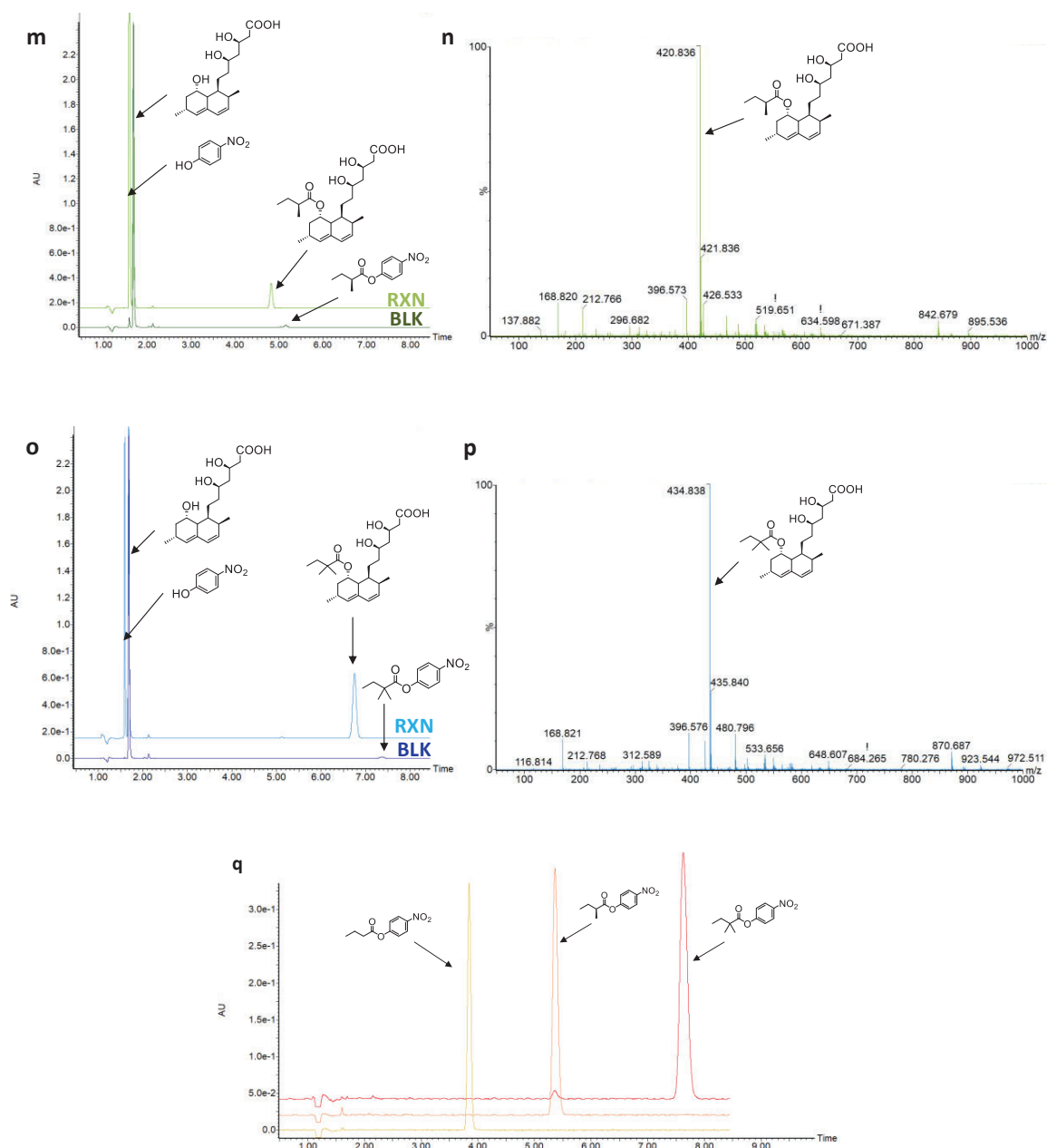
Figure 3. Ultra-performance liquid chromatography (UPLC) traces of samples collected for 24-hour reactions with (reaction, RXN) and without (blank, BLK) 1 μM LovD9, 3 mM monacolin J acid (**5**) and 10 mM vinyl acetate (**2a**) (**a**), vinyl propionate (**2b**) (**e**), vinyl butyrate (**2c**) (**g**), vinyl benzoate (**2f**) (**i**), p-nitrophenyl butyrate (**3c**) (**k**), p-nitrophenyl (S)-2-methylbutyrate (**3d**) (**m**) and p-nitrophenyl 2,2-dimethylbutyrate (**3e**) (**o**) and of **3c**, **3d** and **3e** pure samples (**q**). Mass spectrometry (MS) spectra corresponding to each acylation product: reaction products with **2a** at retention times 2.011 min (**b**), 2.931 min (**c**) and 3.238 min (**d**), reaction product with **2b** at retention time 2.624 min (**f**), reaction product with **2c** at retention time 3.281 min (**h**), reaction product with **2f** at retention time 4.596 min (**j**), reaction product with **3c** at retention time 3.456 min (**l**), reaction product with **3d** at retention time 4.595 min (**n**) and reaction product with **3e** at retention time 6.479 min (**p**).

132

6.4.2. LovD9 kinetic parameters for new acyl donors

The Michaelis constant (K_M) and the turnover number (k_{cat}) for the LovD9-catalyzed ester conversion (combined esterase and transferase activities) with different acyl donors were determined. All experiments were conducted in the presence of MJA at a fixed concentration to mitigate hydrolysis of the acyl-enzyme complex through the direct competition between the acyl acceptor and water.
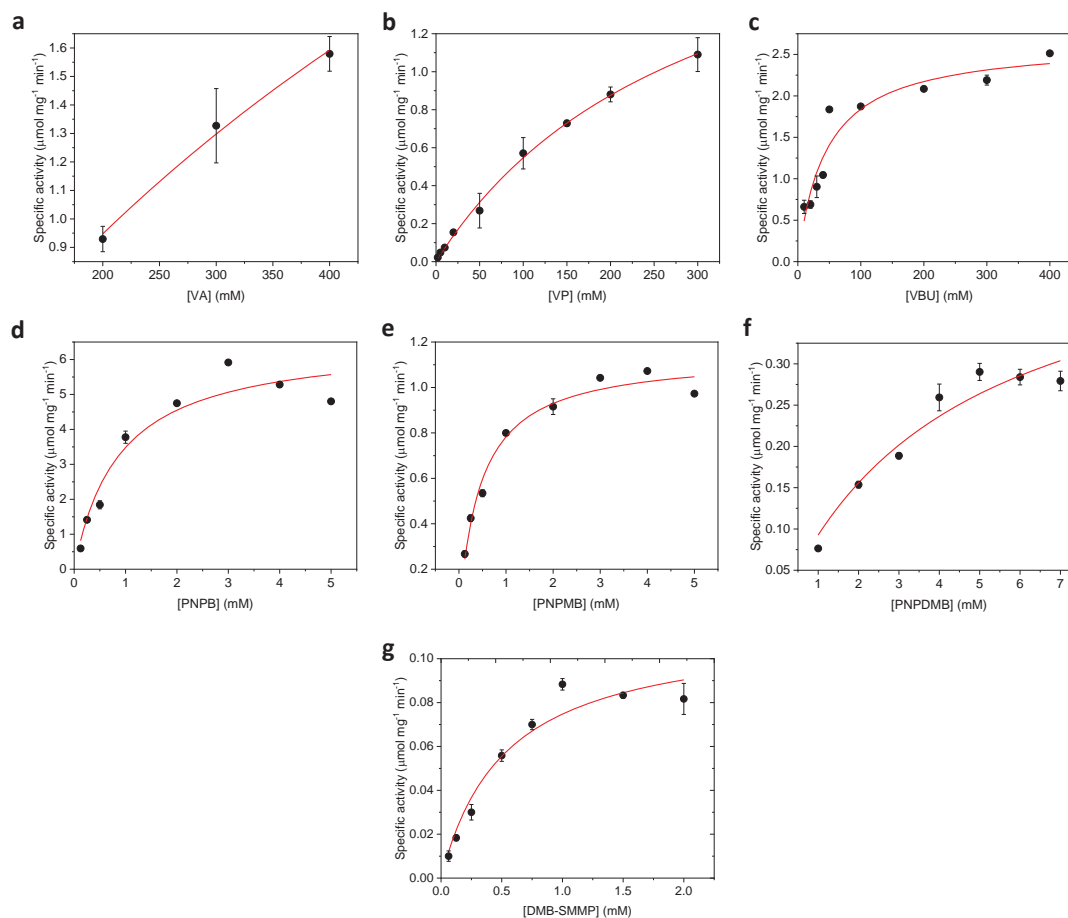
Figure 4. Michaelis-Menten plots for LovD9-catalyzed reaction of vinyl acetate (VA) (**a**), vinyl propionate (VP) (**b**), vinyl butyrate (VBU) (**c**), *p*-nitrophenyl butyrate (*p*NPB) (**d**), *p*-nitrophenyl (*S*)-2-methylbutyrate (*p*NPMB) (**e**), *p*-nitrophenyl 2,2-dimethyl butyrate (*p*NPDMB) (**f**) and α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP) (**g**). The experimental data (black dots) were fitted to the Michalis-Menten model (red line). 1 mM monacolin J acid (MJA) was incubated with 0-400 mM acyl donor and 0.5 µM LovD9.

Table 1. Kinetic parameters for the LovD9-catalyzed ester conversion of different acyl donors: vinyl acetate (**2a**), vinyl propionate (**2b**) and vinyl butyrate (**2c**), *p*-nitrophenyl butyrate (**3c**), *p*-nitrophenyl (*S*)-2-methyl butyrate (**3d**), *p*-nitrophenyl 2,2-dimethyl butyrate (**3e**) and α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (**4e**), in the presence of 1 mM monacolin J acid (**5**).

Structure	k_{cat} (min^{-1})	K_M (mM)	k_{cat} / K_M (min^{-1} mM^{-1})
2a	229.1 (± 56.1)	851.0 (± 288.0)	0.3 (± 0.0)
2b	101.0 (± 7.4)	301.0 (± 36.0)	0.3 (± 0.0)
2c	121.9 (± 9.2)	44.0 (± 11.0)	2.8 (± 0.5)
3c	301.2 (± 29.8)	0.9 (± 0.3)	343.0 (± 78.2)
3d	52.6 (± 1.9)	0.5 (± 0.1)	114.1 (± 12.6)
3e	22.5 (± 3.8)	4.3 (± 1.5)	5.3 (± 0.9)
4e	5.2 (± 0.6)	0.5 (± 0.2)	9.9 (± 1.9)

As observed in **Figure 4** and **Table 1**, vinyl esters with small alkyl chains (**2a**, **2b**) showed K_M values one order of magnitude higher than butyryl donors (**2c**). When compared to vinyl esters, *p*NP derivatives and **4e** exhibited even lower K_M values. These results suggest that LovD9 preferably binds four-carbon acyl thioester donors or esters that are activated with a *p*-nitrophenyl rather than vinyl group and are in agreement with those presented in **Figure 2**. Since **2a** shows the highest K_M, its binding to the enzyme is less favored, and so **5** competitive inhibition is presumably higher. **5** concentration increase affects the 24-hour statin yield to a lesser extent when the acyl donor is **2b**, as the K_M of this substrate is lower than the one of **2a**. Finally, **2c** and **4e** K_M values are closer to that of **5** (0.6 mM, reported in **Chapter 4**), which translates into a superior statin yield at higher **5** concentrations due to less significant substrate inhibition.

On the other hand, vinyl esters showed higher turnover numbers than **4e**. Of note, **3c** features the same acyl group than **2c** and displayed a 3-fold higher k_{cat}. When methyl and dimethyl substituents are introduced at the alpha carbon (**3d** and **3e**), the turnover is significantly reduced. Following the aforementioned trend, **3e**, whose acyl group is the same than that of **4e**, showed a 4-fold increase in turnover number compared to the natural LovD9 surrogate; this compensates the worse binding properties of **3e**, so that the catalytic efficiency of **4e** is still 2-fold higher.

Despite presenting lower K_M values than vinyl esters and higher turnover numbers than **4e**, *p*NP esters afforded reduced statin yields after 24-hour reactions. This could be explained by the fact that kinetic parameters in this study were only measured for the first step of the reaction. Hence, *p*NP derivatives (**3c** and **3d** to a larger extent than **3e**) achieve excellent binding and reactivity parameters, but most of the acyl-enzyme complex is hydrolyzed rather than attacked by **5**, and/or less sterically hindered statins undergo faster hydrolysis compared to SVA. Also, the superior reactivity of the more hydrophilic acyl-Ser76 intermediates and statin products towards water could be at the origin of the observed behavior.

6.4.3. Computational study of LovD catalytic triad acylation mechanisms

The differences of experimental catalytic rate constants (k_{cat}) of LovD9 towards the different acyl donors tested in this study were investigated through quantum mechanics (QM) calculations (**Figure 5** and **Table 2**).
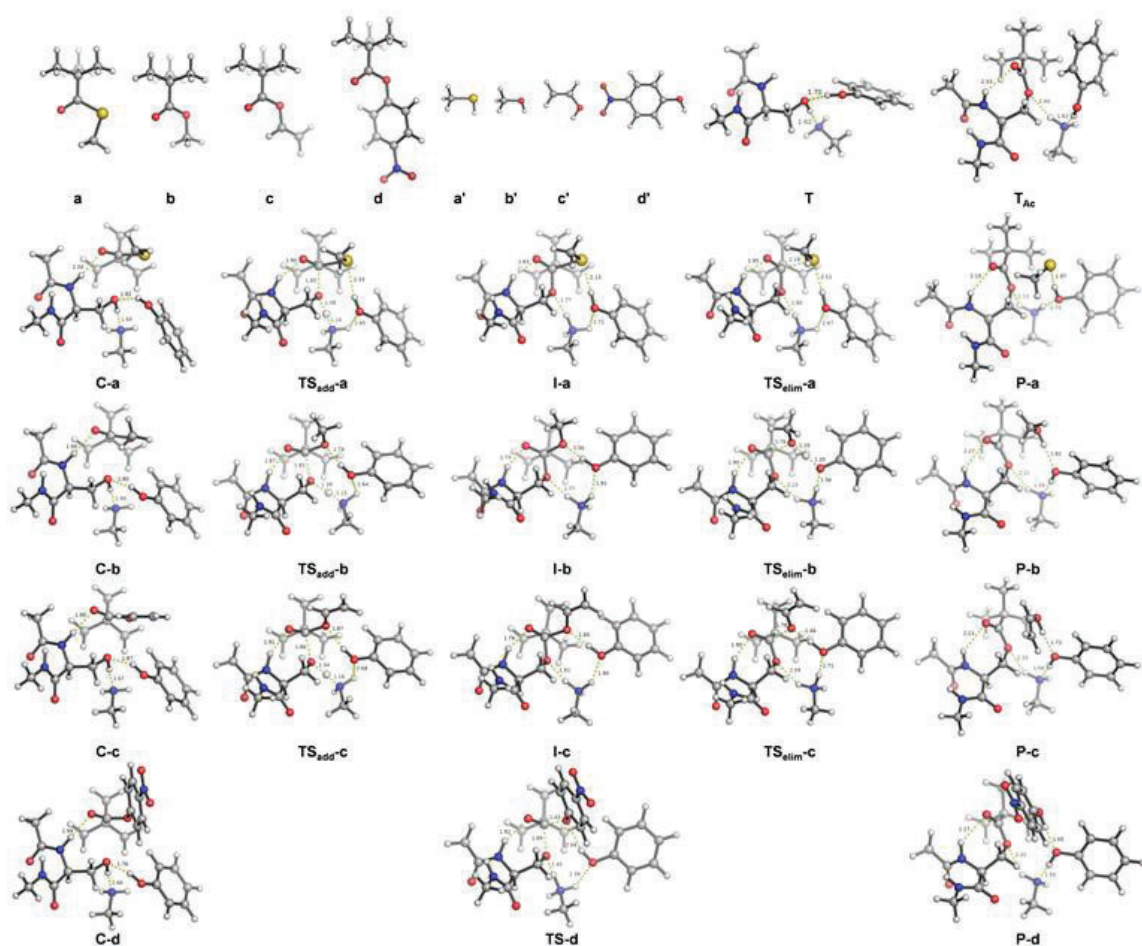

Figure 5. Guide to compound numbering of calculated structures (only the lowest energy conformers are shown).

Table 2. Energies, entropies, and lowest frequencies of the lowest energy calculated structures.

Structure	E_{elec} (Hartree)	E_{elec} + ZPE (Hartree)	H (Hartree)	S (cal mol^{-1} K^{-1})	G (Hartree)	Lowest freq. (cm^{-1})	# of imag. freq.
T	−974.087537	−973.720999	−973.695509	182.1	−973.771822	8.9	0
a	−709.131712	−708.960251	−708.948795	101.6	−708.994904	10.1	0
C-a	−1683.233823	−1682.693385	−1682.656394	230.2	−1682.754726	12.0	0
TS$_{add}$-a	−1683.202907	−1682.664045	−1682.628521	219.7	−1682.723830	−400.6	1
I-a	−1683.209508	−1682.666924	−1682.630973	221.6	−1682.727134	20.3	0
TS$_{elim}$-a	−1683.209493	−1682.667210	−1682.631811	219.7	−1682.726908	−77.1	1
P-a	−1683.221858	−1682.679834	−1682.642870	233.0	−1682.741431	12.9	0
a'	−438.639443	−438.592901	−438.588350	60.6	−438.617142	230.5	0
b	−386.180085	−386.004250	−385.993448	95.6	−386.037892	35.7	0
C-b	−1360.283310	−1359.737760	−1359.701757	223.0	−1359.798204	16.8	0
TS$_{add}$-b	−1360.249734	−1359.706011	−1359.671599	213.5	−1359.764826	−355.6	1
I-b	−1360.254035	−1359.707652	−1359.672878	215.9	−1359.766907	15.2	0
TS$_{elim}$-b	−1360.248112	−1359.706989	−1359.672104	220.4	−1359.766088	−664.8	1
P-b	−1360.274550	−1359.730629	−1359.693803	232.5	−1359.792135	10.8	0
b'	−115.677798	−115.625938	−115.621676	56.8	−115.648681	324.8	0
c	−424.248968	−424.068782	−424.057164	100.9	−424.103721	35.5	0
C-c	−1398.352622	−1397.802977	−1397.765931	229.3	−1397.864436	17.4	0
TS$_{add}$-c	−1398.323578	−1397.776086	−1397.740651	219.2	−1397.835801	−474.8	1
I-c	−1398.332276	−1397.780780	−1397.745067	221.5	−1397.840837	13.2	0
TS$_{elim}$-c	−1398.330880	−1397.781222	−1397.745715	222.2	−1397.841002	−170.0	1
P-c	−1398.356045	−1397.807020	−1397.769790	234.2	−1397.868653	13.6	0
c'	−153.755371	−153.698268	−153.693721	61.0	−153.722708	470.8	0
d	−782.289457	−782.058372	−782.042013	128.8	−782.098898	20.2	0
C-d	−1756.391395	−1755.790834	−1755.749015	256.2	−1755.856523	10.4	0
TS-d	−1756.368278	−1755.769219	−1755.728725	249.1	−1755.833282	−321.5	1
P-d	−1756.406146	−1755.806405	−1755.764546	260.8	−1755.871928	9.0	0
d'	−511.802371	−511.694013	−511.685148	88.8	−511.726772	55.2	0
T$_{Ac}$	−1244.584880	−1244.094843	−1244.062376	207.4	−1244.152191	21.5	0
T	−974.087537	−973.720999	−973.695509	182.1	−973.771822	8.9	0

Figure 6.a-c shows that the thioester model *S*-methyl 2,2-dimethylpropanethioate, the non-activated ester model methyl pivalate and the activated ester model vinyl pivalate, undergo the canonic stepwise acylation mechanism: a nucleophilic addition followed by the elimination of the leaving group (a thiol, an alcohol and an enol, respectively). In all three cases, the Michaelis complex entails a hydrogen bond network involving the S/O atoms of the thioester/ester donor and the hydroxyl group of Tyr188 sidechain (modeled as phenol), and the carbonyl group of the acyl donor and the backbone amide of Ser76 (modeled as Ac-Ser-NHMe). The first step in the enzyme acylation reaction involves nucleophilic attack of Ser76 hydroxyl group with neutral Lys79 (modeled as methylamine) acting as a base, to the carbonyl carbon of the acyl donor (transition states **TS$_{add}$ a-c**), leading to a negatively charged high-energy tetrahedral intermediate (**I-a-c**). The calculated activation barriers for the thioester, non-activated alkyl ester and activated vinyl ester models are 19.4, 20.9 and 18.0 kcal mol^{-1}, respectively. Next, the very reactive tetrahedral intermediates quickly undergo C-S/O bond breaking and concomitant hydrogen transfer through the Lys79-Tyr188 proton shuttle (transition states **TS$_{elim}$ a-c**). For the three models, the energy of TS$_{elim}$ is slightly lower than TS$_{add}$, which becomes the rate-limiting step. The calculated reaction energies suggest that acylation is slightly exergonic with thioesters and nearly thermoneutral with vinyl esters; however, acylation with non-activated esters is calculated to be highly endergonic (i.e. thermodynamically unfeasible), reflecting the leaving group abilities of each reagent.

The acylation mechanism changes when a *p*-nitrophenyl ester model was used as an acyl donor, undergoing a concerted mechanism (**Figure 6.d**). In this case, the *p*-nitrophenyl activated ester binds to the active site through the same hydrogen bond network described above, leading to a much more favored transition state (**TS-d**) with an activation barrier of +14.6 kcal mol^{-1} involving simultaneous O-C and C-O bond formation/breaking and proton shuttling to yield acylated Ser76 and *p*-nitrophenol in a single step. The larger exergonic character of this reaction is related to the change in the reaction mechanism and the much lower activation energy, as *p*-nitrophenoxyde is very polarizable anion and a weak base, and hence a very good leaving group. Intrinsic Reaction Coordinate (IRC) calculations (**Figure 7**) probe the stepwise and concerted nature of the calculated mechanisms.

