

Université du Québec  
Institut National de la Recherche Scientifique  
Centre INRS-Institut Armand-Frappier

## **PROTEIN ENGINEERING OF THE CALB LIPASE TO SYNTHESIZE FRAGRANCE COMPOUNDS**

Par  
Ying Lian Chew Fajardo

Mémoire présenté pour l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en microbiologie appliquée

### **Jury d'évaluation**

Président du jury et  
examineur interne

Pr. Jonathan Perreault, Ph. D.  
INRS- Institut Armand-Frappier

Examineur externe

Pr. David Kwan, Ph. D.  
Department of biology  
Concordia, University, Montreal, QC

Directeur de recherche

Pr. Nicolas Doucet, Ph. D.  
INRS-Institut Armand-Frappier



## ACKNOWLEDGMENTS

**Most of what exist** in the universe, our actions, and all the other forces, resources, and ideas, **has little value** and yields little result; on the other hand, a **few things** work fantastically well and **have tremendous impact**.

–Richard Koch

This modest work represents, so far, the platform that -I can already say- has enhanced the most my personal and professional paths. I am extremely honored of having had received the intellectual, social and capital support of: my research director Nicolas Doucet, who trusted me enough to let me handle by myself really expensive reagents and lab equipment; Guillaume Brault, who first shared his scientific dream with me and energetically engaged me to this project; and especially Yossef L de Los Santos, who challenged me, introduced me to new ideas, alternative points of view, and entirely different ways of thinking; so he played not only as my scientific mentor, but a great friend in moments of well-known-crisis of us studying-abroad kids missing home. I also found lab mates that resulted to be friends that I am sure I'll keep on cherish in the years to come. #Youknowwhoyouare.

Merci également aux professeurs de l'axe microbiologie qui m'ont donné des cours, qui ont fait preuve de patience et diligence pendant mon étape d'incompétence temporaire au niveau de la langue française. Ce paradoxe montre que pour construire nos aptitudes et nos compétences à venir, il faut accepter d'être incompétent temporairement. Il y a donc lieu de remercier tous ceux qui m'ont accompagnée dans le début de mon franco-chemin d'apprentissage -merci les amis, merci mon complice de langue française-.

Finalmente, quiero agradecer a mi familia. Los 3,000 Km de distancia entre nosotros, no nos separaron ni un poco. Sus visitas, llamadas, y fotos me hicieron sentir el calor de esta familia que me apoya y alienta donde sea que esté y en lo que sea que haga. Gracias tíos, primo y primas, hermanos y hermana, papá y mamá. Otra etapa que terminamos juntos. Gracias.



## ABSTRACT

A computationally-guided semi-rational protein design approach was used to improve the enzymatic selectivity and catalytic efficiency of *Pseudozyma antarctica* lipase B (CalB) to synthesize methyl salicylate and methyl cinnamate. These fatty acid esters have significant relevance as flavoring and fragrance compounds in the biotechnological industry. Moreover, CalB is a highly active lipase that is widely used for the enzymatic hydrolysis and synthesis of esters, offering potential for the biological production of flavoring agents. However, the relatively confined organization of its active site precludes the recognition of bulky and aromatic substrates. To overcome this limitation, *in silico* docking analyses of CalB were undertaken to identify amino acid residues involved in precursor binding and recognition. These “hot spots” were subjected to combinatorial mutagenesis to yield three generations of CalB libraries per substrate. A surrogate substrate was used to screen for synthetic activity and evaluation of the new CalB variants revealed mutations giving rise to significant increase in synthetic activity relative to wild-type CalB. Ultimately, the best CalB variant could serve as a template to develop an *E. coli* whole-cell biocatalyst suitable for industrial enzymatic synthesis of methyl salicylate.

Keywords: CalB, semi-rational protein design, molecular docking, methyl salicylate, directed evolution, protein engineering, lipase activity, enzyme catalysis.

## RÉSUMÉ

Une approche semi-rationnelle d'évolution dirigée de protéines couplée à des méthodes *in silico* a été utilisée pour améliorer la sélectivité enzymatique et l'efficacité catalytique de la lipase B de *Pseudozyma antarctica* (CalB) en vue de synthétiser le salicylate de méthyle et le cinnamate de méthyle. Ces esters d'acides gras jouent un rôle critique en tant que composés aromatisants et parfumés dans l'industrie biotechnologique. De plus, CalB est une lipase largement utilisée dans le processus d'hydrolyse enzymatique et la synthèse d'esters, offrant un potentiel de production biologique d'agents aromatisants. Néanmoins, l'architecture particulière de son site actif empêche la reconnaissance de substrats plus volumineux et aromatiques. Pour surmonter cette limitation, nous avons effectué des analyses *in silico* de CalB dans le but d'identifier les sites de résidus acides aminés les plus significativement impliqués dans la liaison et la reconnaissance du précurseur de ces composés. Ces "*hot spots*" ont été soumis à une mutagenèse combinatoire pour générer trois générations de bibliothèques de variants de CalB par substrat. Un substrat substitut a été utilisé pour mesurer l'activité de la protéine et l'évaluation de ces nouvelles générations de CalB a révélé une augmentation significative de l'activité de synthèse par rapport à l'enzyme sauvage. Le meilleur variant de CalB pourrait servir de modèle pour développer un système de catalyse *E. coli* en cellules entières adapté à la synthèse industrielle enzymatique du salicylate de méthyle.