The calculated activation barriers are in good agreement with the observed experimental trend in terms of relative reactivity of the different donors bearing the same or similar acyl groups: *p*-nitrophenyl esters > vinyl esters > thioesters >> alkyl esters (**Table 1**).
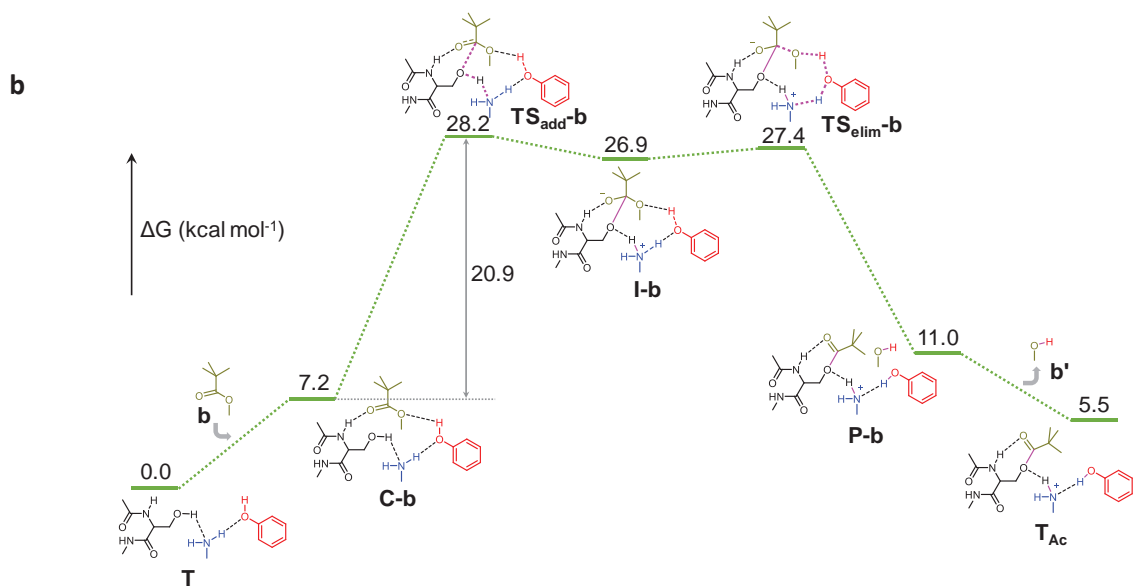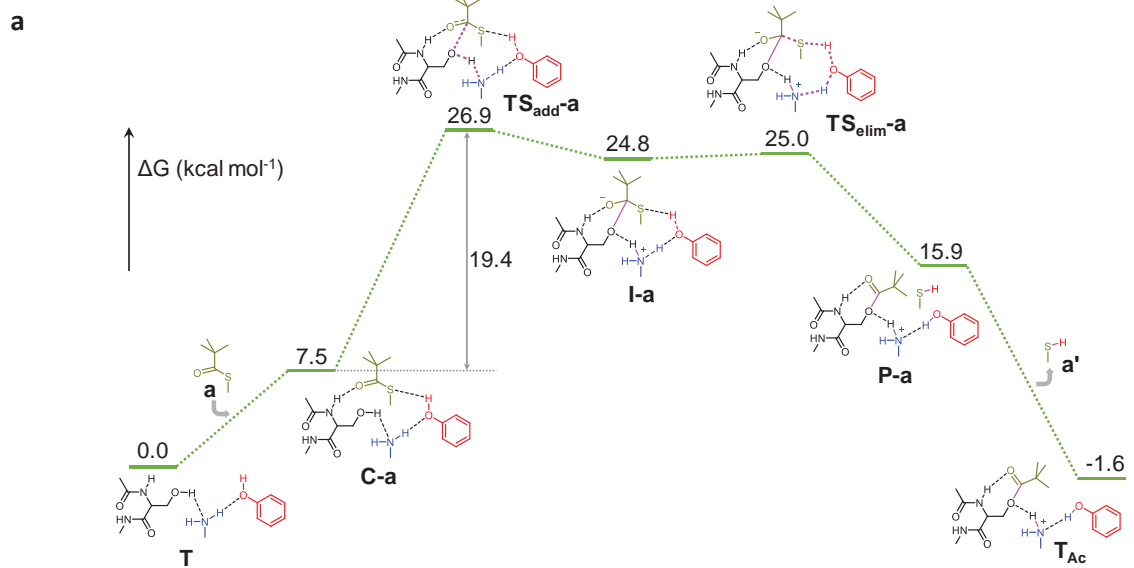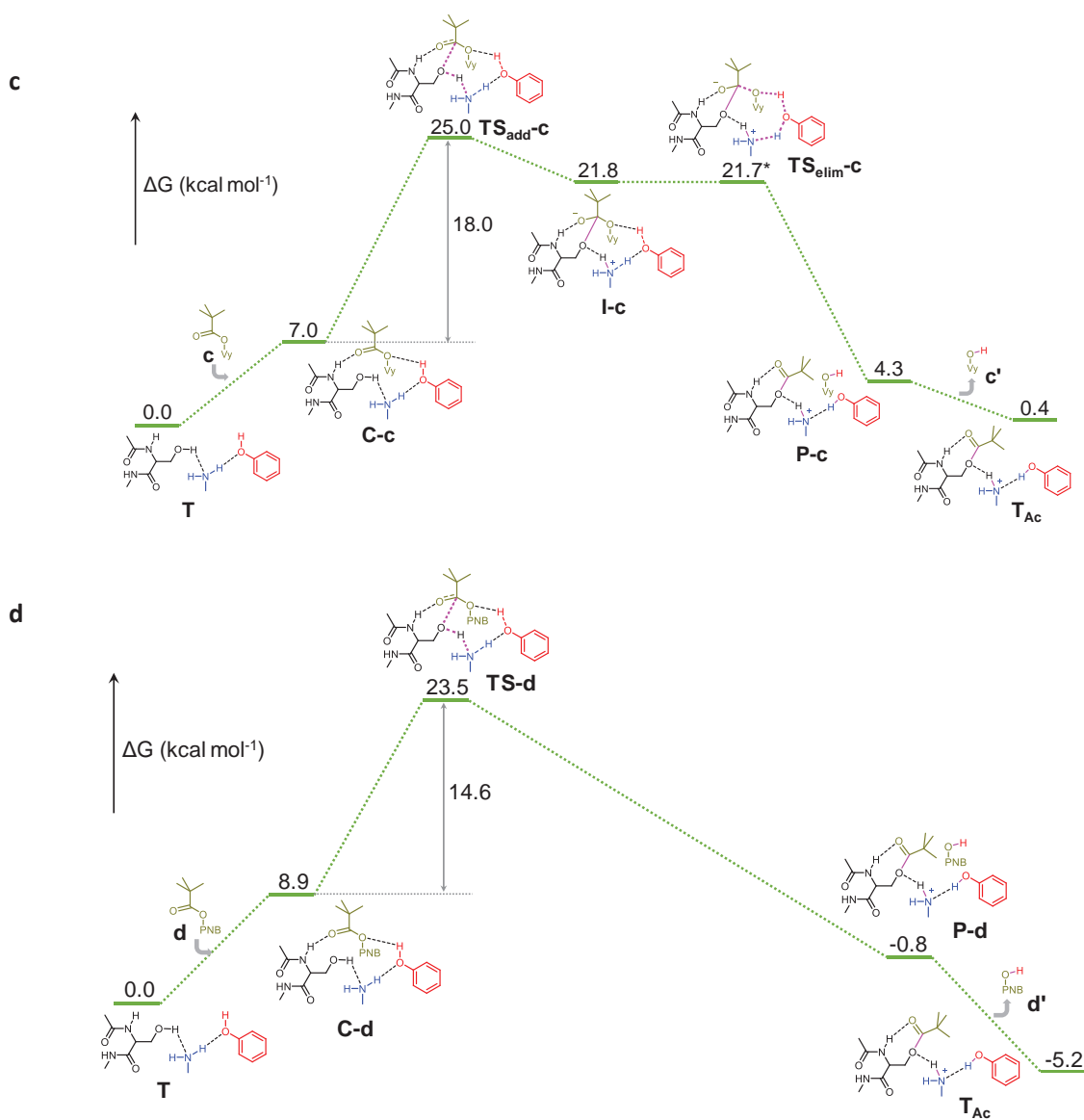
Figure 6. Calculated lowest-energy profiles of LovD catalytic triad (Ser76-Lys79-Tyr188) acylation mechanism with *S*-methyl 2,2-dimethylpropanethioate (**a**), methyl pivalate (**b**), vinyl pivalate (**c**) and *p*-nitrophenyl pivalate (**d**) using abbreviated models.
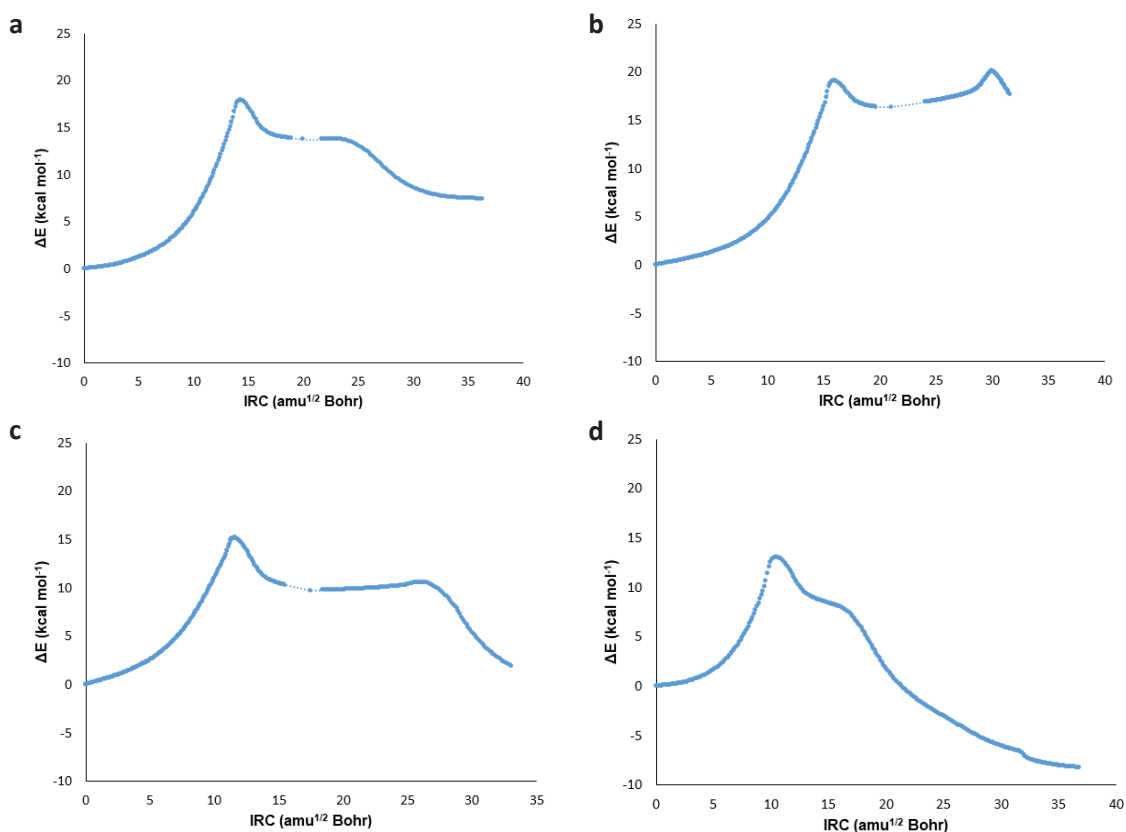
Figure 7. Intrinsic Reaction Coordinate (IRC) plot calculated with PCM(H_2O)/M06-2X/6-31+G(d,p) from the lowest energy nucleophilic addition and elimination (**a**, **b** and **c**) or addition-elimination (**d**) transition states (**TS**, see **Figure 6**) with model acylating reagents *S*-methyl 2,2-dimethylpropanethioate (**a**), methyl pivalate (**b**), vinyl pivalate (**c**) and *p*-nitrophenyl pivalate (**d**). Dashed lines indicate regions of the Potential Energy Surface (PES) that could not be fully explored with IRC calculations due to convergence problems. The individual points located between the nucleophilic addition and elimination PES represent the fully optimized tetrahedral intermediates (**I-a (a)**, **I-b (b)** and **I-c (c)**, see **Figure 6**) connecting both surfaces and are located arbitrarily in the IRC profiles. The zero of the relative energy along the IRC has been arbitrarily set at the first calculated IRC point.

6.4.4. LovD9 substrate scope towards different acyl acceptors

LovD9 substrate scope towards the second acyl transfer step was also studied, with the aim of exploring biocatalytic transformations different from the synthesis of cholesterol-lowering drugs. As the natural substrate of this enzyme is MJA, which is constituted by a hydroxylated unsaturated decalin core, we selected as acyl acceptor candidates various cyclic alcohols with potential commercial interest, such as phenol, 1-naphthol and (+)-menthol.

LovD9 was not able to accept none of the tested acyl acceptors after 24-hour reactions with α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP) as an acyl donor.

In this way, LovD9 has proven to be a quite promiscuous enzyme regarding the first acylation step, but very specific towards MJA as the acyl acceptor. Therefore, the potential industrial application of this enzyme is still limited to statins manufacture.

6.5. Conclusions

The development of new catalytic pathways for the biosynthesis of pharmaceutically relevant statins has been analyzed. In this work, we have characterized the broad substrate specificity of the highly active simvastatin synthase LovD9 towards different commercially available or synthetically accessible acyl donors.

The reactivity (k_{cat}) of vinyl and p-nitrophenyl esters in the presence of LovD9 is even higher than that of the substrate for which the enzyme was evolved (DMB-SMMP). Vinyl esters showed much worse specific binding (K_M) than DMB-SMMP, but their high reactivity make them suitable acyl donor candidates when used in large excess compared to the acyl acceptor. p-Nitrophenyl esters displayed both good k_{cat} and K_M values, but the hydrolysis of the acyl-enzyme complex outcompeted MJA acylation. However, pNPDMB achieved SVA yields comparable to those of DMB-SMMP at a 100:1 acyl donor/MJA ratio thus posing an alternative for SVA manufacturing.

The mechanism for Ser76 acylation with different thioester/ester surrogate models was studied through QM calculations. A gradual shift from the canonical stepwise acylation mechanism with thioesters to a concerted mechanism in the case of p-nitrophenyl esters, was found. These calculations also supported the superior kinetic and thermodynamic properties of p-nitrophenyl esters when compared to the rest of the studied acyl donors, including thioesters. Our work, therefore, provides valuable insights into substrate promiscuity of engineered LovD9 and into the discovery of novel substrates for statins biosynthesis.

CHAPTER 7:

Immobilization of the
acyltransferase LovD-BuCh2
for simvastatin manufacturing
in a continuous flow system

7.1. Introduction

In the last decades, the high demand in delivering new and successful active pharmaceutical ingredients (APIs) to the market has put a pressure into the development of novel research techniques to improve production processes. Out of all the advances in synthetic chemistry, **continuous flow processing (CFP)** has emerged as one of the most useful techniques to improve APIs manufacturing.[193–195] As already mentioned in **Chapter 1**, main advantages offered by a CFP compared with a batch reactor are the high control in the reaction variables, the increased safety, the possibility of automation, the reproducibility, the flexibility of the production volume and the possibility to do in line downstream process for product purification and isolation, among others.

CFP with enzymes are particular interesting in biocatalysis since enzymes catalyze the conversion of multiple target molecules to a broad variety of products of industrial pharmaceutical interest. A relatively large number of APIs already on the market have been directly manufactured (or at least indirectly through one of their intermediates) with an enzymatic process.[38,334,335] Nevertheless, the integration of enzymes into CFP present considerable limitations inherent to their biological nature that must be faced. Among them, their solubility, lability and substrate inhibition seriously limit their direct implementation in a chemical CFP.

The most habitual way of implement enzymes in CFP is through the immobilization of purified enzymes in solid materials, developing **heterogeneous biocatalysts**. In first place, a heterogeneous biocatalyst facilitates the separation of substrates and products with a downstream process. Furthermore, immobilization of enzymes in solid materials often stabilizes the protein scaffold although sometimes also drives to an activity reduction. On this way, the stability/activity ratio is the parameter that will determine the potential of an immobilized enzyme for its industrial application. Once the enzyme stabilization is achieved, the enzyme reuse is a clear benefit in terms of productivity, considering that the catalyst is ready for subsequent reactions without the need of time-consuming and costly extraction or purification procedures. Due to enzyme production costs, few biocatalytic processes are fully implemented in an industrial scale. However, most of them involve immobilized enzymes as this methodology makes the process economically more efficient. One of the most representative examples is the design of an *in vitro* biocatalytic cascade for the manufacturing of the human immunodeficiency virus (HIV) treatment islatravir.[336] Moreover, lipases constitute the higher proportion of CFP achievements due to their large scale availability, promiscuity and ability to operate under a wide variety of conditions. In fact, the most broadly used heterogeneous biocatalyst, both at laboratory and industrial scales, is the Novozyme 435 (lipase B from *C. antarctica* immobilized in a hydrophobic acrylic resin). This industrial heterogeneous biocatalyst has been implemented in CFP for the production of alkyl esters,[41] the oxidation of alkenes,[337] the ring opening polymerization of ε-caprolactone to polycaprolactone,[338] the chemoenzymatic esterification of 3-amino-1,2-propanediol to afford pseudo-ceramides[339] or the multi-step chemoenzymatic synthesis of chiral cyanohydrins.[340] Other relevant examples of CFP with immobilized enzymes for APIs

manufacturing are transaminases immobilization in the landmark Merck process for the production of sitagliptin,[341] *Halomonas elongata* transaminase immobilization onto cobalt-derivatized epoxy-resin for the synthesis of amines and aldehydes[342] or the KRED/NADPH simultaneous immobilization on agarose-based cationic carriers for the continuous synthesis of (trifluoromethyl)benzyl alcohol.[343]

As already reported,[344] immobilization strategies and CFP can be especially helpful in systems presenting undesired side-reactions as substrate inhibition or product hydrolysis. Normally, these reactions are slower than the principal one, so they are diminished at low residence times in the reactor, as their time in contact with the catalyst is reduced. As described in previous chapters, acyltransferase LovD and their engineered variants catalyze simvastatin synthesis from monacolin J acid (MJA), whose efficiency is limited by three undesired factors: competitive inhibition by MJA, hydrolysis of the acyl-enzyme complex (thioesterase activity) and hydrolysis of the final product (esterase activity), as described in **Chapter 4**.[251,252,260,261]

As this enzyme has never been tested as a heterogeneous biocatalyst, the aims of this study are:

1) the evaluation of its **activity and stability** in different porous carriers with distinctive functional groups

2) the characterization of the **spatial distribution** of the enzyme in the material

3) the implementation in a **CFP process**.

Our designed LovD-BuCh2 variant was selected for this study due to its adequate catalytic properties and thermal stability described in **Chapter 4** (**Scheme 1**).
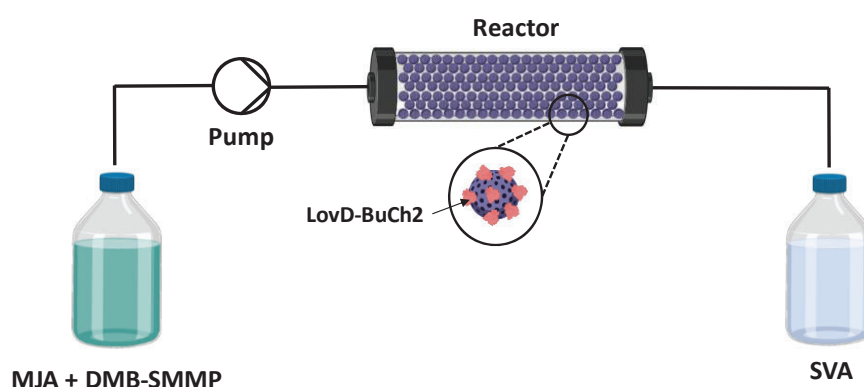
Scheme 1. Continuous flow system for simvastatin manufacturing using the acyltransferase LovD-BuCh2.

7.2. Experimental section

7.2.1. Materials

Common materials used in previous chapters are already described in the **Experimental section** of **Chapter 4**. Agarose microbeads 4BCL (AG, particle size 50-150 µm, pore size 50 nm) with and without cobalt chelates (30 µmol of Co^{2+} g support $^{-1}$) were purchased from Agarose Bead Technologies (Madrid, Spain). Porous glass microbeads EziG1 (particle size 75-150 µm, pore size 50 nm, 10 µmol Fe^{3+} g support $^{-1}$) were purchased from EnginZyme (Wincester, UK). $NaBH_4$, $CoCl_2$, iminodiacetic acid (IDA), rhodamine B isothiocyanate (RBITC) and fluorescein isothiocyanate (FITC) were purchased from Sigma-Aldrich (St.Louis, IL, USA).

7.2.2. Expression of LovD-BuCh2

Expression of LovD-BuCh2 was performed as described in the **Experimental Section** of **Chapter 4**.

7.2.3. Functionalization of agarose microbeads with epoxide groups and cobalt chelates

1 g of agarose 4BCL were carefully diluted in 1.6 mL of acetone and 44 mL of water containing 0.2 g of $NaBH_4$ and 0.1 M of NaOH at 4 °C. 11 mL of (±)-epichloridrin were added very slowly at the same temperature to avoid epoxide hydrolysis. The mixture was incubated for 16 h at room temperature with mild stirring. Finally, the suspension was filtered off and washed with 10 mL of water. Then, 1 g of epoxi-agarose was modified by adding 0.5 M of iminodiacetic acid (IDA) in water at pH 11 and stirring at room termperature for 1 h with mild agitation. The suspension was filtered off and washed with 10 mL of water. Under these conditions the activation degree is 20.2 µmol g^{-1} of IDA and 18.5 µmol g^{-1} of epoxides. 1 g of IDA-epoxi-agarose was additionally modified by adding 10 mL of a $CoCl_2$ solution (30 mg mL^{-1} in water) and incubated for 1 h at 25 °C with mild stirring. Finally, the suspension was filtered off and washed with 10 mL of water.

7.2.4. Immobilization and one-step purification of LovD-BuCh2 in different carriers

1 mL of the clear crude extract containing His-tagged LovD-BuCh2 was diluted (1:5) in 4 mL of sodium phosphate buffer 50 mM pH 8 and incubated with 0.5 g of agarose or porous glass microbeads activated with cobalt or iron chelates (AG-Co^{2+} and EziG1) for 1 h at 4 °C. The protein was also immobilized in agarose-Co(II) functionalized with epoxide groups (AG-Co^{2+}/E). Then, the resulting colloid was incubated for 1 h in 5 mL of a 1 M glycine solution in phosphate buffer 50 mM pH 8 to cap the remaining epoxide groups thus avoiding unspecific interactions. Then, the beads were washed with 10 mL of 50 mM sodium phosphate buffer, vacuum dried, and stored at 4 °C for further use. A fraction of the immobilized enzyme was eluted with 250 mM imidazole and 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) at pH 8 to assess further assays with both soluble and immobilized enzymes. Immobilization rate for each carrier was qualitatively assessed through sodium dodecyl

147

sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and quantitatively determined through the Bradford protein assay,[270] calculating the difference of protein concentration between the crude extract and the flow-through recovered after immobilization.

7.2.5. Calculation of immobilization parameters

The immobilization parameters characterized in this study were calculated as follows:

- The **load** is the mass of immobilized protein per gram of carrier.

$$\text{Load (mg g}^{-1}) = \left(\text{offered protein (mg mL}^{-1}) - \text{protein in supernatant (mg mL}^{-1})\right) \times \frac{\text{immobilization volume (mL)}}{\text{carrier mass (g)}} \quad \text{(1)}$$

- The **immobilization yield (Ψ)** is defined as the rate of the offered enzyme immobilized on the carrier.

$$\Psi = \frac{\text{offered protein (mg mL}^{-1}) - \text{protein in supernatant (mg mL}^{-1})}{\text{offered protein (mg mL}^{-1})} \times 100 \quad \text{(2)}$$

- The **recovered activity (RA)** is defined as the real enzyme activity per gram of carrier and is expressed in U g^{-1}.

- The **immobilized specific activity (iSA)** is defined as the activity per gram of immobilized enzyme.

$$\text{iSA (U mg}^{-1}) = \frac{\text{RA (U g}^{-1})}{\text{Load (mg g}^{-1})} \quad \text{(3)}$$

- The **relative recovered activity (rRA)** is defined as the ratio between the specific activity of immobilized and free enzyme.

$$\text{rRA} = \frac{\text{iSA}_{\text{immobilized enzyme}}}{\text{iSA}_{\text{free enzyme}}} \times 100 \quad \text{(4)}$$

7.2.6. Thioesterase spectrophotometric assay

Thioesterase spectrophotometric assay was performed as described in the **Experimental Section** of **Chapter 4**.

7.2.7. Protein labeling with fluorescent probes

An enzyme solution in 100 mM of sodium bicarbonate buffer at pH 8.5 was mixed (1:10 molar ratio) with fluorofore rhodamine B isothiocyanate (RBITC) or fluorescein isothiocyanate (FITC) solutions in DMSO (10 mg/ml) and incubated 1 h at 25 °C in the dark. Labeled enzyme was filtered off using a tangential ultrafiltration unit (10 kDa) and washed with 50 mM HEPES at pH 8 until no coloring was observed in the filtered solution. 1 mL of

the pure fluorophore-labelled LovD-BuCh2 was diluted (1:10) and immobilized in different supports as described above for the crude extract of LovD-BuCh2.

7.2.8. Confocal Laser Scanning Microscopy (CLSM) imaging and analysis

The localization and distribution of fluorophore-labelled LovD-BuCh2 throughout the different supports were analyzed with a confocal microscope Espectral ZEISS LSM 880 equipped with an excitation laser (λ_{ex} = 561 nm for RBITC) and emission filter (LP505 and LP565, respectively). Confocal images were obtained using x20 and x40 objectives. For quantitative optimization of protein migration, we processed CLSM acquired images at different storage times and under different conditions to obtain a normalized fluorescence intensity radial profile of single beads using FIJI software (**Figure 1.a,b**). Subsequently, we developed a single-particle analysis algorithm that was implemented in the above-mentioned image processing program. The developed plugin automatically performs a Gaussian fit on the normalized fluorescence intensity radial profile of every single bead of the confocal image under analysis (**Figure 1.c**). Then, the plugin searches for the fitted data point that corresponds to the radius coordinate ($x_{50\%}$), where fluorescence is the 50% of the maximum normalized fluorescence fitted peak ($y_{50\%}$). Then, $x_{50\%}$ is subtracted from the radius of the analyzed bead to obtain the infiltration distance (μm). Finally, protein migration is defined as relative infiltration which is the average infiltration distance divided by beads by the average radius of the analyzed beads (n = 10), expressed as percentage (%).
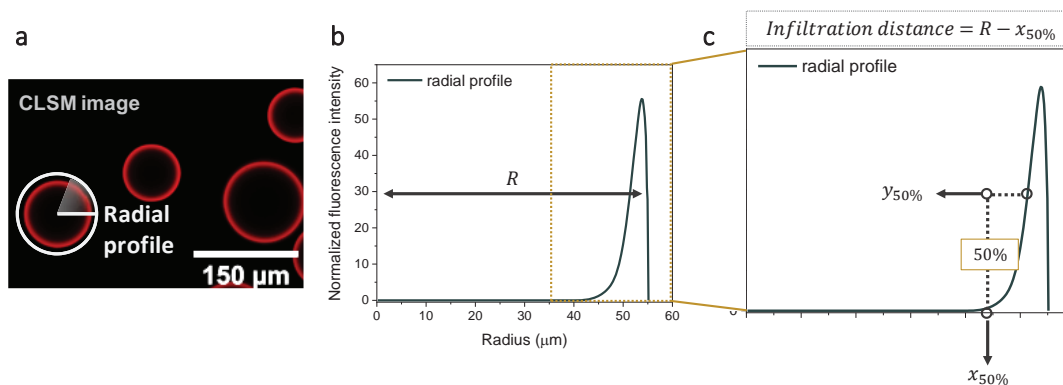
Figure 1. Explanatory illustration of infiltration distance calculation. **(a)** Representation of fluorescence radial profile of a single microbead in a confocal laser scanning microscopy (CLSM) image. **(b)** Obtained normalized fluorescence intensity radial profile (R represents the particle radius) and **(c)** normalized fluorescence intensity radial profile (grey continues line) together with the corresponding Gaussian fit (yellow dots) and explanatory illustration of infiltration distance estimation derived from the obtained fitted radial profile.

7.2.9. Analysis of intrinsic protein fluorescence

70 µg of LovD-BuCh2 both in solution and immobilized in different supports were resuspended in 50 mM HEPES at pH 8 and placed in a 96-well dark plate before and after incubating for 4 h at 40 °C. The plate was irradiated at 280 nm with excitation and fluorescence emission spectra was recorded between 300 and 500 nm; bandwidths of 5 nm were used for both excitation and emission.

7.2.10. Protein fluorescence anisotropy measurements

3 ng of FITC-labelled LovD-BuCh2 both in solution and immobilized in different supports were added in a 96-well dark plate and fluorescence anisotropy was measured in a BioTek Microplate Reader Synergy. Anisotropy values were calculated using the following equation:

$$r = \frac{I_z - I_y}{I_z + I_y} \quad (5)$$

where I_z is the parallel intensity and I_y is the perpendicular intensity. Anisotropy values of immobilized LovD-BuCh2 were normalized to the anisotropy of free LovD-BuCh2. Values lower than 1 mean immobilizate has a higher rotational tumbling than the free enzyme, while values higher than 1 mean immobilizate has a lower rotational tumbling than the free enzyme.

7.2.11. Thermal inactivation assays

3 µM solution of either free or immobilized enzyme in 50 mM HEPES 10 mM MgCl$_2$ (pH 8) was incubated 1 h at different temperatures (25, 37, 40, 45 and 50 °C) and the activity was measured with a spectrophotometric assay performed in 96-well plates for 30 min at 37 °C using 1 µM of enzyme, 3 mM MJA, 2 mM DMB-SMMP and 2 mM 2,2- 'dithiodipyridine (2-DTDP) in 50 mM HEPES 10 mM MgCl$_2$ (pH 8) 10 % DMSO. The thiol attack to 2-DTDP generates 2-thiopyridone, which absorbs at 323 nm with an absorption coefficient of 7600 M^{-1} cm $^{-1}$. The enzyme solution was also incubated at 40 °C and the activity was measured at different times (0; 15; 30; 60; 120 and 240 minutes) of incubation with the same assay described above.

7.2.12. Raman spectroscopy

Raman spectra and maps were acquired using a confocal Raman microscope (alpha 300R, Witec GmbH) which is equipped with four fibre-coupled excitation sources (488, 532, 633 and 785 nm) and 2 fibre-coupled, lens-based spectrometers optimized for visible and near-infrared range, a motorized scanning stage hooked with a piezo-driven scanning platform for high-precision imaging. Raman spectra were collected from solid (dry) samples under excitation with a 532 nm laser with a power of 65 mW focusing through a 20x (numerical aperture, NA=0.5) or through a 50x (NA=0.75) objective using the spectrometer

of 600 mm focal length, a 300 lines/mm grating and a back-illuminated EMCCD camera with 1600 x 200 pixel. For single-point Raman spectra, the integration time was set to 0.2 s and 10 spectra were accumulated and averaged. Alternatively, averaged Raman spectra were also measured by scanning over a determined area and taking single-point spectra with a step size of 1 um in x and y, an integration time of 50 ms or 100 ms and averaging the spectra over the whole area. The local distribution of components in the sample can be visualized by chemical imaging. For this application, the Raman maps were created by plotting the Raman intensity of the fingerprint as a function of the position. Project5/5+ Software (Witec GmbH) was used to process and analyze the data. First, cosmic rays and the background were removed and afterwards the True Component Analysis tool applied to find and identify (different) components. Each component spectrum found can be assigned to a color and the corresponding fingerprint spectrum intensity to a color scale. The plot of each measured pixel (x, y position) as a function of the component and its intensity results in false-color image showing the local distribution and the concentration of components.

7.2.13. Operational stability of LovD-BuCh2 immobilized in AG-Co^{2+}/E and in EziG1 in batch

Batch reaction time-courses were measured for free LovD-BuCh2, LovD-BuCh2@AG-Co^{2+}/E and LovD-BuCh2@EziG1 at 3 mM MJA, 2mM DMB-SMMP in 50 mM HEPES (pH 8) and 10% dimethyl sulfoxide (DMSO) to facilitate solubilization of DMB-SMMP. Reactions were initiated in a 5 mL column by adding 1 µM of enzyme in HEPES 50 mM MgCl$_2$ 10 mM pH 8 and quenched at 1, 2, 4, 8 and 24 h by collecting by filtration 100 µL of the reaction mixture, separating it from the immobilized enzyme. Upon one 24 h reaction cycle, the immobilized enzymes were separated, washed with HEPES 50 mM MgCl$_2$ 10 mM pH 8 and placed back into the column for a consecutive 24 h reaction cycle. Samples were analyzed by Ultra Performance Liquid Chromatography (UPLC) (Waters 2690) equipped with a PDA detector using a ACQUITY UPLC® BEH C18 1.7 µm (2.1 x 50 mm) Waters column coupled to a LCT XE time-of-flight mass spectrometry detector with electrospray ionization source (ESI). Analytes were eluted with an isocratic mobile phase composed of 52 % (v/v) of acetonitrile in water (0.1 % (v/v) formic acid) during 15 min at flow rate of 0.3 mL min^{-1}.

7.2.14. Continuous flow processing for simvastatin synthesis and hydrolysis with LovD-BuCh2 immobilized in EziG1

1 g of LovD-BuCh2 immobilized in EziG1 was packed in a 1 cm^3 column for flow experiments. Flow reactions were performed with 1 mM MJA, 2 mM DMB-SMMP in 50 mM HEPES (pH 8) and 10% DMSO at different flow rates (10, 20, 50, 100 and 200 µL min^{-1}). Collected samples at the outlet of the column were analyzed by UPLC/MS using the same method described above. Other substrate ratios were also tested for the flow microrreactor: 3 mM MJA / 2 mM DMB-SMMP and 1 mM MJA / 10 mM DMB-SMMP. SVA hydrolysis was monitored by passing a solution of 2 mM SVA at flow rates of 10 µL min^{-1} and 200 µL min^{-1}.

7.3.   Results and Discussion

7.3.1. Immobilization and characterization of different LovD-BuCh2 biocatalysts

With the aim of obtaining a productive heterogeneous biocatalyst with LovD-BuCh2, we tested the immobilization of the enzyme in three porous carriers with different physical properties and functionalizations. We chose agarose microbeads functionalized with imidoacetic acid (IDA) and cobalt chelates (AG-Co^{2+}) and porous glass particles functionalized with iron-catechol complexes (EziG1) to immobilize the enzyme through its His-tag in N-terminus through a site-directed manner. Apart from site-direction, rigidification of the protein structure was achieved through the epoxidation of AG-Co^{2+}, switching on a carrier (AG-Co^{2+}/E) to which the enzyme can be attached in a covalent irreversible way. This immobilization occurs in a sequential way: the enzyme quickly interacts with the carrier through the affinity immobilization driven by Co^{2+}-chelates and the His-tag (as described above), and then covalent and irreversible bonds are formed through the nucleophilic attack of positively charged amino acids in the enzyme surface (such as lysines) to epoxide groups at the surface of the carrier (**Figure 2**).
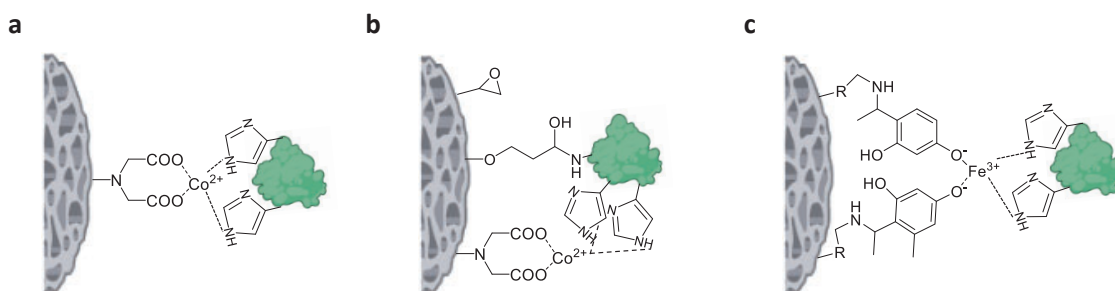
Figure 2. Chemical functionalization of porous carriers for different immobilization strategies. **(a)** Affinity immobilization through poly-His tag coordination with cobalt chelates functionalized with iminodiacetic acid (IDA) in agarose beads. **(b)** Covalent immobilization of affinity-bound enzyme (through poly-His tag coordination with cobalt chelates, as described in **(a)**) through nucleophilic attack of protein sidechains to epoxides of the activated agarose carrier. **(c)** Affinity immobilization through poly-His tag coordination with iron chelates functionalized with chatecol groups in EziG1 porous glass beads. The enzyme is represented in green and the carrier is represented as the porous grey surface.

Cell-free extracts of overexpressed LovD-BuCh2 in *E. coli* were incubated with each carrier, with the selectivity of the binding allowing immobilization and one-step purification. We found immobilization yields of 100 % and 95 % for AG-Co^{2+} and AG-Co^{2+}/E carriers, with protein loads of 1.11 and 1.06 mg ml^{-1}, while EziG1 was only able to immobilize 30.8 %, with an enzyme load of 0.34 mg ml^{-1} (**Table 1**). SDS-PAGE analysis of the immobilization (**Figure 3**) shed light on the difference in the immobilization yield and protein load between the agarose based and EziG1 immobilizates: while AG-Co^{2+} and AG-Co^{2+}/E bound LovD-BuCh2 in a very specific way, the enzyme was not bound with such a high selectivity to EziG1. This phenomenon could be mostly explained by the more unspecific coordination of the His-tag to Fe^{3+} compared to Co^{2+}. [345,346] Moreover, when EziG1 carrier with immobilized LovD-BuCh2

was incubated at 100 °C with β-mercaptoethanol, the enzyme was not leaked from the surface, which suggested that this porous glass carrier could form additional electrostatic interactions with the enzyme that do not occur in agarose-based materials, providing a stronger binding. On the other hand, we observed that in both carrier and eluted fractions, 100 % of the enzyme was leaked from AG-Co^{2+}, while only 50 % of the enzyme was leaked from AG-Co^{2+}/E. This observation suggested that around the half of the immobilized enzyme correctly stablished covalent interactions with the epoxide groups of the carrier.

Table 1. Immobilization parameters of LovD-BuCh2 in different carriers. Ψ : Protein immobilization yield; RA: recovered activity (activity measured upon the immobilization protocol taking free enzyme as reference); iSA: immobilized specific activity. rRA: relative recovered activity, calculated in relation to free LovD-BuCh2 SA (0.41 U mg^{-1}).

Material	Support	Functional group	Density groups (µmol g^{-1})	Particle size (µm)	Ψ (protein) (%)	Load (mg g^{-1} support)	RA (U g^{-1} support)	iSA (U mg^{-1})	rRA (%)
Agarose	AG-Co^{2+}	IDA-Co^{2+}	30	50-150	100	1.11	0.316	0.28	68.3
	AG-Co^{2+}/E	IDA-Co^{2+}/Epoxide	20	50-150	95.0	1.06	0.100	0.09	21.9
Glass	EziG1	Catechol-Fe^{3+}	10	75-125	30.8	0.34	0.114	0.34	82.9

The thioesterase enzyme activity in all the three carriers was assessed through a colorimetric method based on the detection of the product intermediate methyl 3-mercaptopropanoate released during the first reaction step. The enzyme recovered activity per mass of carrier (RA) and the immobilized specific activity (iSA) significantly varied depending on the immobilization strategy and the material of the carrier utilized. LovD-BuCh2 immobilized on AG-Co^{2+} (LovD-BuCh2@AG-Co^{2+}) presented a RA of 0.316 U g^{-1} of support and an iSA of 0.28 U mg^{-1}, considerably higher than for LovD-BuCh2 immobilized in AG-Co^{2+}/E (LovD-BuCh2@AG-Co^{2+}/E) (RA = 0.1 U g^{-1} of support; iSA = 0.09 U mg^{-1}). Because of the irreversibility of the covalent bond formed between scattered residues of the protein and the epoxides in AG-Co^{2+}/E, conformational changes of the protein structure are certainly limiting the proper access of the substrate to the catalytic site, which translates into a relevant loss of activity compared to that of the enzyme in solution (rRA = 21.9%). The lack of these covalent bonds in LovD-BuCh2@AG-Co^{2+} is translated into a weaker binding, but also guarantees a less pronounced activity loss (rRA = 68.3%). On its turn, LovD-BuCh2 immobilized in EziG1 (LovD-BuCh2@EziG1) displayed a lower RA than LovD-BuCh2@AG-Co^{2+} (0.114 U g^{-1} of support), but the highest iSA (0.34 U mg^{-1}), which indicated that less enzyme was immobilized in the carrier, but its activity loss was minor than for the other two biocatalysts (rRA = 82.9%).
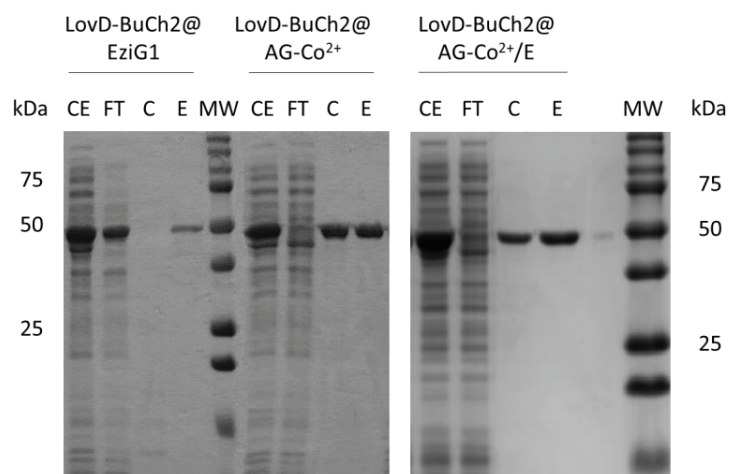
Figure 3. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for protein expression and immobilization of LovD-BuCh2 in different carriers. CE: Crude Extract. FT: Flow-through. C: Carrier. E: Eluted protein. MW: molecular weight marker.

7.3.2. Spatial distribution of LovD-BuCh2 in different carriers

The spatial distribution of the enzyme in the carrier provides very valuable information about the effectivity of the biocatalyst. LovD-BuCh2 was previously labeled with the fluorophore Rhodamine B isothiocyanate (RBITC) and then immobilized onto the three carriers, and spatial distribution was determined by fluorophore detection through confocal laser scanning microscopy (CLSM) (**Figure 4**).