Mots clés: CalB, design semi-rationnel de protéines, arrimage moléculaire, salicylate de méthyle, évolution dirigée, ingénierie des protéines, activité lipolytique, catalyse enzymatique.

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS .....</b>	<b>3</b>
<b>ABSTRACT .....</b>	<b>5</b>
<b>RÉSUMÉ .....</b>	<b>6</b>
<b>LIST OF TABLES.....</b>	<b>9</b>
<b>LIST OF FIGURES .....</b>	<b>10</b>
<b>EQUATION .....</b>	<b>12</b>
<b>ABBREVIATIONS .....</b>	<b>13</b>
<b>1. CHAPTER 1. INTRODUCTION.....</b>	<b>15</b>
1.1. CARBOXYLIC ACID ESTERS.....	15
1.2. CHEMISTRY OF ESTERIFICATION .....	16
1.3. ENZYME CLASSIFICATION AND REACTION MEDIA.....	18
1.4. CALB: STRUCTURAL FRAMEWORK AND CATALYSIS .....	18
1.5. PRECURSOR SUBSTRATE MOLECULES .....	21
1.6. ENZYME ENGINEERING .....	23
1.7. SEMI-RATIONAL DESIGN OF CALB TO OPTIMIZE SYNTHETIC ACTIVITY .....	25
1.8. COMPUTATIONAL APPROACH TO OPTIMIZE TARGETED POSITIONS FOR MUTATION .....	27
<b>2. CHAPTER 2. HYPOTHESIS AND RESEARCH OBJECTIVES.....</b>	<b>28</b>
2.1. METHODOLOGY.....	28
2.1.1. <i>Objective 1. Perform in silico predictions to identify amino acid positions most involved in substrate recognition through molecular docking .....</i>	<i>28</i>
2.1.2. <i>Objective 2. Build and screen CalB libraries for synthetic activity .....</i>	<i>31</i>
<b>3. CHAPTER 3. MANUSCRIPT SECTIONS .....</b>	<b>35</b>
3.1. AUTHOR CONTRIBUTIONS.....	35
3.2. MATERIALS AND METHODS .....	36
3.2.1. <i>DNA constructs and expression systems.....</i>	<i>36</i>
3.2.2. <i>Wild type and mutant CalB overexpression .....</i>	<i>36</i>
3.2.3. <i>Wild type and mutant CalB esterification assays .....</i>	<i>37</i>

3.2.4. Enzyme extraction and lyophilization .....	37
3.2.5. Virtual docking of substrates in the CalB active site .....	38
3.2.6. Iterative semi-rational evolution of CalB towards bulky substrates.....	38
3.2.7. Library Design Strategy.....	39
3.2.8. Screening of mutant libraries and transesterification assays .....	41
3.2.9. Bioinformatics .....	42
3.3. RESULTS .....	42
3.3.1. Optimization of CalB overexpression .....	42
3.3.2. High-throughput screening of lipase transesterification .....	43
3.3.3. Using ISRD to uncover and target active-site 'hot spots' for CalB activity improvement .....	44
3.3.4. Effect of mutations on the catalytic pocket of CalB.....	52
3.3.5. Natural vs. directed evolution of the CalB active site.....	53
3.4. DISCUSSION .....	55
<b>4. CONCLUSION AND PERSPECTIVES .....</b>	<b>58</b>
<b>5. REFERENCES .....</b>	<b>61</b>
<b>ANNEXE 1. SUPPORTING FIGURES .....</b>	<b>67</b>

## LIST OF TABLES

Table 1	Binding energy contributions of enzyme-substrate complexes for WT CalB and the most efficient CalB variants selected after three rounds of Iterative Semi-Rational Design (ISRD) _____	48
Annex 1	Table S1 Synthetic activity increase of CalB variants _____	72
	Table S2 Changes in CalB active-site volume upon mutagenesis and selection _____	73