LovD-BuCh2 exhibited a similar spatial distribution on agarose-based materials (AG-Co^{2+} and AG-Co^{2+}/E), with an external immobilization pattern due to the rapid interaction of the His-tag of the protein with the IDA-Co^{2+} groups in the surface of the bead that avoids the penetration of the enzyme into deeper layers. On the other hand, a larger penetration of the enzyme in the material is observed for EziG1 (**Figure 4.a-c**). 3D profiles (**Figure 4.d-f**) confirmed that the fluorescence in LovD-BuCh2@AG-Co^{2+} was mostly detected on the surface of the particle, while this fluorescence intensity was accumulated in even more external layers in the case of LovD-BuCh2@AG-Co^{2+}/E. The slight difference between both distribution profiles can be explained by the covalent bond generated in LovD-BuCh2@AG-Co^{2+}/E, which hampers a higher diffusion of the enzyme into the inner bead. LovD-BuCh2@EziG1 exhibited a profile with a larger distribution of the fluorescence intensity within the internal layers of the particle. This higher internalization of the enzyme into the bead could be in agreement with the SDS-PAGE analysis (**Figure 3**) and be influenced by the existence of additional electrostatic interactions between the enzyme and the carrier, as suggested before.[347,348] Additionally, the larger infiltration within the material is traduced in a lower enzyme density per μm^3 of carrier, reducing the protein molecular packing, which could explain the superior rRA of LovD-BuCh2@EziG1 (**Table 1**). The infiltration distance

calculated from the radial profile of around 20 particles (**Figure 4.g**) was 6.54 µm for AG-Co^{2+}, 3.29 µm for AG-Co^{2+}/E and 24.42 µm for EziG1.
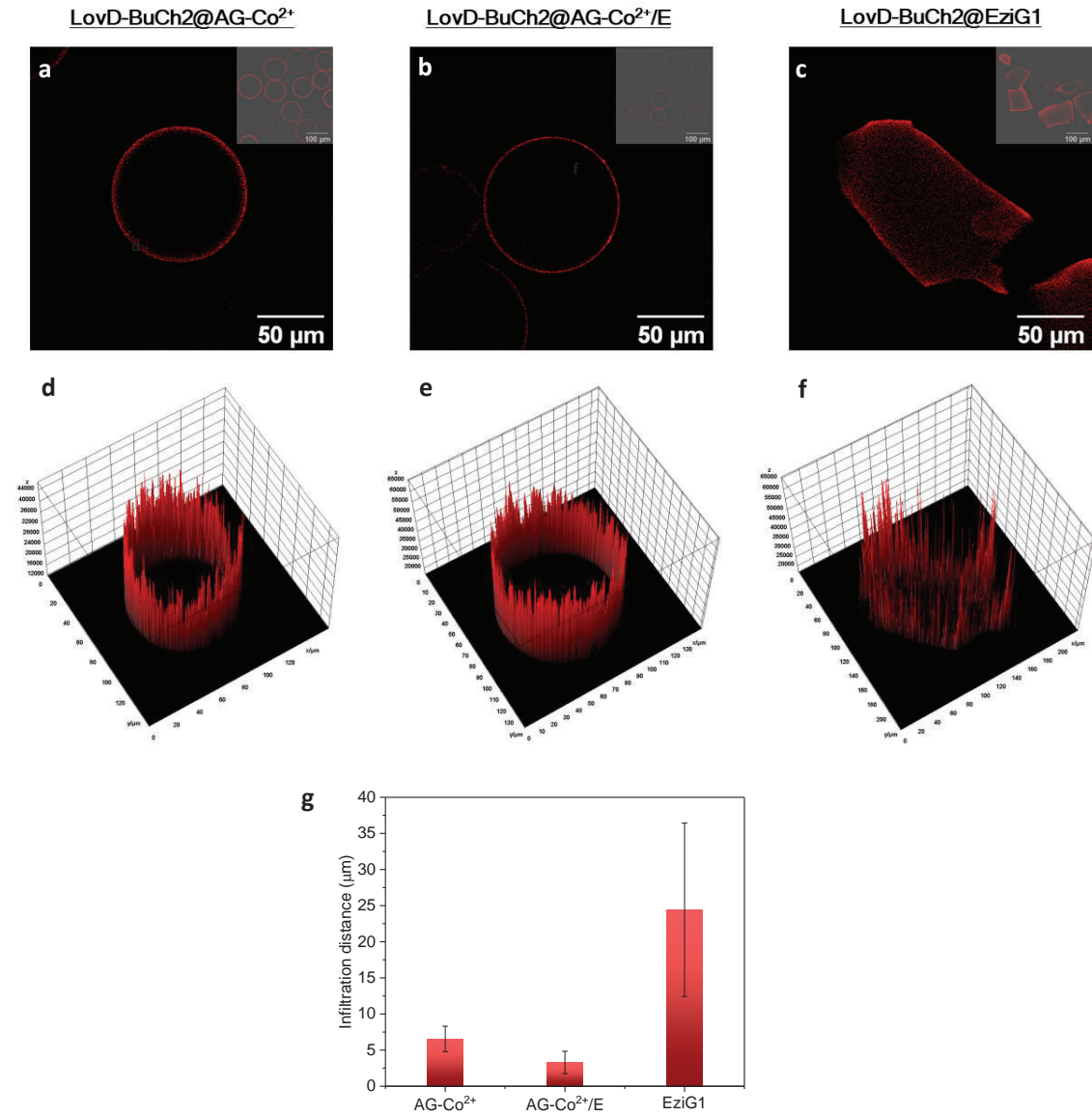
Figure 4. (a-c) Confocal microscope fluorescent images (x40 in the full window and x20 in the right corner, red field) of LovD-BuCh2 labeled with Rhodamine isothiocyanate (RBITC) and immobilized onto cobalt-activated agarose microbeads (AG-Co^{2+}) **(a)**, AG-Co^{2+} functionalized with epoxide groups (AG-Co^{2+}/E) **(b)** and porous glass functionalized with Fe^{3+}-catechol complexes (EziG1) **(c)**. **(d-f)** Fluorescence 3D profiles of single beads of LovD-BuCh2@AG-Co^{2+} **(d)**, LovD-BuCh2@AG-Co^{2+}/E **(e)** and LovD-BuCh2@EziG1 **(f)**. **(g)** Infiltration distance values calculated from fluorescence profiles of 20 single beads of similar size.

7.3.3. Thermal stability analysis

Thermal stability of enzymes is a key parameter for industrial processes in biocatalysis that guarantees the operability of the catalyst under drastic conditions and long reaction times. As discussed above, rigidification of the enzyme backbone by anchoring it to a solid carrier improves its resistance to temperature changes, generally at the expense of decreasing the specific activity of the catalyst.[349,350]

Figure 5.a reveals that LovD-BuCh2@AG-Co^{2+}/E and LovD-BuCh2@EziG1 retained more than 80% of their activity at temperatures up to 45 °C, whereas LovD-BuCh2@AG-Co^{2+} and the free enzyme retained only 50% and 40 % of their activity at 40 °C, respectively. At 50 °C all the biocatalysts are close to be completely inactivated.

Because at 40 °C two out of the four biocatalysts lost around the half of their initial activity, we selected this temperature to perform a kinetic thermal deactivation assay with longer incubation times. **Figure 5.b** shows that, as expected, the soluble enzyme turned to be the most unstable catalyst in terms of residual activity with the higher inactivation constant (k), whereas k for immobilizates presented the following trend: LovD-BuCh2@AG-Co^{2+} >> LovD-BuCh2@EziG1 > LovD-BuCh2@AG-Co^{2+}/E (**Figure 5.c**). This thermal stability kinetic study demonstrates that AG-Co^{2+}/E and EziG1 are good candidates to operate at high temperature conditions. The better retention of activity of EziG1 when compared with AG-Co^{2+} can be due to the major ability to get the enzyme internalized inside the particle, as we have previously seen through CFLM (**Figure 4**), which could guarantee the protection of the enzyme against temperature changes. On the other hand, the retention of activity in AG-Co^{2+}/E could be explained by the rigidification of the protein backbone due to the covalent immobilization occurring on this carrier. Finally, half-life time ($t_{1/2}$) is also a good stabilization descriptor, and it follows the same trend observed for k, being AG-Co^{2+}/E and EziG1 the most long-term stabilized biocatalysts.
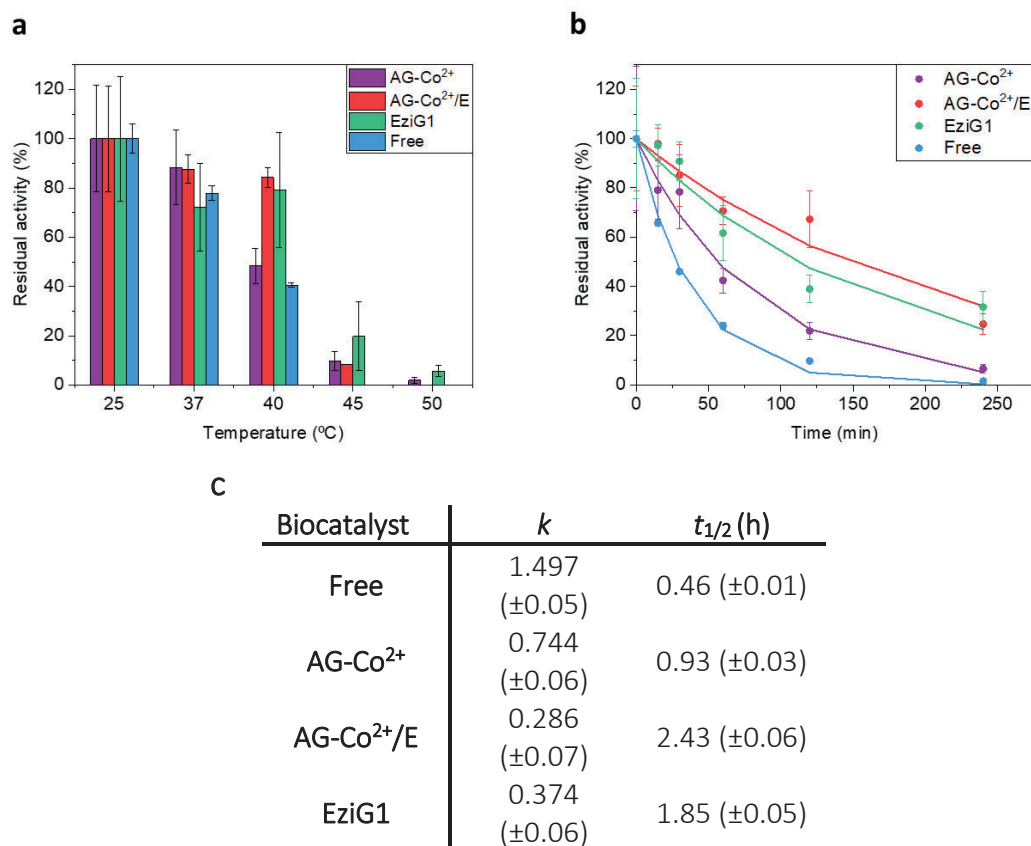
Figure 5. **a.** Thermal inactivation of LovD-BuCh2 both free and immobilized in different carriers at different temperatures. The residual activity (%) was calculated based on the initial activity at 25 °C of each biocatalyst. **b.** Inactivation time-courses of LovD-BuCh2 immobilized in different carriers incubated at 40 °C. All the assays were performed with 1 µM of enzyme; 3 mM MJA and 2 mM DMB-SMMP. **c.** Kinetic inactivation constant (k) and half-life ($t_{1/2}$) of LovD-BuCh2 immobilized in different carriers and incubated at 40 °C.

7.3.4. Structural flexibility and conformational studies on immobilized LovD-BuCh2

Biophysical techniques can be of great use to elucidate the structural rearrangements of the protein upon immobilization in different carriers and along the inactivation process. The intrinsic fluorescence of aromatic residues provides information on their microenvironment within the protein.[351,352] Therefore, structural changes on the protein upon immobilization or when subjected to thermal shock can be detected *via* changes in its fluorescence spectrum (**Figure 6.a-h**).

Figure 6.i shows a slight hypsochromic shift in λ_{max} from 335 nm to 330 nm when LovD-BuCh2 was immobilized in all the carriers and incubated at 25 °C compared to the free enzyme, unveiling a minor solvent exposure of the aromatic residues upon immobilization process that suggest some enzyme conformational changes when interacting with the carriers. Those new conformations may explain the lower specific activity of the immobilized enzyme compared to its free counterpart. When the free and immobilized biocatalysts were

157

incubated at 40 °C for 4 hours, the free enzyme exhibited a 50 nm bathochromic shift λ_{max} because thermal unfolding induces a higher solvent exposure of the aromatic residues. This large shift occurred in a lesser extent for immobilized enzymes: LovD-BuCh2@AG-Co^{2+} displayed a 5 nm red-shift λ_{max}, while LovD-BuCh2@AG-Co^{2+}/E and LovD-BuCh2@EziG1 maintained the same λ_{max} when they are either incubated at 25 °C or 40 °C. These findings prove that temperature increase has a smaller impact in protein structural integrity when the enzyme is immobilized than in its soluble form. Another relevant observation is that besides the alleged rigidification provided by covalent binding in LovD-BuCh2@AG-Co^{2+}/E that might contribute to protein stabilization, EziG1-enzyme interactions also seem to provide a structure more protected to thermal shock. These observations agree with the results of the thermal deactivation assays at 40 °C (**Figure 5.b**), which revealed that EziG1 affinity immobilization improved enzyme thermostability when compared to AG-Co^{2+}.

As shown in **Figure 6.j**, the variation in intrinsic fluorescence upon thermal shock (40 °C) was not the same for all the biocatalysts. 50%, 8% and 12% decrease in maximum fluorescence for the free enzyme, LovD-BuCh2@AG-Co^{2+} and LovD-BuCh2@EziG1, respectively, and 74% increase for LovD-BuCh2@AG-Co^{2+}/E), which could be traduced in a major burial of aromatic residues to the solvent upon thermal incubation. This result suggests that the temperature induced a more compact structure of LovD-BuCh2 when attached to AG-Co^{2+}/E, explaining its higher half-life time compared to the other immobilizates (**Figure 5.c**). Unlike for the soluble enzyme, temperature driven conformational rearrangement tend to pack the conformation of LovD-BuCh2 His-tag oriented and irreversibly attached to AG-Co^{2+}/E.

Besides conformational changes in tertiary structure, the possible changes in the enzyme flexibility upon immobilization process were analyzed through protein fluorescence anisotropy measurements using fluorescein isothiocyanate (FITC) as a probe.[353] **Figure 6.k** shows that the relative anisotropy of the immobilized catalysts was higher than that of the free enzyme, suggesting that immobilization reduced enzyme mobility in all cases. LovD-BuCh2@AG-Co^{2+} exhibited a relative anisotropy two times higher than the free enzyme, whereas the reduction of the mobility of the structure is slightly higher for LovD-BuCh2@EziG1 and LovD-BuCh2@AG-Co^{2+}/E, with relative anisotropy values of 2.45 and 2.94, respectively. This could be a plausible explanation for the recovered activity trends discussed above (**Table 1**). Furthermore, a clear correlation between the relative anisotropy values for the four biocatalysts and their half-life times when incubated 4 h at 40 °C was observed, suggesting that thermostability of the protein is related to its flexibility.

Irreversible and multipoint covalent immobilization promoted by AG-Co^{2+}/E explains the higher compactness and the lower mobility of the immobilized enzymes according to Trp fluorescence and anisotropy studies, entailing the high stability of this biocatalyst. This correlation of the valency of the attachment with the protein deformation and its increasing effect on protein stability under denaturing conditions has already been observed and demonstrated in literature.[354–356]
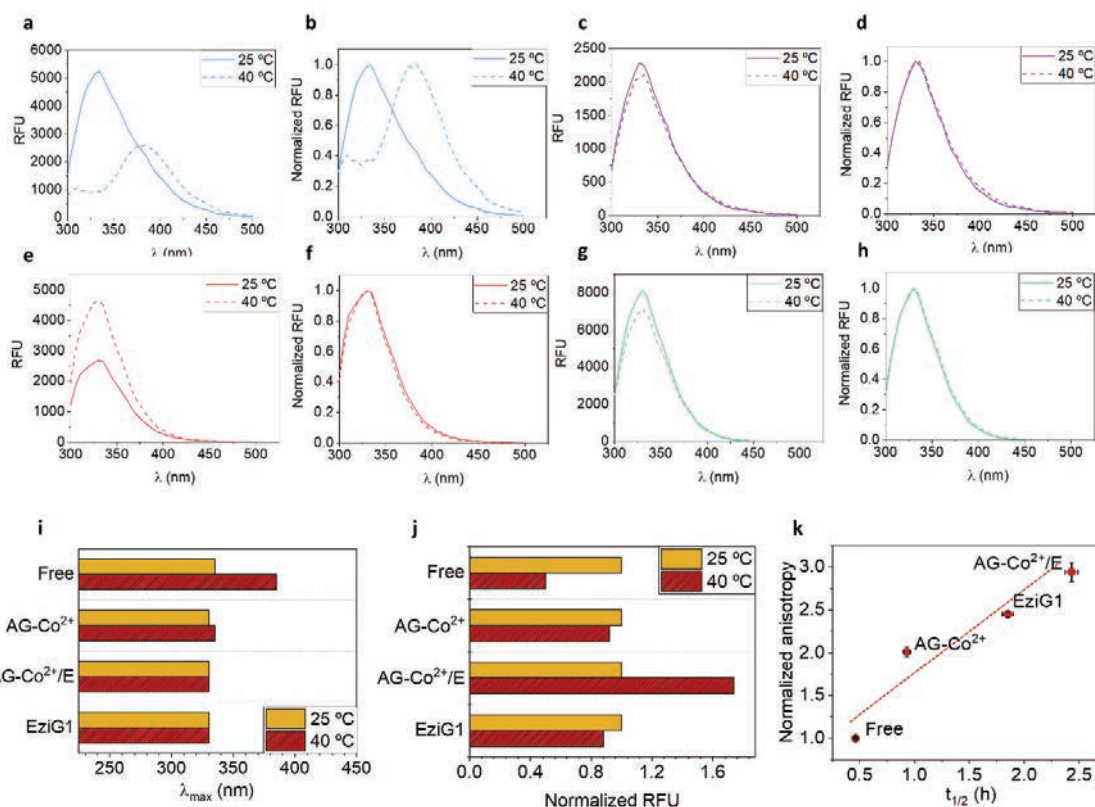
Figure 6. (a-h) Spectra of LovD- BuCh2 intrinsic protein fluorescence (280 nm). **a, c, e** and **g** show relative fluorescence units (RFU) vs λ whereas **b, d, f** and **h** show the spectra normalized to the maximum RFU. **(a, b)** Spectra of free LovD-BuCh2. **(c, d)** Spectra of LovD-BuCh2@AG-Co^{2+}. **(e, f)** Spectra of LovD-BuCh2@AG-Co^{2+}/E. **(g, h)** spectra of LovD-BuCh2@EziG1. **(i)** λ$_{max}$ of each biocatalyst after incubation at 25 °C and 40 °C for 4 h. **(j)** Normalized RFU of each biocatalyst after incubation at 25 °C and 40 °C for 4 h. The RFU before incubation was set to a reference value of 1. **(k).** Normalized anisotropy as a function of the half-life time at 40 °C. The anisotropy of free LovD-BuCh2 was set to a reference value of 1.

7.3.5. LovD-BuCh2@EziG1 thermostability by Raman spectroscopy

Raman spectroscopy has been used for decades to identify the secondary structure and sidechain environment of proteins.[357–360] Spectroscopic applications in structural biology are possible due to robust assignments of the spectral features and correlations between the band positions and intensities and the structure of the biomolecule. As large polypeptides, vibrational spectra show multiple overlapping bands. Nevertheless, the protein backbone provides dominant peaks in the spectra, such as those related to deformation of C-H bonds (δ-CH) at 1447 cm^{-1} and the amide I region at 1656 cm^{-1}. As already reported,[360] Raman spectroscopy is emerging as a useful technique to study protein conformational changes. Raman maps of immobilized enzymes have been already reported as an effective way to study enzyme distribution in the carrier with no need of protein labeling.[361] Moreover, True Component Analysis (TCA) is an outstanding approach to spatially and chemically resolve the different chemical species within the Raman hyperspectral imaging data.[362]

LovD-BuCh2 immobilization in EziG1 has shown the best activity/stability balance and hence we selected this biocatalyst to study it through Raman spectroscopy, which provides a spatio-temporal resolution of the immobilized protein structure when subjected to thermal shock. We used an excitation wavelength of 532 nm to register the spectra of the immobilized enzyme before and after incubation at different temperatures: 40 °C for 4 h and 100 °C for 1 h, with the purpose of imaging temperature induced conformational changes at different microscopic regions of the carrier. Therefore, we constructed Raman maps of EziG1 beads and analyzed them through TCA (**Figure 7**).

Raman scans of free enzyme and EziG1 (as a control) were acquired to identify the aforementioned characteristic bands of the protein present in the free (**Figure 7.a**). Apart from the characteristic protein bands, the signal of C-H stretching (σ-CH) at 2930 cm^{-1} is notably more intense in the spectra corresponding to protein sample than in the EziG1 control, probably due to a much larger number of σ-CH interactions in the former.

We therefore obtained averaged spectra and Raman maps of LovD-BuCh2@EziG1 with a TCA, which was able to identify the protein Raman spectrum as one of the components (**Figure 7.b-j**). This spectrum was characterized by the afore described presence of δ-CH at 1447 cm^{-1} and the amide I region at 1656 cm^{-1}. As already mentioned, σ-CH band was present in all the spectra, but in a higher extent in those corresponding to the protein. LovD-BuCh2@EziG1 TCA generates a map with a non-homogeneous distribution where one chemical specie (corresponding to the protein Raman spectrum shown in **Figure 7.a**) is clearly located at the outer surface of particles (**Figure 7.f**) in agreement with the spatial distribution elicited by CLSM for the same sample (**Figure 4.c**).

The incubation of the immobilized enzyme sample at 40 °C for 4 h revealed a Raman map and average spectrum with a lower intensity of the protein signal (**Figure 7.d,h**), suggesting conformational changes in the protein that could be responsible of its loss in activity observed when incubated at these conditions (**Figure 5**). When the same sample is incubated at 100 °C for 1 h (**Figure 7.d,j**), the Raman signal corresponding to the active configuration of the enzyme was completely lost, supporting the denaturing effect of the temperature on the immobilized protein structure.
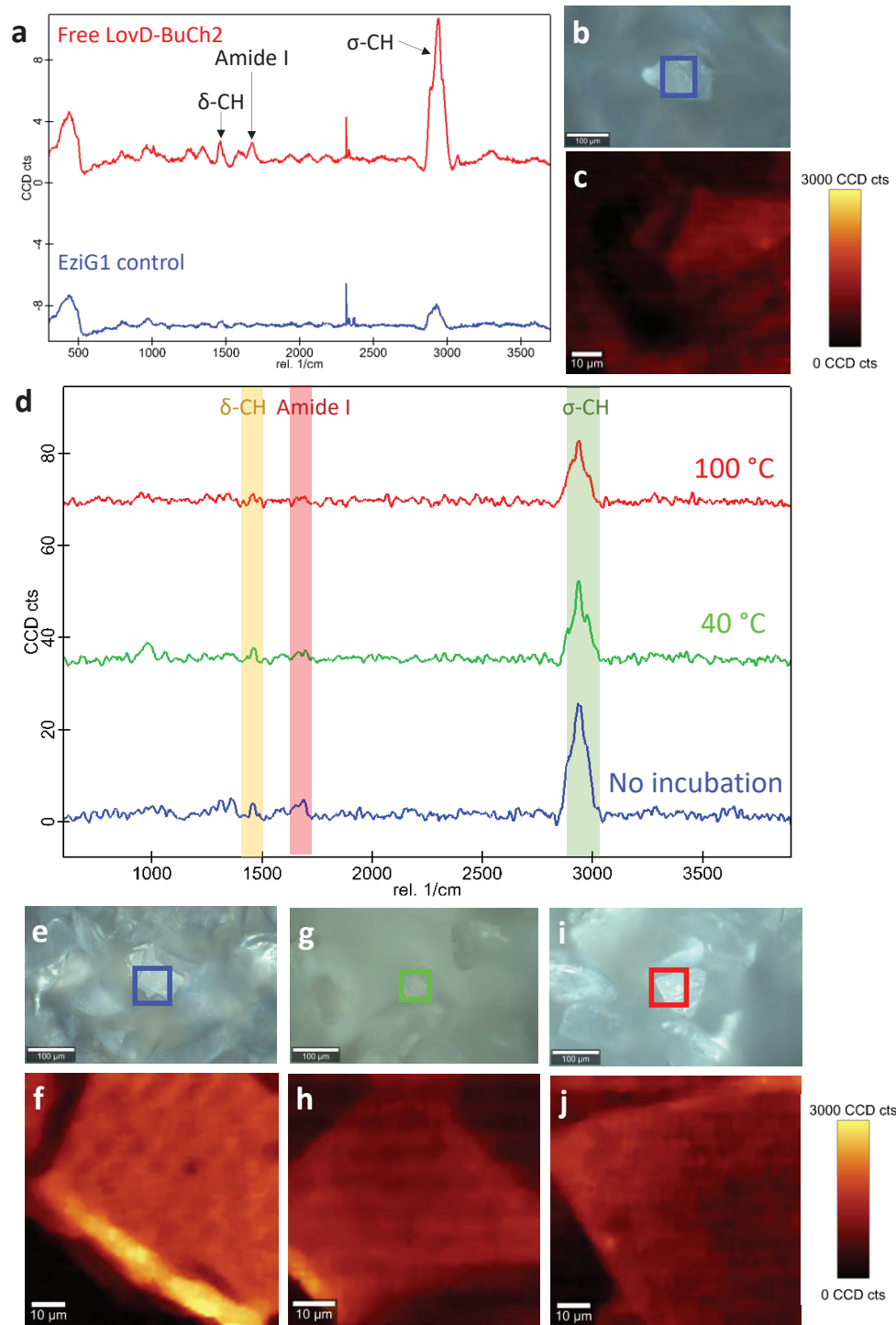
Figure 7. (a) Single Raman spectra of free LovD-BuCh2 (in red) and EziG1 control (in blue). (b) EziG1 control bright-field (BF) Raman image. (c) EziG1 control Raman 2-D map obtained with True Component Analysis (TCA). (d) Raman average spectra of LovD-BuCh2@EziG1 before incubation and after incubation at 40 °C and 100 °C. LovD-BuCh2@EziG1 pre-incubation BF image (e) and Raman map obtained with TCA (f). LovD-BuCh2@EziG1 incubated at 40 °C BF image (g) and Raman map obtained with TCA (h). LovD-BuCh2@EziG1 incubated at 100 °C BF image (i) and Raman map obtained with TCA (j).

161

7.3.6. Operational stability of LovD-BuCh2@EziG1

In addition to thermal stability, enzyme recycling is also fundamental to design an industrially useful biocatalyst in terms of productivity. The efficiency of the enzyme to produce simvastatin (SVA) was studied by measuring batch reaction time-courses with both the free enzyme and the two most thermostable biocatalysts, LovD-BuCh2@AG-Co^{2+}/E and LovD-BuCh2@EziG1 (**Figure 8.a**). SVA yield achieved by the immobilized catalysts was also determined in subsequent cycles to assess enzyme operational stability (**Figure 8.b**).
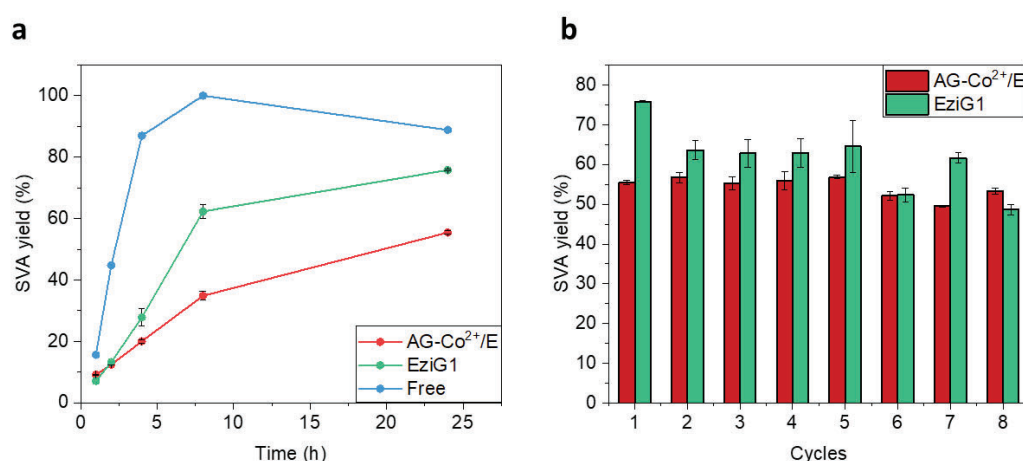
Figure 8. (a) Batch reaction time-courses for free LovD-BuCh2, LovD-BuCh2@AG-Co^{2+}/E and LovD-BuCh2@EziG1 with 1 µM of enzyme, 3 mM Monacolin J acid (MJA) and 2 mM α-dimethylbutyryl-*S*-methyl-mercaptopropionate (DMB-SMMP). **(b)** 24 h batch cycles under the same aforementioned reaction conditions.

As observed in **Figure 8.a**, the free enzyme reached 100% yield of SVA production in 8 h and was the fastest biocatalyst with an initial velocity (v$_0$) of 2.34 nmol mg^{-1} min^{-1}. However, 16 h later product hydrolysis decreased SVA yield to 85%. In contrast, LovD-BuCh2@EziG1 and LovD-BuCh2@AG-Co^{2+}/E exhibited v$_0$ values of 0.80 and 0.29 nmol mg^{-1} min^{-1}, and SVA yields of 75% and 50% after 24 h, respectively. As expected from the immobilization parameters (**Table 1**), the immobilization reduced the initial velocity and 24 h yield when the reaction was catalyzed with the heterogeneous biocatalysts. Those reductions were more dramatic for the enzyme irreversibly immobilized on AG-Co^{2+}/E and aligned with the lower specific activity of the immobilized preparations.

The study of the biocatalyst reusability through 24 h reaction cycles (**Figure 8.b**) shows that both LovD-BuCh2@AG-Co^{2+}/E and LovD-BuCh2@EziG1 exhibited a good operational stability: LovD-BuCh2@AG-Co^{2+}/E maintained almost the same SVA yield (around 50%), while LovD-BuCh2@EziG1 only lowered SVA yield from 75% to 50 % after 8 operation cycles. These data along with the thermostability results (**Figure 5**) suggest that both biocatalysts are stable enough for a continuous flow system for SVA production. Due to the best operational activity/stability balance of LovD-BuCh2@EziG1, we selected this heterogeneous biocatalyst to be exploited for continuous synthesis of SVA.

7.3.7. Continuous synthesis of SVA with a LovD-BuCh2@EziG1 packed-flow reactor

With the aim of producing a continuous synthesis of SVA and having a better control of the reaction, we packed a bed of 1 g of 0.34 mg g^{-1} LovD-BuCh2@EziG1 in a 1 cm^3 column and tested the reaction with 1 mM MJA and 2 mM DMB-SMMP at different flow rates. As shown in **Figure 9.a**, whenever the flow rate was increased, SVA yield was also increased (reaching the maximum yield at 100 µL min^{-1}). As LovD-BuCh2 catalyzes a kinetically controlled synthesis of SVA, unwanted hydrolysis of both SVA and DMB-SMMP affect the final yield obtained with the immobilized enzyme when operating at long residence times, explaining the lower SVA yields at low flow rates. These side reactions were also detected using the soluble enzyme under batch conditions. Kinetically controlled synthesis in biocatalytic continuous systems has already been described in literature as an efficient way to optimize the synthesis/hydrolysis balance in this type of reactions.[344] Looking at the yield rates, we could assume that from 10 to 100 µL min^{-1}, the lower the flow rate, the higher product hydrolysis and the lower SVA synthesis. However, at 200 µL min^{-1} flow rate, the residence time of MJA and DMB-SMMP was too short to allow their maximal diffusion towards and binding to the active sites of the immobilized enzymes, resulting in too low synthetic rate to further maximize the SVA yield. Therefore, 100 µL min^{-1} is the optimal flow rate in terms of SVA yield under the assayed conditions. Nevertheless, the specific productivity of the reactor is doubled from 100 to 200 µL min^{-1} flow rate, which means that in terms of mg of SVA produced per mg of enzyme and per minute, the 200 µL min^{-1} flow rate is the most productive one.

To confirm the effect of flow rate on the hydrolysis/synthesis trade-off, we operated the packed-bed reactor (PBR) with a 1 mM SVA solution at two different flow rates (10 and 100 µL min^{-1}) (**Figure 9.b**). There, we observed that for the same volume passed through the PBR at 10 µL min^{-1} the MJA yield was 54%. When the flow rate was increased to 100 µL min^{-1}, MJA yield was 26%. The comparison of MJA yields in both cases proved that at higher flow rates, product hydrolysis is mitigated and SVA recovery is improved.

After these preliminary assays, we developed a first continuous flow process (CFP-1) attempting to improve the productivity of the reactor by mitigating product hydrolysis. For that, we set a flow rate of 100 µL min^{-1} using 1 g of catalyst (residence time of 10 min) and keeping an excess of DMB-SMMP over MJA (2:1 ratio). Under these conditions, the reactor achieved a productivity of 29.2 mg SVA mL^{-1} min^{-1} and a maximum yield of 70 % after 2 h of operation (**Figure 9.c, f**). CFP-1 maintained SVA yield values close to its maximum for at least 6 reactor cycles of 30 min each. However, total SVA conversion was not achieved, probably because the hydrolysis reaction was still taking place.
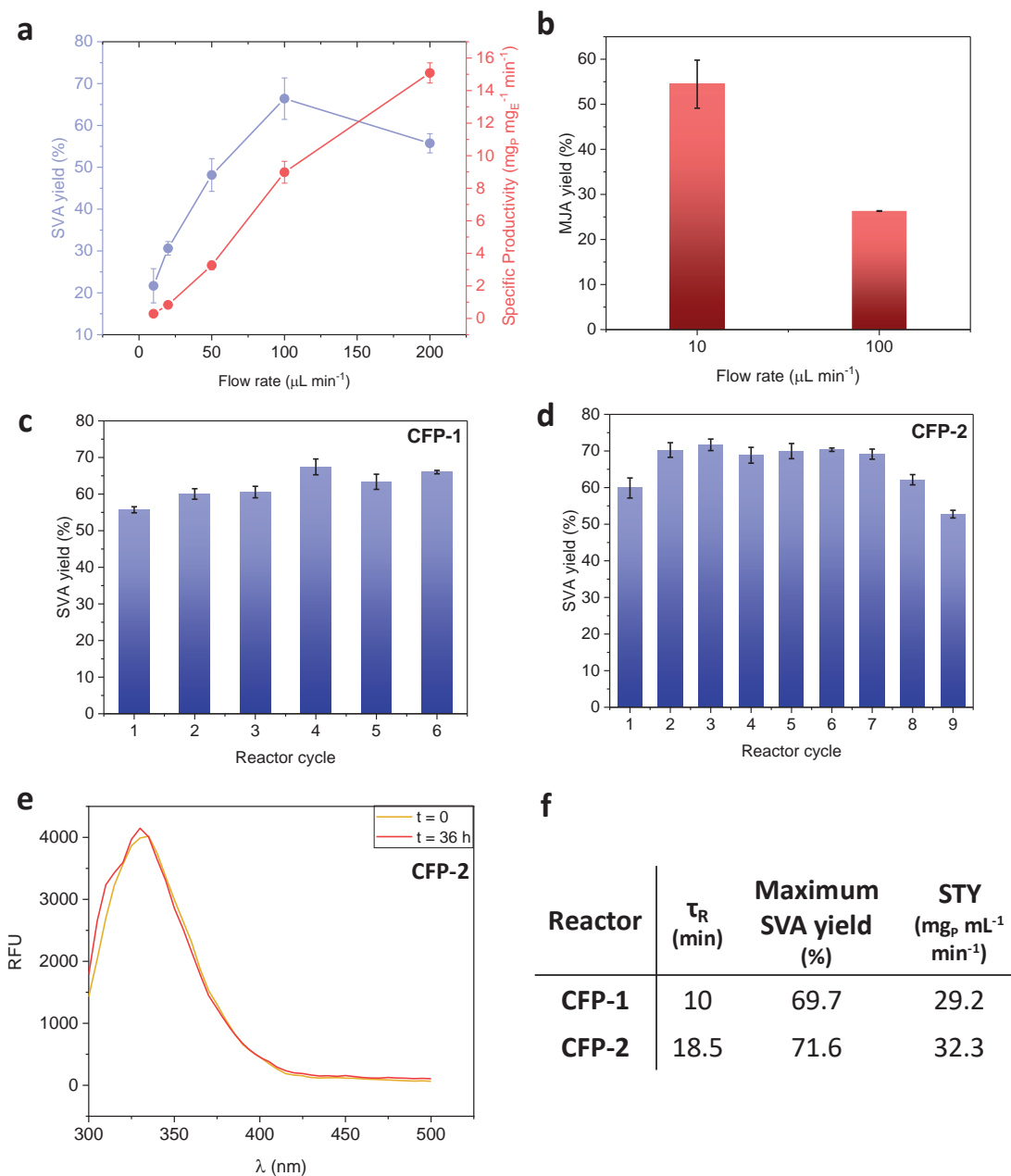
Figure 9. a. Simvastatin acid (SVA) yield (%) and Specific Productivity (in mg product x mg enzyme^{-1} min^{-1}) in a continuous flow process (CFP) at different flow rates (from 10 to 200 μL min^{-1}) with 1 μM of immobilized enzyme, 1 mM MJA and 2 mM dimethylbutyryl-*S*-methyl-mercaptopropionate (DMB-SMMP). **b.** MJA yield (%) after maximum conversion was achieved in a CFP at 10 and 100 μL min^{-1} with 1 μM of enzyme and 1 mM of SVA. **c.** SVA yield of CFP-1 after consecutive 30 min reaction cycles with a flow rate of 100 μL min^{-1} (1 mM MJA and 2 mM DMB-SMMP). **d.** SVA yield of CFP-2 after consecutive 4 h reaction cycles with a flow rate of 20 μL min^{-1} (3 mM MJA and 2 mM DMB-SMMP). **e.** Spectra of intrinsic protein fluorescence of LovD-BuCh2@EziG1 before and after 36 h (9 cycles) of operation under CFP-2 conditions. **f.** CFP parameters for CFP-1 and CFP-2, where τ_R is the residence time and STY is the Space-Time Yield.

164

With the aim of assessing LovD-BuCh2@EziG1 operability under long-term conditions, we set a second continuous flow process (CFP-2) with a flow rate of 20 μL min^{-1} using 300 mg of biocatalyst (residence time of 18.5 min) and a 2:3 DMB-SMMP: MJA ratio. The excess of MJA over DMB-SMMP was envisioned to shift the equilibrium towards SVA synthesis, thus minimizing hydrolysis. Our hypothesis was confirmed since 70% of SVA yield was achieved after 8 hours of operation (**Figure 9.d,f**). Furthermore, similar SVA yields were maintained for 7 reactor cycles of 4 h each, indicating a considerable operational stability of LovD-BuCh2@EziG1 in long-term continuous flow processes. After 36 hours of operation, the yield decayed to 52%, revealing the partial inactivation of the enzyme. This observation was supported through the slight shift in λ_{max} from 340 to 335 nm observed in protein intrinsic fluorescence measurements of the biocatalyst before and after 36 hours of operation (**Figure 9.e**).

Finally, despite operating at different flows, the STY of CFP-1 and -2 is similar, due to the influence of both DMB-SMMP: MJA ratio and flow rate in the kinetically controlled synthesis of SVA: the highest the flow and the lowest the DMB-SMMP: MJA ratio, the highest the STY of the reactor (**Figure 9.f**).

7.4. Conclusions

Three heterogeneous biocatalysts based on the very active variant LovD-BuCh2 of the simvastatin synthase LovD were prepared by immobilizing the enzyme onto three different solid porous carriers, AG-Co^{2+}, AG-Co^{2+}/E and EziG1, with the aim of testing different materials and immobilization strategies. The enzyme was successfully immobilized in all the three materials with considerably high immobilization rates and recovered activities. Confocal imaging studies revealed that in both AG-Co^{2+} and AG-Co^{2+}/E, the enzyme was attached to the carrier in the outer surface, while internalization was larger with EziG1.

Biophysical techniques such as intrinsic protein fluorescence or anisotropy experiments shed light on the stability of the enzyme tertiary structure in different biocatalysts. Enzyme immobilization in both AG-Co^{2+}/E and EziG1 stabilized the enzyme to a larger extent than AG-Co^{2+}, as demonstrated by thermal stability studies. Raman hyperspectral imaging was used to monitor protein structural changes promoted by thermal shock at different carrier microscopic regions.

LovD-BuCh2@AG-Co^{2+}/E and LovD-BuCh2@EziG1 displayed considerable reusability in batch mode, as they maintained high production yields after several operation cycles. The process was tested in continuous flow reactors under different conditions. The detrimental effect of SVA hydrolysis in the global yield was minimized by optimizing the flow rate and acyl donor/acceptor ratio. Finally, LovD-BuCh2@EziG1 showed a long operation time (more than 24 hours) before starting to lose activity, supporting the suitability of this biocatalyst for *in vitro* production of simvastatin.

CAPÍTULO 8:

Conclusiones

En la presente tesis doctoral se ha estudiado el potencial de la aciltransferasa LovD de *Aspergillus terreus* para la síntesis del fármaco anticolesterolémico simvastatina, comercializada con el nombre de Zocor®. Dicho potencial ha sido abordado mediante cuatro aproximaciones diferentes:

1- Se llevó a cabo una **deconvolución parcial** de la variante LovD9, obtenida por evolución dirigida a través de la introducción de 29 mutaciones repartidas en toda la estructura proteica. Estas 29 mutaciones fueron divididas en tres "clusters": las mutaciones internas cercanas al centro activo, las mutaciones localizadas en el canal de entrada del sustrato y las mutaciones superficiales. Para dilucidar el impacto de cada uno de estos "clusters" de mutaciones en las distintas propiedades enzimáticas y en las reacciones laterales subyacentes a este proceso, se produjeron cuatro nuevas variantes jamás reportadas: LovD-Bu, con las 8 mutaciones más internas; LovD-Ch, con las 6 mutaciones del canal de entrada; y LovD-BuCh1 y LovD-BuCh2, con la combinación de las 8 mutaciones internas y las 6 del canal de entrada, y diferenciándose entre sí por la posición 261 (Val en LovD-BuCh1 e His en LovD-BuCh2). En este estudio se observó que las mutaciones internas son fundamentales para mejorar el paso de transferencia del grupo acilo del complejo acil-enzima al ácido monacolínico J (MJA), mientras que las mutaciones en el canal de entrada mitigan la hidrólisis del producto final (simvastatina). Además, la polaridad de la posición 261 localizada en el canal de entrada parece ser de gran relevancia para la acilación de la enzima, siendo la variante con una His en esta posición (LovD-BuCh2) más activa que la variante con una Val (LovD-BuCh1). Finalmente, LovD-BuCh2 presentó unos valores de termoestabilidad comparables a los de la variante obtenida por evolución dirigida LovD9, indicando un efecto cumulativo de los dos "clusters" de mutaciones propuestos.

2- Se trató de obtener nuevas variantes de LovD mediante un **diseño *in silico***, como alternativa a la evolución dirigida. Para ello, se propusieron tres métodos computacionales diferentes, todos ellos apoyados en dinámicas moleculares (MD) y en la correcta configuración del centro activo como método de selección:

a) Se seleccionaron las 29 posiciones totales y las 8 posiciones internas de LovD9 para realizar mutagénesis *in silico* con el "software" RosettaDesign en un proceso iterativo con MD para estabilizar las estructuras. Tras esto, se realizaron MD más largas para estudiar la conformación catalítica del centro activo y aplicarlo como criterio de selección de los mutantes más prometedores. De los 10 mutantes seleccionados, 8 se expresaron en forma soluble y todos ellos mostraron capacidad para aceptar el donador del grupo acilo α-dimetilbutiril *(S)*-metil-mercaptopropanoato DMB-SMMP e hidrolizarlo. Sin embargo, ninguno presentó actividad de síntesis de simvastatina, probablemente debido a problemas de estabilidad estructural.

b) Se identificaron nuevos "hotspots" de mutaciones mediante un análisis dinámico de redes de alostería en el cual se seleccionaron los residuos cuya movilidad correlacionaba mejor con la configuración del centro activo. Para introducir las mutaciones también se hizo uso del método de RosettaDesign. De entre los dos nuevos "hotspots" identificados, se aisló varios mutantes (N43R, A123P, D119S/A123P y N43T/D119S/A123P), algunos de ellos ya identificados por evolución dirigida, de 2 a 6 veces más activa que la enzima "wild-type".

c) Se combinó un análisis de intermediación (basado en las redes de alostería calculadas) con un análisis de homología de secuencia. El cambio más relevante identificado con este método fue F363W, que consigue aumentar 2 veces la capacidad tanto de síntesis como de hidrólisis de simvastatina con respecto a la enzima "wild-type".

3- Se exploró la promiscuidad de la enzima evolucionada LovD9 frente a nuevos sustratos, tanto donadores como aceptores del grupo acilo, para la obtención de simvastatina y sus análogos. Se identificaron los *p*-nitrofenil ésteres como altamente reactivos y específicos; sin embargo, la mayor parte del complejo acil-enzima es hidrolizado y no llega a transferirse a MJA. No obstante, el sustrato *p*-nitrofenil 2,2-dimetilbutirato alcanzó niveles de conversión similares a los de DMB-SMMP. Por otro lado, los vinil ésteres mostraron una baja afinidad, pero una muy alta reactividad, por lo que también pueden ser candidatos prometedores si se utilizan en un exceso con respecto a MJA en la reacción. El mecanismo de acilación de los distintos sustratos utilizados fue estudiado mediante QM, revelando un cambio del mecanismo de acilación canónico en dos pasos que ocurre en tioésteres y vinil ésteres a un mecanismo concertado para los *p*-nitrofenil ésteres. Estos últimos también presentaron propiedades cinéticas y termodinámicas superiores, de acuerdo con los resultados experimentales.

4- Se desarrollaron biocatalizadores heterogéneos en diferentes soportes (agarosa-cobalto, AG-Co^{2+}; agarosa-cobalto-epóxido, AG-Co^{2+}/E; y el soporte comercial EziG1) con el diseño obtenido por deconvolución LovD-BuCh2. La enzima mostró una mayor actividad recuperada relativa a la enzima soluble cuando fue inmovilidad en EziG1 (83 %). Mediante microscopía confocal de fluorescencia, se observó que la enzima se inmovilizó de forma superficial en AG-Co^{2+} y en AG-Co^{2+}/E, mientras que la infiltración fue mayor en EziG1. Se llevaron a cabo diferentes estudios biofísicos, así como ensayos de termoestabilidad y espectroscopía Raman para observar que la estabilización de la enzima ocurrió en mayor medida para los biocatalizadores realizados con AG-Co^{2+}/E y EziG1. Además, estos dos biocatalizadores también mostraron una elevada estabilidad operacional, manteniendo más del 50 % de su actividad inicial después de 8 ciclos de 24 h de reacción. Por todo ello, se escogió el biocatalizador de EziG1 para llevar a cabo un ensayo de producción de simvastatina en flujo. Se consiguió minimizar la hidrólisis de simvastatina mediante la optimización del flujo de reacción y de la ratio de los sustratos MJA/DMB-SMMP. Este sistema

presentó un elevado tiempo de operación, manteniendo su actividad máxima durante más de 24 h de reacción.

REFERENCES

1. Muller, P. Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994). *Pure Appl. Chem.* **66**, 1077–1184 (1994).

2. Masel, R. I. *Chemical kinetics and catalysis / Richard I. Masel.* (Wiley-Interscience, 2001).

3. Milas, N. A. SOME STUDIES ON HOMOGENEOUS CATALYSIS. *Proc. Natl. Acad. Sci.* **15**, 596 LP – 601 (1929).

4. García-Álvarez, J. Special Issue: "Advances in Homogeneous Catalysis". *Molecules* **25**, (2020).

5. Shelke, Y. G., Yashmeen, A., Gholap, A. V. A., Gharpure, S. J. & Kapdi, A. R. Homogeneous Catalysis: A Powerful Technology for the Modification of Important Biomolecules. *Chem. – An Asian J.* **13**, 2991–3013 (2018).

6. Heitbaum, M., Glorius, F. & Escher, I. Asymmetric Heterogeneous Catalysis. *Angew. Chemie Int. Ed.* **45**, 4732–4762 (2006).

7. Liang, J., Liang, Z., Zou, R. & Zhao, Y. Heterogeneous Catalysis in Zeolites, Mesoporous Silica, and Metal–Organic Frameworks. *Adv. Mater.* **29**, 1701139 (2017).

8. Védrine, J. C. Metal Oxides in Heterogeneous Oxidation Catalysis: State of the Art and Challenges for a More Sustainable World. *ChemSusChem* **12**, 577–588 (2019).

9. Schlögl, R. Heterogeneous Catalysis. *Angew. Chemie Int. Ed.* **54**, 3465–3520 (2015).

10. Su, D. S. *et al.* Metal-Free Heterogeneous Catalysis for Sustainable Chemistry. *ChemSusChem* **3**, 169–180 (2010).

11. MacMillan, D. W. C. The advent and development of organocatalysis. *Nature* **455**, 304–308 (2008).

12. Terrett, J. A., Clift, M. D. & MacMillan, D. W. C. Direct β-Alkylation of Aldehydes via Photoredox Organocatalysis. *J. Am. Chem. Soc.* **136**, 6858–6861 (2014).

13. Oscar, L. On the Chemical Nature of Enzymes. *Science (80-.).* **10**, 955–961 (1899).

14. H., K. J. On the Vital Activity of the Enzymes. *Science (80-.).* **13**, 765–771 (1901).

15. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).

16.   Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 399–412 (2005).

17.   Messmore, J. M., Fuchs, D. N. & Raines, R. T. Ribonuclease A: Revealing Structure-Function Relationships with Semisynthesis. *J. Am. Chem. Soc.* **117**, 8057–8060 (1995).

18.   Jollès, P. & Perutz, M. F. Relationship between chemical structure and biological activity of hen egg-white lysozyme and lysozymes of different species. *Proc. R. Soc. London. Ser. B. Biol. Sci.* **167**, 350–364 (1967).

19.   Enzyme nomenclature: Recommendations (1992) of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology. Pp 862. Academic Press, San Diego. 1992 ISBN 0-12-227165-3. *Biochem. Educ.* **21**, 102 (1993).

20.   Dagmar, R. & A., P. G. How Enzymes Work. *Science (80-.).* **320**, 1428–1429 (2008).

21.   Pravda, L. *et al.* Anatomy of enzyme channels. *BMC Bioinformatics* **15**, 379 (2014).

22.   Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der Dtsch. Chem. Gesellschaft* **27**, 2985–2993 (1894).

23.   Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci.* **44**, 98 LP – 104 (1958).

24.   Johnson, K. A. & Goody, R. S. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**, 8264–8269 (2011).

25.   Luk, L. Y. P., Loveridge, E. J. & Allemann, R. K. Protein motions and dynamic effects in enzyme catalysis. *Phys. Chem. Chem. Phys.* **17**, 30817–30827 (2015).

26.   Loria, J. P., Berlow, R. B. & Watt, E. D. Characterization of Enzyme Motions by Solution NMR Relaxation Dispersion. *Acc. Chem. Res.* **41**, 214–221 (2008).

27.   Kohen, A. Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Acc. Chem. Res.* **48**, 466–473 (2015).

28.   Yamada, H. & Kobayashi, M. Nitrile Hydratase and Its Application to Industrial Production of Acrylamide. *Biosci. Biotechnol. Biochem.* **60**, 1391–1400 (1996).

29.   Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351 (2002).

30.   Jensen, V. J. & Rugh, S. Industrial-scale production and application of immobilized glucose isomerase. *Immobilized Enzymes and Cells, Part C* 356–370 (1987).

31.	Buchholz, K. A breakthrough in enzyme technology to fight penicillin resistance—industrial application of penicillin amidase. *Appl. Microbiol. Biotechnol.* **100**, 3825–3839 (2016).

32.	Vojcic, L. *et al.* Advances in protease engineering for laundry detergents. *N. Biotechnol.* **32**, 629–634 (2015).

33.	Gröger, H. Enzyme catalysis in the synthesis of pharmaceuticals. *Bioorg. Med. Chem.* **26**, 1239–1240 (2018).

34.	Falck, G. & Müller, K. M. Enzyme-Based Labeling Strategies for Antibody–Drug Conjugates and Antibody Mimetics. *Antibodies* **7**, (2018).

35.	Adsul, M. *et al.* Designing a cellulolytic enzyme cocktail for the efficient and economical conversion of lignocellulosic biomass to biofuels. *Enzyme Microb. Technol.* **133**, 109442 (2020).

36.	Verho, O. & Bäckvall, J.-E. Chemoenzymatic Dynamic Kinetic Resolution: A Powerful Tool for the Preparation of Enantiomerically Pure Alcohols and Amines. *J. Am. Chem. Soc.* **137**, 3996–4009 (2015).

37.	Jiao, S., Li, F., Yu, H. & Shen, Z. Advances in acrylamide bioproduction catalyzed with Rhodococcus cells harboring nitrile hydratase. *Appl. Microbiol. Biotechnol.* **104**, 1001–1012 (2020).

38.	Bell, E. L. *et al.* Biocatalysis. *Nat. Rev. Methods Prim.* **1**, 46 (2021).

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

40.	Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409**, 258–268 (2001).

41.	Woodcock, L. L. *et al.* Enzymatic synthesis of a series of alkyl esters using novozyme 435 in a packed-bed, miniaturized, continuous flow reactor. *Biocatal. Biotransformation* **26**, 466–472 (2008).

42.	Kortemme, T. *et al.* Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379 (2004).

43.	Giver, L., Gershenson, A., Freskgard, P.-O. & Arnold, F. H. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci.* **95**, 12809 LP – 12813 (1998).

44.	Annika, R. & Alexander, S. Acyltransferases in Bacteria. *Microbiol. Mol. Biol. Rev.* **77**,

277–321 (2013).

45.   Kruis, A. J. *et al.* Microbial production of short and medium chain esters: Enzymes, pathways, and applications. *Biotechnol. Adv.* **37**, 107407 (2019).

46.   Hoyos, P., Pace, V. & Alcántara, A. R. Biocatalyzed Synthesis of Statins: A Sustainable Strategy for the Preparation of Valuable Drugs. *Catalysts* **9**, (2019).

47.   Jonathan, K. *et al.* Modulation of Polyketide Synthase Activity by Accessory Proteins During Lovastatin Biosynthesis. *Science (80-.).* **284**, 1368–1372 (1999).

48.   Zeymer, C. & Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).

49.   Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).

50.   Cox, E. C. BACTERIAL MUTATOR GENES AND THE CONTROL OF SPONTANEOUS MUTATION. *Annu. Rev. Genet.* **10**, 135–156 (1976).

51.   Greener, A., Callahan, M. & Jerpseth, B. An efficient random mutagenesis technique using anE.coli mutator strain. *Mol. Biotechnol.* **7**, 189–195 (1997).

52.   Leung, D. W., Chen, E. Y. & Goeddel, D. V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. in (1989).

53.   Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. An Approach to Random Mutagenesis of DNA Using Mixtures of Triphosphate Derivatives of Nucleoside Analogues. *J. Mol. Biol.* **255**, 589–603 (1996).

54.   Eckert, K. A. & Kunkel, T. A. High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase . *Nucleic Acids Res.* **18**, 3739–3744 (1990).

55.   Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* **5**, 939–942 (2008).

56.   Wells, J. A., Vasser, M. & Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **34**, 315–323 (1985).

57.   Reetz, M. T. *et al.* Expanding the Substrate Scope of Enzymes: Combining Mutations Obtained by CASTing. *Chem. − A Eur. J.* **12**, 6031–6038 (2006).

58.   Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).

59. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).

60. Ostermeier, M., Shim, J. H. & Benkovic, S. J. A combinatorial approach to hybrid enzymes independent of DNA homology . *Nat. Biotechnol.* **17**, 1205–1209 (1999).

61. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science (80-.).* **339**, 307–310 (2013).

62. Watt, A. P., Morrison, D., Locker, K. L. & Evans, D. C. Higher Throughput Bioanalysis by Automation of a Protein Precipitation Assay Using a 96-Well Format with Detection by LC–MS/MS. *Anal. Chem.* **72**, 979–984 (2000).

63. Behrendorff, J. B. Y. H., Vickers, C. E., Chrysanthopoulos, P. & Nielsen, L. K. 2,2-Diphenyl-1-picrylhydrazyl as a screening tool for recombinant monoterpene biosynthesis. *Microb. Cell Fact.* **12**, 76 (2013).

64. Mack, M., Burger, M., Pietschmann, P. & Hock, B. A high-throughput microtiter plate-based screening method for the detection of full-length recombinant proteins. *Protein Expr. Purif.* **61**, 92–98 (2008).

65. He, Y.-C., Ma, C.-L., Xu, J.-H. & Zhou, L. A high-throughput screening strategy for nitrile-hydrolyzing enzymes based on ferric hydroxamate spectrophotometry. *Appl. Microbiol. Biotechnol.* **89**, 817–823 (2011).

66. J., F. M. Electronic Separation of Biological Cells by Volume. *Science (80-.).* **150**, 910–911 (1965).

67. Yang, G. & Withers, S. G. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *Chembiochem* **10**, 2704–2715 (2009).

68. Lee, S. Y., Choi, J. H. & Xu, Z. Microbial cell-surface display. *Trends Biotechnol.* **21**, 45–52 (2003).

69. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).

70. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* **108**, 11399 LP – 11404 (2011).

71. Catherine, C., Lee, K.-H., Oh, S.-J. & Kim, D.-M. Cell-free platforms for flexible

expression and screening of enzymes. *Biotechnol. Adv.* **31**, 797–803 (2013).

72.   Stapleton, J. A. & Swartz, J. R. Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. *PLoS One* **5**, e15275–e15275 (2010).

73.   Speight, R. E., Hart, D. J., Sutherland, J. D. & Blackburn, J. M. A new plasmid display technology for the in vitro selection of functional phenotype–genotype linked proteins. *Chem. Biol.* **8**, 951–965 (2001).

74.   Bessette, P. H., Rice, J. J. & Daugherty, P. S. Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng. Des. Sel.* **17**, 731–739 (2004).

75.   McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).

76.   Fernandez-Gacio, A., Uguen, M. & Fastrez, J. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **21**, 408–414 (2003).

77.   Keefe, A. D. Protein Selection Using mRNA Display. *Curr. Protoc. Mol. Biol.* **53**, 24.5.1-24.5.34 (2001).

78.   Hanes, J. & Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci.* **94**, 4937 LP – 4942 (1997).

79.   Amstutz, P. *et al.* In Vitro Selection for Catalytic Activity with Ribosome Display. *J. Am. Chem. Soc.* **124**, 9396–9403 (2002).

80.   Hall, B. G. Predicting evolution by in vitro evolution requires determining evolutionary pathways. *Antimicrob. Agents Chemother.* **46**, 3035–3038 (2002).

81.   Firestine, S. M., Salinas, F., Nixon, A. E., Baker, S. J. & Benkovic, S. J. Using an AraC-based three-hybrid system to detect biocatalysts in vivo. *Nat. Biotechnol.* **18**, 544–547 (2000).

82.   Dixon, N. *et al.* Reengineering orthogonally selective riboswitches. *Proc. Natl. Acad. Sci.* **107**, 2830 LP – 2835 (2010).

83.   Mate, D. M. *et al.* Thermostability enhancement of the Pseudomonas fluorescens esterase I by in vivo folding selection in Thermus thermophilus. *Biotechnol. Bioeng.* **117**, 30–38 (2020).

84.   Rothman, S. C., Voorhies, M. & Kirsch, J. F. Directed evolution relieves product inhibition and confers in vivo function to a rationally designed tyrosine

aminotransferase. *Protein Sci.* **13**, 763–772 (2004).

85. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4552–4557 (2001).

86. Ellefson, J. W. *et al.* Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **32**, 97–101 (2014).

87. Garnier, J. Protein structure prediction. *Biochimie* **72**, 513–524 (1990).

88. Ginalski, K. Comparative modeling for protein structure prediction. *Curr. Opin. Struct. Biol.* **16**, 172–177 (2006).

89. David, R., Korenberg, M. J. & Hunter, I. W. 3D-1D threading methods for protein fold recognition. *Pharmacogenomics* **1**, 445–455 (2000).

90. Fox, R. J. *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **25**, 338–344 (2007).

91. Morris, G. M., Huey, R. & Olson, A. J. Using AutoDock for Ligand-Receptor Docking. *Curr. Protoc. Bioinforma.* **24**, 8.14.1-8.14.40 (2008).

92. Notonier, S., Gricman, Ł., Pleiss, J. & Hauer, B. Semirational Protein Engineering of CYP153AM.aq.-CPRBM3 for Efficient Terminal Hydroxylation of Short- to Long-Chain Fatty Acids. *ChemBioChem* **17**, 1550–1557 (2016).

93. Maria-Solano, M. A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J. & Osuna, S. Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.* **54**, 6622–6634 (2018).

94. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**, 808–814 (2009).

95. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6679 LP – 6685 (2005).

96. Salsbury, F. R. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr. Opin. Pharmacol.* **10**, 738–744 (2010).

97. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304 (2008).

98. Ponder, J. W. & Case, D. A. B. T.-A. in P. C. Force Fields for Protein Simulations. in

Protein Simulations **66**, 27–85 (Academic Press, 2003).

99.   MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

100.   Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **3**, 198–210 (2013).

101.   Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988).

102.   Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004).

103.   Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **6**, e19230 (2011).

104.   Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A. & Schulten, K. Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure* **14**, 437–449 (2006).

105.   E., S. D. *et al.* Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (80-.).* **330**, 341–346 (2010).

106.   Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140–150 (1997).

107.   Olsson, S. & Noé, F. Mechanistic Models of Chemical Exchange Induced Relaxation in Protein NMR. *J. Am. Chem. Soc.* **139**, 200–210 (2017).

108.   Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466**, 756–760 (2010).

109.   Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120**, 11919–11929 (2004).

110.   Chen, J., Yin, B., Wang, W. & Sun, H. Effects of Disulfide Bonds on Binding of Inhibitors to β-Amyloid Cleaving Enzyme 1 Decoded by Multiple Replica Accelerated Molecular Dynamics Simulations. *ACS Chem. Neurosci.* **11**, 1811–1826 (2020).

111.   Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci.* **99**, 12562

LP – 12566 (2002).

112. Kuzmanic, A. *et al.* Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *Elife* **6**, e22175 (2017).

113. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).

114. Yang, M. & MacKerell, A. D. Conformational Sampling of Oligosaccharides Using Hamiltonian Replica Exchange with Two-Dimensional Dihedral Biasing Potentials and the Weighted Histogram Analysis Method (WHAM). *J. Chem. Theory Comput.* **11**, 788–799 (2015).

115. Lan, N. T. *et al.* Prediction of AChE-ligand affinity using the umbrella sampling simulation. *J. Mol. Graph. Model.* **93**, 107441 (2019).

116. Monza, E. *et al.* Insights into Laccase Engineering from Molecular Simulations: Toward a Binding-Focused Strategy. *J. Phys. Chem. Lett.* **6**, 1447–1453 (2015).

117. Rivalta, I. *et al.* Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci.* **109**, E1428 LP-E1436 (2012).

118. Romero-Rivera, A., Garcia-Borràs, M. & Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **7**, 8524–8532 (2017).

119. Osuna, S., Jiménez-Osés, G., Noey, E. L. & Houk, K. N. Molecular Dynamics Explorations of Active Site Structure in Designed and Evolved Enzymes. *Acc. Chem. Res.* **48**, 1080–1089 (2015).

120. Clarke, J. & Fersht, A. R. Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32**, 4322–4329 (1993).

121. Pikkemaat, M. G., Linssen, A. B. M., Berendsen, H. J. C. & Janssen, D. B. Molecular dynamics simulations as a tool for improving protein stability. *Protein Eng. Des. Sel.* **15**, 185–192 (2002).

122. Reetz, M. T., Carballeira, J. D. & Vogel, A. Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability. *Angew. Chemie Int. Ed.* **45**, 7745–7751 (2006).

123. Acharya, P., Rajakumara, E., Sankaranarayanan, R. & Rao, N. M. Structural Basis of Selection and Thermostability of Laboratory Evolved Bacillus subtilis Lipase. *J. Mol.*

Biol. **341**, 1271–1281 (2004).

124. Augustyniak, W. *et al.* Biophysical characterization of mutants of Bacillus subtilis lipase evolved for thermostability: Factors contributing to increased activity retention. *Protein Sci.* **21**, 487–497 (2012).

125. Kamal, M. Z. *et al.* In Vitro Evolved Non-Aggregating and Thermostable Lipase: Structural and Thermodynamic Investigation. *J. Mol. Biol.* **413**, 726–741 (2011).

126. Rathi, P. C., Jaeger, K.-E. & Gohlke, H. Structural Rigidity and Protein Thermostability in Variants of Lipase A from Bacillus subtilis. *PLoS One* **10**, e0130289 (2015).

127. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).

128. Mulholland, A. J. Computational enzymology: modelling the mechanisms of biological catalysts. *Biochem. Soc. Trans.* **36**, 22–26 (2008).

129. Vasilevskaya, T., Khrenova, M. G., Nemukhin, A. V & Thiel, W. Methodological aspects of QM/MM calculations: A case study on matrix metalloproteinase-2. *J. Comput. Chem.* **37**, 1801–1809 (2016).

130. Vidal-Limón, A., Águila, S., Ayala, M., Batista, C. V & Vazquez-Duhalt, R. Peroxidase activity stabilization of cytochrome P450BM3 by rational analysis of intramolecular electron transfer. *J. Inorg. Biochem.* **122**, 18–26 (2013).

131. Frushicheva, M. P. & Warshel, A. Towards Quantitative Computer-Aided Studies of Enzymatic Enantioselectivity: The Case of Candida antarctica Lipase A. *ChemBioChem* **13**, 215–223 (2012).

132. Velasco-Lozano, S. *et al.* Selective oxidation of alkyl and aryl glyceryl monoethers catalysed by an engineered and immobilised glycerol dehydrogenase. *Chem. Sci.* **11**, 12009–12020 (2020).

133. Silva, J. R. A., Roitberg, A. E. & Alves, C. N. Catalytic Mechanism of L,D-Transpeptidase 2 from Mycobacterium tuberculosis Described by a Computational Approach: Insights for the Design of New Antibiotics Drugs. *J. Chem. Inf. Model.* **54**, 2402–2410 (2014).

134. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational Enzyme Design. *Angew. Chemie Int. Ed.* **52**, 5700–5725 (2013).

135. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

136. Weitzner, B. D., Kipnis, Y., Daniel, A. G., Hilvert, D. & Baker, D. A computational method for design of connected catalytic networks in proteins. *Protein Sci.* **28**, 2036–2041 (2019).

137. Khersonsky, O. *et al.* Evolutionary Optimization of Computationally Designed Enzymes: Kemp Eliminases of the KE07 Series. *J. Mol. Biol.* **396**, 1025–1042 (2010).

138. Khersonsky, O. *et al.* Optimization of the In-Silico-Designed Kemp Eliminase KE70 by Computational Design and Directed Evolution. *J. Mol. Biol.* **407**, 391–412 (2011).

139. Khersonsky, O. *et al.* Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci.* **109**, 10358 LP – 10363 (2012).

140. Lin, J. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science (80-.).* **319**, 1387–1391 (2008).

141. Garrabou, X., Beck, T. & Hilvert, D. A Promiscuous De Novo Retro-Aldolase Catalyzes Asymmetric Michael Additions via Schiff Base Intermediates. *Angew. Chemie Int. Ed.* **54**, 5609–5612 (2015).

142. Linder, M., Johansson, A. J., Olsson, T. S. G., Liebeschuetz, J. & Brinck, T. Computational design of a Diels–Alderase from a thermophilic esterase: the importance of dynamics. *J. Comput. Aided. Mol. Des.* **26**, 1079–1095 (2012).

143. Basler, S. *et al.* Efficient Lewis acid catalysis of an abiological reaction in a de novo protein scaffold. *Nat. Chem.* **13**, 231–235 (2021).

144. Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural Enzyme Evolution and Rational Enzyme Engineering. *Biochemistry* **51**, 3933–3940 (2012).

145. Koebke, K. J. & Pecoraro, V. L. Development of de Novo Copper Nitrite Reductases: Where We Are and Where We Need To Go. *ACS Catal.* **8**, 8046–8057 (2018).

146. Siedhoff, N. E., Schwaneberg, U. & Davari, M. D. Chapter Twelve - Machine learning-assisted enzyme engineering. in *Enzyme Engineering and Evolution: General Methods* (ed. Tawfik, D. S. B. T.-M. in E.) **643**, 281–315 (Academic Press, 2020).

147. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci.* **116**, 8852 LP – 8858 (2019).

148. Jegannathan, K. R., Abang, S., Poncelet, D., Chan, E. S. & Ravindra, P. Production of

Biodiesel Using Immobilized Lipase—A Critical Review. *Crit. Rev. Biotechnol.* **28**, 253–264 (2008).

149. Joshi, K. A. *et al.* A Disposable Biosensor for Organophosphorus Nerve Agents Based on Carbon Nanotubes Modified Thick Film Strip Electrode. *Electroanalysis* **17**, 54–58 (2005).

150. Jesionowski, T., Zdarta, J. & Krajewska, B. Enzyme immobilization by adsorption: a review. *Adsorption* **20**, 801–821 (2014).

151. Nelson, J. M. & Griffin, E. G. ADSORPTION OF INVERTASE. *J. Am. Chem. Soc.* **38**, 1109–1115 (1916).

152. Mateo, C., Abian, O., Fernandez-Lafuente, R. & Guisan, J. M. Reversible enzyme immobilization via a very strong and nondistorting ionic adsorption on support–polyethylenimine composites. *Biotechnol. Bioeng.* **68**, 98–105 (2000).

153. Lalonde, J. & Margolin, A. *Immobilization of enzymes*. *Enzyme Catalysis in Organic Synthesis: Second Edition* (2008).

154. Sardar, M., Roy, I. & Gupta, M. N. Simultaneous purification and immobilization of Aspergillus niger xylanase on the reversibly soluble polymer EudragitTM L-100. *Enzyme Microb. Technol.* **27**, 672–679 (2000).

155. Barbosa, O. *et al.* Strategies for the one-step immobilization–purification of enzymes as industrial biocatalysts. *Biotechnol. Adv.* **33**, 435–456 (2015).

156. Abbaszadeh, M. & Hejazi, P. Metal affinity immobilization of cellulase on Fe3O4 nanoparticles with copper as ligand for biocatalytic applications. *Food Chem.* **290**, 47–55 (2019).

157. Mateo, C. *et al.* Immobilization of enzymes on heterofunctional epoxy supports. *Nat. Protoc.* **2**, 1022–1033 (2007).

158. Soleimani, M., Khani, A. & Najafzadeh, K. α-Amylase immobilization on the silica nanoparticles for cleaning performance towards starch soils in laundry detergents. *J. Mol. Catal. B Enzym.* **74**, 1–5 (2012).

159. Damnjanović, J. J. *et al.* Covalently immobilized lipase catalyzing high-yielding optimized geranyl butyrate synthesis in a batch and fluidized bed reactor. *J. Mol. Catal. B Enzym.* **75**, 50–59 (2012).

160. Jin, W. & Brennan, J. D. Properties and applications of proteins encapsulated within sol–gel derived materials. *Anal. Chim. Acta* **461**, 1–36 (2002).

161. Tsai, H. & Doong, R. Preparation and characterization of urease-encapsulated biosensors in poly(vinyl alcohol)-modified silica sol–gel materials. *Biosens. Bioelectron.* **23**, 66–73 (2007).

162. Rother, C. & Nidetzky, B. Enzyme Immobilization by Microencapsulation: Methods, Materials, and Technological Applications. *Encyclopedia of Industrial Biotechnology* 1–21 (2014).

163. Chen, W.-H., Vázquez-González, M., Zoabi, A., Abu-Reziq, R. & Willner, I. Biocatalytic cascades driven by enzymes encapsulated in metal–organic framework nanoparticles. *Nat. Catal.* **1**, 689–695 (2018).

164. Cao, L., Langen, L. van & Sheldon, R. A. Immobilised enzymes: carrier-bound or carrier-free? *Curr. Opin. Biotechnol.* **14**, 387–394 (2003).

165. Sheldon, R. A. Characteristic features and biotechnological applications of cross-linked enzyme aggregates (CLEAs). *Appl. Microbiol. Biotechnol.* **92**, 467–477 (2011).

166. Velasco-Lozano, S., López-Gallego, F., Mateos-Díaz, J. C. & Favela-Torres, E. Cross-linked enzyme aggregates (CLEA) in enzyme improvement – a review. *Biocatalysis* **1**, 166–177 (2016).

167. Lee, S. W., Cheon, S. A., Kim, M. Il & Park, T. J. Organic–inorganic hybrid nanoflowers: types, characteristics, and future prospects. *J. Nanobiotechnology* **13**, 54 (2015).

168. Romero-Fernández, M. *et al.* Designing continuous flow reaction of xylan hydrolysis for xylooligosaccharides production in packed-bed reactors using xylanase immobilized on methacrylic polymer-based supports. *Bioresour. Technol.* **266**, 249–258 (2018).

169. Chen, B., Miller, E. M., Miller, L., Maikner, J. J. & Gross, R. A. Effects of Macroporous Resin Size on Candida antarctica Lipase B Adsorption, Fraction of Active Molecules, and Catalytic Activity for Polyester Synthesis. *Langmuir* **23**, 1381–1387 (2007).

170. Johnson, C. R. & Bis, S. J. Enzymatic asymmetrization of meso-2-cycloalken-1,4-diols and their diacetates in organic and aqueous media. *Tetrahedron Lett.* **33**, 7287–7290 (1992).

171. De Maio, A. *et al.* Influence of the spacer length on the activity of enzymes immobilised on nylon/polyGMA membranes: Part 2: Non-isothermal conditions. *J. Mol. Catal. B Enzym.* **21**, 253–265 (2003).

172. Bakker, M., van de Velde, F., van Rantwijk, F. & Sheldon, R. A. Highly efficient

immobilization of glycosylated enzymes into polyurethane foams. *Biotechnol. Bioeng.* **70**, 342–348 (2000).

173. van de Velde, F., Lourenço, N. D., Pinheiro, H. M. & Bakker, M. Carrageenan: A Food-Grade and Biocompatible Support for Immobilisation Techniques. *Adv. Synth. Catal.* **344**, 815–835 (2002).

174. de Souza, S. P. *et al.* Cellulose as an efficient matrix for lipase and transaminase immobilization. *RSC Adv.* **6**, 6665–6671 (2016).

175. K. de S. Lira, R. *et al.* Agroindustrial Wastes as a Support for the Immobilization of Lipase from Thermomyces lanuginosus: Synthesis of Hexyl Laurate. *Biomolecules* **11**, (2021).

176. Benítez-Mateos, A. I., Bertella, S., Behaghel de Bueren, J., Luterbacher, J. S. & Paradisi, F. Dual Valorization of Lignin as a Versatile and Renewable Matrix for Enzyme Immobilization and (Flow) Bioprocess Engineering. *ChemSusChem* **14**, 3198–3207 (2021).

177. Yang, S.-Y. *et al.* Production of glutaric acid from 5-aminovaleric acid by robust whole-cell immobilized with polyvinyl alcohol and polyethylene glycol. *Enzyme Microb. Technol.* **128**, 72–78 (2019).

178. Tripathi, A. *et al.* Improved bio-catalytic conversion by novel immobilization process using cryogel beads to increase solvent production. *Enzyme Microb. Technol.* **47**, 44–51 (2010).

179. Sieber, S. *et al.* Immobilization of Enzymes on PLGA Sub-Micrometer Particles by Crosslinked Layer-by-Layer Deposition. *Macromol. Biosci.* **17**, 1700015 (2017).

180. Hudson, S., Cooney, J. & Magner, E. Proteins in Mesoporous Silicates. *Angew. Chemie Int. Ed.* **47**, 8582–8594 (2008).

181. Cabrera, K. Applications of silica-based monolithic HPLC columns. *J. Sep. Sci.* **27**, 843–852 (2004).

182. David, A. E., Wang, N. S., Yang, V. C. & Yang, A. J. Chemically surface modified gel (CSMG): An excellent enzyme-immobilization matrix for industrial processes. *J. Biotechnol.* **125**, 395–407 (2006).

183. Engelmark Cassimjee, K. *et al.* A general protein purification and immobilization method on controlled porosity glass: biocatalytic applications. *Chem. Commun.* **50**, 9134–9137 (2014).

184. Daniel, M.-C. & Astruc, D. Gold Nanoparticles: Assembly, Supramolecular Chemistry, Quantum-Size-Related Properties, and Applications toward Biology, Catalysis, and Nanotechnology. *Chem. Rev.* **104**, 293–346 (2004).

185. Kluchova, K. *et al.* Superparamagnetic maghemite nanoparticles from solid-state synthesis – Their functionalization towards peroral MRI contrast agent and magnetic carrier for trypsin immobilization. *Biomaterials* **30**, 2855–2863 (2009).

186. Kouassi, G. K., Irudayaraj, J. & McCarty, G. Activity of glucose oxidase functionalized onto magnetic nanoparticles. *Biomagn. Res. Technol.* **3**, 1 (2005).

187. Tsang, S. C., Yu, C. H., Gao, X. & Tam, K. Silica-Encapsulated Nanomagnetic Particle as a New Recoverable Biocatalyst Carrier. *J. Phys. Chem. B* **110**, 16914–16922 (2006).

188. Klis, M., Karbarz, M., Stojek, Z., Rogalski, J. & Bilewicz, R. Thermoresponsive Poly(N-isopropylacrylamide) Gel for Immobilization of Laccase on Indium Tin Oxide Electrodes. *J. Phys. Chem. B* **113**, 6062–6067 (2009).

189. Ye, N. *et al.* Metal-Organic Frameworks: A New Platform for Enzyme Immobilization. *ChemBioChem* **21**, 2585–2590 (2020).

190. De Santis, P., Meyer, L.-E. & Kara, S. The rise of continuous flow biocatalysis – fundamentals, very recent developments and future perspectives. *React. Chem. Eng.* **5**, 2155–2184 (2020).

191. Mallia, C. J. & Baxendale, I. R. The Use of Gases in Flow Synthesis. *Org. Process Res. Dev.* **20**, 327–360 (2016).

192. Lawrence, J., O'Sullivan, B., Lye, G. J., Wohlgemuth, R. & Szita, N. Microfluidic multi-input reactor for biocatalytic synthesis using transketolase. *J. Mol. Catal. B Enzym.* **95**, 111–117 (2013).

193. Benítez-Mateos, A. I., Contente, M. L., Roura Padrosa, D. & Paradisi, F. Flow biocatalysis 101: design, development and applications. *React. Chem. Eng.* **6**, 599–611 (2021).

194. Britton, J., Majumdar, S. & Weiss, G. A. Continuous flow biocatalysis. *Chem. Soc. Rev.* **47**, 5891–5918 (2018).

195. Santi, M. *et al.* Flow Biocatalysis: A Challenging Alternative for the Synthesis of APIs and Natural Compounds. *International Journal of Molecular Sciences* **22**, (2021).

196. Wohlgemuth, R., Plazl, I., Žnidaršič-Plazl, P., Gernaey, K. V & Woodley, J. M. Microscale technology and biocatalytic processes: opportunities and challenges for synthesis.

Trends Biotechnol. **33**, 302–314 (2015).

197. Yao, X., Zhang, Y., Du, L., Liu, J. & Yao, J. Review of the applications of microreactors. *Renew. Sustain. Energy Rev.* **47**, 519–539 (2015).

198. Tomaszewski, B., Schmid, A. & Buehler, K. Biocatalytic Production of Catechols Using a High Pressure Tube-in-Tube Segmented Flow Microreactor. *Org. Process Res. Dev.* **18**, 1516–1526 (2014).

199. Ringborg, R. H., Toftgaard Pedersen, A. & Woodley, J. M. Automated Determination of Oxygen-Dependent Enzyme Kinetics in a Tube-in-Tube Flow Reactor. *ChemCatChem* **9**, 3273 (2017).

200. Cen, Y.-K., Liu, Y.-X., Xue, Y.-P. & Zheng, Y.-G. Immobilization of Enzymes in/on Membranes and their Applications. *Adv. Synth. Catal.* **361**, 5500–5515 (2019).

201. Ranieri, G., Mazzei, R., Wu, Z., Li, K. & Giorno, L. Use of a Ceramic Membrane to Improve the Performance of Two-Separate-Phase Biocatalytic Membrane Reactor. *Molecules* **21**, (2016).

202. Salvi, H. M., Kamble, M. P. & Yadav, G. D. Synthesis of Geraniol Esters in a Continuous-Flow Packed-Bed Reactor of Immobilized Lipase: Optimization of Process Parameters and Kinetic Modeling. *Appl. Biochem. Biotechnol.* **184**, 630–643 (2018).

203. Bolivar, J. M. & López-Gallego, F. Characterization and evaluation of immobilized enzymes for applications in flow reactors. *Curr. Opin. Green Sustain. Chem.* **25**, 100349 (2020).

204. Farkas, E. *et al.* Chemoenzymatic Dynamic Kinetic Resolution of Amines in Fully Continuous-Flow Mode. *Org. Lett.* **20**, 8052–8056 (2018).

205. Šalić, A. & Zelić, B. ADH-catalysed hexanol oxidation with fully integrated NADH regeneration performed in microreactors connected in series. *RSC Adv.* **4**, 41714–41721 (2014).

206. Ender, F. *et al.* Microfluidic Multiple Cell Chip Reactor Filled with Enzyme-Coated Magnetic Nanoparticles — An Efficient and Flexible Novel Tool for Enzyme Catalyzed Biotransformations. *J. Flow Chem.* **6**, 43–52 (2016).

207. Babich, L., Hartog, A. F., van Hemert, L. J. C., Rutjes, F. P. J. T. & Wever, R. Synthesis of Carbohydrates in a Continuous Flow Reactor by Immobilized Phosphatase and Aldolase. *ChemSusChem* **5**, 2348–2353 (2012).

208. Adamczak, M., Bornscheuer, U. T. & Bednarski, W. The application of biotechnological

methods for the synthesis of biodiesel. *Eur. J. Lipid Sci. Technol.* **111**, 800–813 (2009).

209. Gibellini, F. & Smith, T. K. The Kennedy pathway—De novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life* **62**, 414–428 (2010).

210. Hofmann, K. A superfamily of membrane-bound O-acyltransferases with implications for Wnt signaling. *Trends Biochem. Sci.* **25**, 111–112 (2000).

211. Stöveken, T. & Steinbüchel, A. Bacterial Acyltransferases as an Alternative for Lipase-Catalyzed Acylation for the Production of Oleochemicals and Fuels. *Angew. Chemie Int. Ed.* **47**, 3688–3694 (2008).

212. Röttig, A., Wenning, L., Bröker, D. & Steinbüchel, A. Fatty acid alkyl esters: perspectives for production of alternative biofuels. *Appl. Microbiol. Biotechnol.* **85**, 1713–1733 (2010).

213. Kalscheuer, R., Stölting, T. & Steinbüchel, A. Microdiesel: Escherichia coli engineered for fuel production. *Microbiology* **152**, 2529–2536 (2006).

214. Santín, O., Galié, S. & Moncalián, G. Directed evolution of a bacterial WS/DGAT acyltransferase: improving tDGAT from Thermomonospora curvata. *Protein Eng. Des. Sel.* **32**, 25–32 (2019).

215. Chang, T.-Y., Li, B.-L., Chang, C. C. Y. & Urano, Y. Acyl-coenzyme A:cholesterol acyltransferases. *Am. J. Physiol. Metab.* **297**, E1–E9 (2009).

216. Qian, H. *et al.* Structural basis for catalysis and substrate specificity of human ACAT1. *Nature* **581**, 333–338 (2020).

217. Chek, M. F. *et al.* Structure of polyhydroxyalkanoate (PHA) synthase PhaC from Chromobacterium sp. USM2, producing biodegradable plastics. *Sci. Rep.* **7**, 5312 (2017).

218. Taguchi, S. & Doi, Y. Evolution of Polyhydroxyalkanoate (PHA) Production System by "Enzyme Evolution": Successful Case Studies of Directed Evolution. *Macromol. Biosci.* **4**, 145–156 (2004).

219. Mason, A. B. & Dufour, J.-P. Alcohol acetyltransferases and the significance of ester synthesis in yeast. *Yeast* **16**, 1287–1298 (2000).

220. Nancolas, B., Bull, I. D., Stenner, R., Dufour, V. & Curnow, P. Saccharomyces cerevisiae Atf1p is an alcohol acetyltransferase and a thioesterase in vitro. *Yeast* **34**, 239–251 (2017).

221. Stribny, J., Querol, A. & Pérez-Torrado, R. Differences in Enzymatic Properties of the Saccharomyces kudriavzevii and Saccharomyces uvarum Alcohol Acetyltransferases and Their Impact on Aroma-Active Compounds Production. *Front. Microbiol.* **7**, 897 (2016).

222. Rodriguez, G. M., Tashiro, Y. & Atsumi, S. Expanding ester biosynthesis in Escherichia coli. *Nat. Chem. Biol.* **10**, 259–265 (2014).

223. Saerens, S. M. G. *et al.* The Saccharomyces cerevisiae EHT1 and EEB1 Genes Encode Novel Enzymes with Medium-chain Fatty Acid Ethyl Ester Synthesis and Hydrolysis Capacity *. *J. Biol. Chem.* **281**, 4446–4456 (2006).

224. Kruis, A. J. *et al.* Ethyl acetate production by the elusive alcohol acetyltransferase from yeast. *Metab. Eng.* **41**, 92–101 (2017).

225. 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).

226. Knight, M. J., Bull, I. D. & Curnow, P. The yeast enzyme Eht1 is an octanoyl-CoA:ethanol acyltransferase that also functions as a thioesterase. *Yeast* **31**, 463–474 (2014).

227. Patinios, C. *et al.* Eat1-Like Alcohol Acyl Transferases From Yeasts Have High Alcoholysis and Thiolysis Activity. *Front. Microbiol.* **11**, 2737 (2020).

228. Müller, J., Sowa, M. A., Fredrich, B., Brundiek, H. & Bornscheuer, U. T. Enhancing the Acyltransferase Activity of Candida antarctica Lipase A by Rational Design. *ChemBioChem* **16**, 1791–1796 (2015).

229. Mathews, I. *et al.* Structure of a Novel Enzyme That Catalyzes Acyl Transfer to Alcohols in Aqueous Conditions. *Biochemistry* **46**, 8969–8979 (2007).

230. Weissman, K. J. *et al.* The Thioesterase of the Erythromycin-Producing Polyketide Synthase: Influence of Acyl Chain Structure on the Mode of Release of Substrate Analogues from the Acyl Enzyme Intermediates. *Angew. Chemie Int. Ed.* **37**, 1437–1440 (1998).

231. de Leeuw, N. *et al.* Ester Synthesis in Water: Mycobacterium smegmatis Acyl Transferase for Kinetic Resolutions. *Adv. Synth. Catal.* **360**, 242–249 (2018).

232. Contente, M. L., Pinto, A., Molinari, F. & Paradisi, F. Biocatalytic N-Acylation of Amines in Water Using an Acyltransferase from Mycobacterium smegmatis. *Adv. Synth. Catal.* **360**, 4814–4819 (2018).

233. Contente, M. L., Roura Padrosa, D., Molinari, F. & Paradisi, F. A strategic Ser/Cys exchange in the catalytic triad unlocks an acyltransferase-mediated synthesis of thioesters and tertiary amides. *Nat. Catal.* **3**, 1020–1026 (2020).

234. Risdian, C., Mozef, T. & Wink, J. Biosynthesis of Polyketides in Streptomyces. *Microorganisms* **7**, 124 (2019).

235. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–295 (2003).

236. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).

237. Joshi, V. C. & Wakil, S. J. Studies on the mechanism of fatty acid synthesis: XXVI. Purification and properties of malonyl-coenzyme A—Acyl carrier protein transacylase of Escherichia coli. *Arch. Biochem. Biophys.* **143**, 493–505 (1971).

238. Dunn, B. J. & Khosla, C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *J. R. Soc. Interface* **10**, 20130297 (2013).

239. Musiol-Kroll, E. M. & Wohlleben, W. Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiot. (Basel, Switzerland)* **7**, 62 (2018).

240. Oliynyk, M., Brown, M. J. B., Cortés, J., Staunton, J. & Leadlay, P. F. A hybrid modular polyketide synthase obtained by domain swapping. *Chem. Biol.* **3**, 833–839 (1996).

241. Liu, L., Thamchaipenet, A., Fu, H., Betlach, M. & Ashley, G. Biosynthesis of 2-Nor-6-deoxyerythronolide B by Rationally Designed Domain Substitution. *J. Am. Chem. Soc.* **119**, 10553–10554 (1997).

242. Ranganathan, A. *et al.* Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues. *Chem. Biol.* **6**, 731–741 (1999).

243. Lau, J., Fu, H., Cane, D. E. & Khosla, C. Dissecting the Role of Acyltransferase Domains of Modular Polyketide Synthases in the Choice and Stereochemical Fate of Extender Units. *Biochemistry* **38**, 1643–1651 (1999).

244. Sundermann, U. *et al.* Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase. *ACS Chem. Biol.* **8**, 443–450 (2013).

245. Bergeret, F. *et al.* Biochemical and Structural Study of the Atypical Acyltransferase Domain from the Mycobacterial Polyketide Synthase Pks13. *J. Biol. Chem.* **287**, 33675–

33690 (2012).

246.   Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996–1047 (2010).

247.   Muth, T. *et al.* JDet: interactive calculation and visualization of function-related conservation patterns in multiple sequence alignments and structures. *Bioinformatics* **28**, 584–586 (2012).

248.   W., L. S. & Rama, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science (80-.).* **286**, 295–299 (1999).

249.   Koch, A. A. *et al.* A Single Active Site Mutation in the Pikromycin Thioesterase Generates a More Effective Macrocyclization Catalyst. *J. Am. Chem. Soc.* **139**, 13456–13465 (2017).

250.   Narayan, A. R. H. *et al.* Enzymatic hydroxylation of an unactivated methylene C–H bond guided by molecular dynamics simulations. *Nat. Chem.* **7**, 653–660 (2015).

251.   Gao, X. *et al.* Directed Evolution and Structural Characterization of a Simvastatin Synthase. *Chem. Biol.* **16**, 1064–1074 (2009).

252.   Jiménez-Osés, G. *et al.* The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* **10**, 431–436 (2014).

253.   S., I. E. & Johann, D. Structural Mechanism for Statin Inhibition of HMG-CoA Reductase. *Science (80-.).* **292**, 1160–1164 (2001).

254.   Istvan, E. Statin inhibition of HMG-CoA reductase: a 3-dimensional view. *Atheroscler. Suppl.* **4**, 3–8 (2003).

255.   Campbell, C. D. & Vederas, J. C. Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes. *Biopolymers* **93**, 755–763 (2010).

256.   Wang, J. *et al.* Structural basis for the biosynthesis of lovastatin. *Nat. Commun.* **12**, 867 (2021).

257.   Syed, M. B. & Ponnusamy, T. Bioconversion of mevastatin to pravastatin by various microorganisms and its applications – A review. *Biocatal. Agric. Biotechnol.* **13**, 62–74 (2018).

258.   Hoffman, W. F. *et al.* 3-Hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors. 4. Side-chain ester derivatives of mevinolin. *J. Med. Chem.* **29**, 849–852 (1986).

259. Askin, D., Verhoeven, T. R., Liu, T. M. H. & Shinkai, I. Synthesis of synvinolin: extremely high conversion alkylation of an ester enolate. *J. Org. Chem.* **56**, 4929–4932 (1991).

260. Xie, X., Watanabe, K., Wojcicki, W. A., Wang, C. C. C. & Tang, Y. Biosynthesis of Lovastatin Analogs with a Broadly Specific Acyltransferase. *Chem. Biol.* **13**, 1161–1169 (2006).

261. Xie, X. & Tang, Y. Efficient synthesis of simvastatin by use of whole-cell biocatalysis. *Appl. Environ. Microbiol.* **73**, 2054–2060 (2007).

262. Xie, X., Wong, W. W. & Tang, Y. Improving simvastatin bioconversion in Escherichia coli by deletion of bioH. *Metab. Eng.* **9**, 379–386 (2007).

263. Xie, X. *et al.* Rational improvement of simvastatin synthase solubility in Escherichia coli leads to higher whole-cell biocatalytic activity. *Biotechnol. Bioeng.* **102**, 20–28 (2009).

264. Huang, X., Liang, Y., Yang, Y. & Lu, X. Single-step production of the simvastatin precursor monacolin J by engineering of an industrial strain of Aspergillus terreus. *Metab. Eng.* **42**, 109–114 (2017).

265. Bond, C. M. & Tang, Y. Engineering Saccharomyces cerevisiae for production of simvastatin. *Metab. Eng.* **51**, 1–8 (2019).

266. Acosta, A., Filice, M., Fernandez-Lorente, G., Palomo, J. M. & Guisan, J. M. Kinetically controlled synthesis of monoglyceryl esters from chiral and prochiral acids methyl esters catalyzed by immobilized Rhizomucor miehei lipase. *Bioresour. Technol.* **102**, 507–512 (2011).

267. Švedas, V., Guranda, D., van Langen, L., van Rantwijk, F. & Sheldon, R. Kinetic study of penicillin acylase from Alcaligenes faecalis. *FEBS Lett.* **417**, 414–418 (1997).

268. Reetz, M. T. The Importance of Additive and Non-Additive Mutational Effects in Protein Engineering. *Angew. Chemie Int. Ed.* **52**, 2658–2666 (2013).

269. Van Die, I. M., Bergmans, H. E. N. & Hoekstra, W. P. M. Transformation In Escherichia coli: Studies On The Role Of The Heat Shock In Induction Of Competence. *Microbiology* **129**, 663–670 (1983).

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

271. Grassetti, D. R. Determination of Sulfhydryl Groups with 2,2'- or 4,4'- Dithiodipyridine. *Arch. Biochem. Biophys.* **119**, 41–49 (1967).

272. Semisotnov, G. V *et al.* Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119–128 (1991).

273. Schrödinger, L., & DeLano, W. (2020). PyMOL. Retrieved from http://www.pymol.org/pymol

274. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

275. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

276. Case, D. A. et al. AMBER 2020 Reference Manual. *Univ Calif San Fr CA*, USA. 1–923 (2020).

277. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

278. Vanderveen, J. R., Durelle, J. & Jessop, P. G. Design and evaluation of switchable-hydrophilicity solvents. *Green Chem.* **16**, 1187 (2014).

279. 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.* **79**, 926–935 (1983).

280. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72**, 2384–2393 (1980).

281. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).

282. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

283. Chovancova, E. *et al.* CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLOS Comput. Biol.* **8**, e1002708 (2012).

284. Pavelka, A. *et al.* CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **13**, 505–517 (2016).

285. Nivón, L. G., Moretti, R. & Baker, D. A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLoS One* **8**, e59004 (2013).

286. Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E. & Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* **23**, 47–55 (2014).

287. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

288. Park, H. *et al.* Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).

289. Yoshino, M. & Murakami, K. Analysis of the substrate inhibition of complete and partial types. *Springerplus* **4**, (2015).

290. Müller, H. *et al.* Sequence-Based Prediction of Promiscuous Acyltransferase Activity in Hydrolases. *Angew. Chem. Int. Ed. Engl.* **59**, 11607–11612 (2020).

291. Li, G., Zhang, H., Sun, Z., Liu, X. & Reetz, M. T. Multiparameter Optimization in Directed Evolution: Engineering Thermostability, Enantioselectivity, and Activity of an Epoxide Hydrolase. *ACS Catal.* **6**, 3679–3687 (2016).

292. Huang, P. *et al.* Evaluating Protein Engineering Thermostability Prediction Tools Using an Independently Generated Dataset. *ACS Omega* **5**, 6487–6493 (2020).

293. Luo, Y. *et al.* Enhancing the biocatalytic manufacture of the key intermediate of atorvastatin by focused directed evolution of halohydrin dehalogenase. *Sci. Rep.* **7**, 42064 (2017).

294. Howard, M. J. Protein NMR spectroscopy. *Curr. Biol.* **8**, R331–R333 (1998).

295. Cavalli, A., Salvatella, X., Dobson, C. M. & Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci.* **104**, 9615 LP − 9620 (2007).

296. Fowler, N. J., Sljoka, A. & Williamson, M. P. A method for validating the accuracy of NMR protein structures. *Nat. Commun.* **11**, 6321 (2020).

297. Smyth, M. S. & Martin, J. H. x ray crystallography. *Mol. Pathol.* **53**, 8–14 (2000).

298. Yee, A. A. *et al.* NMR and X-ray Crystallography, Complementary Tools in Structural Proteomics of Small Proteins. *J. Am. Chem. Soc.* **127**, 16512–16517 (2005).

299. Gawas, U. B., Mandrekar, V. K. & Majik, M. S. Chapter 5 - Structural analysis of proteins

using X-ray diffraction technique. in (eds. Meena, S. N. & Naik, M. M. B. T.-A. in B. S. R.) 69–84 (Academic Press, 2019).

300. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).

301. Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim. Biophys. Acta - Gen. Subj.* **1862**, 324–334 (2018).

302. Jonic, S. & Vénien-Bryan, C. Protein structure determination by electron cryo-microscopy. *Curr. Opin. Pharmacol.* **9**, 636–642 (2009).

303. Childers, M. C. & Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* **2**, 9–33 (2017).

304. Liu, Y. & Kuhlman, B. RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, W235–W238 (2006).

305. Aaron, K., E., B. M., David, B. & L., S. B. Computational Thermostabilization of an Enzyme. *Science (80-.).* **308**, 857–860 (2005).

306. Butterfoss, G. L. & Kuhlman, B. COMPUTER-BASED DESIGN OF NOVEL PROTEIN STRUCTURES. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 49–65 (2006).

307. Van Wart, A. T., Durrant, J., Votapka, L. & Amaro, R. E. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J. Chem. Theory Comput.* **10**, 511–517 (2014).

308. Li, Y. *et al.* Hierarchical Decomposition for Betweenness Centrality Measure of Complex Networks. *Sci. Rep.* **7**, 46491 (2017).

309. Ni, J., Takehara, M. & Watanabe, H. Identification of activity related amino acid mutations of a GH9 termite cellulase. *Bioresour. Technol.* **101**, 6438–6443 (2010).

310. Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr. Protoc. Bioinforma.* **00**, 2.3.1-2.3.22 (2003).

311. Ehren, J., Govindarajan, S., Morón, B., Minshull, J. & Khosla, C. Protein engineering of improved prolyl endopeptidases for celiac sprue therapy. *Protein Eng. Des. Sel.* **21**, 699–707 (2008).

312. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

313. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

314. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

315. Liang, Y. & Lu, X. Structural insights into the catalytic mechanism of lovastatin hydrolase. *J. Biol. Chem.* **295**, 1047–1055 (2020).

316. Soares, V. *et al.* Umbelliferone esters with antibacterial activity produced by lipase-mediated biocatalytic pathway. *Biotechnol. Lett.* **43**, 469–477 (2021).

317. Paravidino, M. & Hanefeld, U. Enzymatic acylation: assessing the greenness of different acyl donors. *Green Chem.* **13**, 2651–2657 (2011).

318. Chênevert, R., Pelchat, N. & Morin, P. Lipase-mediated enantioselective acylation of alcohols with functionalized vinyl esters: acyl donor tolerance and applications. *Tetrahedron: Asymmetry* **20**, 1191–1196 (2009).

319. SONKARIA, S. *et al.* Evidence for 'lock and key' character in an anti-phosphonate hydrolytic antibody catalytic site augmented by non-reaction centre recognition: variation in substrate selectivity between an anti-phosphonate antibody, an anti-phosphate antibody and two hydro. *Biochem. J.* **381**, 125–130 (2004).

320. Li, J.-J. & Bugg, T. D. H. Investigation of a general base mechanism for ester hydrolysis in C–C hydrolase enzymes of the α/β-hydrolase superfamily: a novel mechanism for the serine catalytic triad. *Org. Biomol. Chem.* **5**, 507–513 (2007).

321. Ju, X., Li, J., Hou, M. & Tao, J. A lipase-catalyzed process for green synthesis of temsirolimus. *Eng. Life Sci.* **15**, 229–233 (2015).

322. Marchese, A. *et al.* Antibacterial and antifungal activities of thymol: A brief review of the literature. *Food Chem.* **210**, 402–414 (2016).

323. Lee, S. H. *et al.* Stereoselective Amination of Chiral Benzylic Ethers Using Chlorosulfonyl Isocyanate: Total Synthesis of (+)-Sertraline. *J. Org. Chem.* **76**, 10011–10019 (2011).

324. Roy, B. *et al.* A novel process for synthesis of Atovaquone. *Indian J. Chem. - Sect. B Org. Med. Chem.* **52**, 1299–1312 (2013).

325. Nair, V. *et al.* Menthol in electronic cigarettes: A contributor to respiratory disease? *Toxicol. Appl. Pharmacol.* **407**, 115238 (2020).

326. Eccles, R. Menthol and Related Cooling Compounds. *J. Pharm. Pharmacol.* **46**, 618–630 (1994).

327. Pergolizzi Jr, J. V, Taylor Jr, R., LeQuang, J.-A., Raffa, R. B. & Group, the N. R. The role and mechanism of action of menthol in topical analgesic products. *J. Clin. Pharm. Ther.* **43**, 313–319 (2018).

328. Klumpp, D. A. *et al.* Synthesis of Menthol Glycinates and Their Potential as Cooling Agents. *ACS Omega* **5**, 4043–4049 (2020).

329. Raj, S. B., Ramaswamy, S. & Plapp, B. V. Yeast Alcohol Dehydrogenase Structure and Catalysis. *Biochemistry* **53**, 5791–5803 (2014).

330. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **120**, 215–241 (2008).

331. Scalmani, G. & Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **132**, 114110 (2010).

332. Ribeiro, R. F., Marenich, A. V, Cramer, C. J. & Truhlar, D. G. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem. B* **115**, 14556–14562 (2011).

333. Gonzalez, C. & Schlegel, H. B. An improved algorithm for reaction path following. *J. Chem. Phys.* **90**, 2154–2161 (1989).

334. Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chemie Int. Ed.* **60**, 88–119 (2021).

335. Rosenthal, K. & Lütz, S. Recent developments and challenges of biocatalytic processes in the pharmaceutical industry. *Curr. Opin. Green Sustain. Chem.* **11**, 58–64 (2018).

336. A., H. M. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science (80-.).* **366**, 1255–1259 (2019).

337. Wiles, C., Hammond, M. J. & Watts, P. The development and evaluation of a continuous flow process for the lipase-mediated oxidation of alkenes. *Beilstein J. Org. Chem.* **5**, 27 (2009).

338. Kundu, S. *et al.* Continuous Flow Enzyme-Catalyzed Polymerization in a Microreactor.

J. Am. Chem. Soc. **133**, 6006–6011 (2011).

339. Le Joubioux, F., Bridiau, N., Sanekli, M., Graber, M. & Maugard, T. Continuous lipase-catalyzed production of pseudo-ceramides in a packed-bed bioreactor. *J. Mol. Catal. B Enzym.* **109**, 143–153 (2014).

340. Brahma, A. *et al.* An Orthogonal Biocatalytic Approach for the Safe Generation and Use of HCN in a Multistep Continuous Preparation of Chiral O-Acetylcyanohydrins. *Synlett* **27**, 262–266 (2016).

341. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science (80-.).* **329**, 305–309 (2010).

342. Planchestainer, M. *et al.* Continuous flow biocatalysis: production and in-line purification of amines by immobilised transaminase from Halomonas elongata. *Green Chem.* **19**, 372–375 (2017).

343. Benítez-Mateos, A. I. *et al.* Asymmetric Reduction of Prochiral Ketones by Using Self-Sufficient Heterogeneous Biocatalysts Based on NADPH-Dependent Ketoreductases. *Chem. – A Eur. J.* **23**, 16843–16852 (2017).

344. Ruzic, L., Bolivar, J. M. & Nidetzky, B. Glycosynthase reaction meets the flow: Continuous synthesis of lacto-N-triose II by engineered β-hexosaminidase immobilized on solid support. *Biotechnol. Bioeng.* **117**, 1597–1602 (2020).

345. Arnold, F. H. Metal-Affinity Separations: A New Dimension in Protein Processing. *Bio/Technology* **9**, 151–156 (1991).

346. Gaberc-Porekar, V. & Menart, V. Potential for Using Histidine Tags in Purification of Proteins at Large Scale. *Chem. Eng. Technol.* **28**, 1306–1314 (2005).

347. Ampon, K. Distribution of an enzyme in porous polymer beads. *J. Chem. Technol. Biotechnol.* **55**, 185–190 (1992).

348. Arana-Peña, S. *et al.* Effects of Enzyme Loading and Immobilization Conditions on the Catalytic Features of Lipase From Pseudomonas fluorescens Immobilized on Octyl-Agarose Beads. *Front. Bioeng. Biotechnol.* **8**, 36 (2020).

349. Masuda, Y., Kugimiya, S. & Kato, K. Improvement of thermal-stability of enzyme immobilized onto mesoporous zirconia. *J. Asian Ceram. Soc.* **2**, 11–19 (2014).

350. Talebi, M., Vaezifar, S., Jafary, F., Fazilati, M. & Motamedi, S. Stability Improvement of Immobilized α-amylase using Nano Pore Zeolite. *Iran. J. Biotechnol.* **14**, 33–38 (2016).

351. Pokalsky, C., Wick, P., Harms, E., Lytle, F. E. & Van Etten, R. L. Fluorescence Resolution of the Intrinsic Tryptophan Residues of Bovine Protein Tyrosyl Phosphatase (∗). *J. Biol. Chem.* **270**, 3809–3815 (1995).

352. Ghisaidoobe, A. B. T. & Chung, S. J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *Int. J. Mol. Sci.* **15**, 22518–22538 (2014).

353. Gijsbers, A., Nishigaki, T. & Sánchez-Puig, N. Fluorescence Anisotropy as a Tool to Study Protein-protein Interactions. *J. Vis. Exp.* 54640 (2016).

354. Weltz, J. S., Kienle, D. F., Schwartz, D. K. & Kaar, J. L. Reduced Enzyme Dynamics upon Multipoint Covalent Immobilization Leads to Stability-Activity Trade-off. *J. Am. Chem. Soc.* **142**, 3463–3471 (2020).

355. Pedroche, J. *et al.* Effect of the support and experimental conditions in the intensity of the multipoint covalent attachment of proteins on glyoxyl-agarose supports: Correlation between enzyme–support linkages and thermal stability. *Enzyme Microb. Technol.* **40**, 1160–1166 (2007).

356. Santiago-Arcos, J., Velasco-Lozano, S., Diamanti, E., Cortajarena, A. L. & López-Gallego, F. Immobilization Screening and Characterization of an Alcohol Dehydrogenase and its Application to the Multi-Enzymatic Selective Oxidation of 1,-Omega-Diols. *Front. Catal.* **1**, 9 (2021).

357. Tuma, R. Raman spectroscopy of proteins: from peptides to large assemblies. *J. Raman Spectrosc.* **36**, 307–319 (2005).

358. Benevides, J. M., Overman, S. A. & Thomas Jr., G. J. Raman Spectroscopy of Proteins. *Curr. Protoc. Protein Sci.* **33**, 17.8.1-17.8.35 (2003).

359. Rygula, A. *et al.* Raman spectroscopy of proteins: a review. *J. Raman Spectrosc.* **44**, 1061–1076 (2013).

360. Kuhar, N., Sil, S. & Umapathy, S. Potential of Raman spectroscopic techniques to study proteins. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **258**, 119712 (2021).

361. López-Gallego, F. & Yate, L. Selective biomineralization of Co3(PO4)2-sponges triggered by His-tagged proteins: efficient heterogeneous biocatalysts for redox processes. *Chem. Commun.* **51**, 8753–8756 (2015).

362. Smith, J. P. *et al.* Raman hyperspectral imaging with multivariate analysis for investigating enzyme immobilization. *Analyst* **145**, 7571–7581 (2020).