## LIST OF FIGURES

Figure 1.1	Short-chain esters as flavoring agents. _____	15
Figure 1.2	Acid-catalyzed esterification of fatty acids _____	17
Figure 1.3	Acid-catalyzed transesterification of lipids _____	17
Figure 1.4	Active-site view of the CalB lipase from <i>Pseudozyma antarctica</i> _____	20
Figure 1.5	Catalytic mechanism of CalB from <i>Pseudozyma antarctica</i> _____	21
Figure 1.6	Precursor molecules for flavoring esters synthesized in this study _____	22
Figure 1.7	Semi-rational design is an enzyme engineering approach to access useful phenotypes on an accelerated time scale _____	25
Figure 1.8	Semi-rational evolution strategy to increase CalB fitness for the synthesis of methyl salicylate and methyl cinnamate _____	26
Figure 1.9	Example of Iterative Saturation Mutagenesis strategy (ISM) _____	27
Figure 2.1	Flowchart of substrate-imprinted docking strategy _____	30
Figure 2.2	Model reaction example on vinyl salicylate for the evaluation of lipase synthetic activity assay _____	33
Figure 2.3	General screening strategy employed in the current study _____	34
Figure 3	Schematic representation of the Iterative Semi-Rational Design (ISRD) strategy employed in the present work. _____	40
Figure 3.1.1	Structural changes in the active-site cavity of CalB upon mutagenesis and selection (A-B). _____	45
Figure 3.1.2	Structural changes in the active-site cavity of CalB upon mutagenesis and selection (C-D). _____	46
Figure 3.2	Methyl cinnamate and methyl salicylate production by the CalB lipase subjected to 3 rounds of Iterative Semi-Rational Design (ISRD) _____	50
Figure 3.3	Comparative analysis of $\alpha/\beta$ hydrolase sequence diversity with ISRD-selected active-site mutations that increase CalB transesterification on bulky substrates _____	54
Annex 1	Figure S1 Wild-type CalB overexpression in <i>E. coli</i> . _____	67

	Figure S2	Combined effects of E. coli strain, IPTG induction, and incubation temperature on WT CalB production yields.	68
Annex 1	Figure S3	Optimization of the CalB library screening methodology for the detection of methyl cinnamate and vinyl salicylate	69
	Figure S4	Methyl cinnamate production by CalB variants obtained from G1, G2 and G3 library screening	70
	Figure S5	Hierarchical modular organization of CalB using a Residue Interacting Network (RIN) analysis	71

## EQUATION

Equation 1. Poisson distribution for estimation of equiprobable variants diversity in protein-encoding library. \_\_\_\_\_31

## ABBREVIATIONS

CalB: *Pseudozyma antarctica* fraction B

WT: Wild type

MBTH: 3-methyl-2-benzothialinone

TAPMC: tetraaza-pentamethincyanine

IPTG: Isopropyl  $\beta$ -D-1-thiogalactopyranoside

ISM: Saturation Mutagenesis strategy

Enz1: Enzyme

Sub1: Substrate

Enz1-Sub1: Enzyme-substrate complex

ISRDL: Iterative Semi-Rational Design

RIN: Residue Interaction Network

G1: First generation

G2: Second generation

G3: Third generation

ETA: Evolutionary Trace Analysis

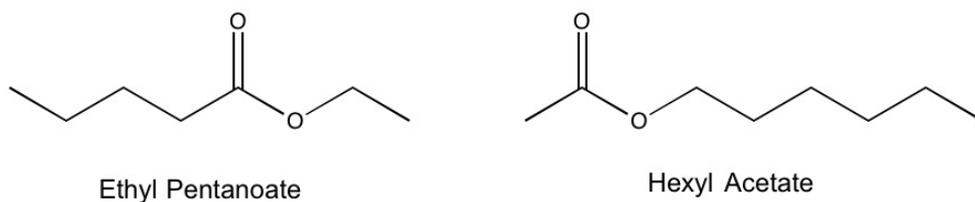
HTS: High-Throughput Screening



# 1. CHAPTER 1. INTRODUCTION

## 1.1. Carboxylic acid esters

Carboxylic acid esters form a significant group of compounds with broad applicability in the industrial world. Depending on their carbon chain length, they have different properties. They can be used in environmentally sustainable processes, e.g. for the production of biodiesel, as flavoring agents in the food industry, or as fragrance compounds in the cosmetic industry [1,2,3,4]. Specifically, short-chain esters are important flavoring agents and fragrance compounds (Figure 1.1).



**Figure 1.1** Short-chain esters as flavoring agents. Ethyl valerate with a typical fragrance compound of green apple and hexyl acetate with a pear flavour property are in high demand and are widely used in food, cosmetic and pharmaceutical industries [5].

Flavours and fragrances roughly represent 25% of the worldwide market of food additives [6]. As a result, efficient methods for the production of carboxylic acids are needed in order to meet global industrial demand. Traditionally, these compounds were extracted from natural sources, but the discovery of their molecular structures prompted chemical synthesis as an efficient production method. However, chemical synthesis of fragrances is environmentally deleterious due to the hazardous compounds used to extract feedstock. Other drawbacks include reduction of process efficiency and increase in downstream costs and synthesis of side products. Indeed, many compounds used for flavouring exist as optical isomers and further yield undesirable racemic mixture as side products [7]. Finally, additional limitations include lack of substrate selectivity and specificity, yielding undesirable enantiomer side products. These drawbacks significantly reduce process efficiency and increase downstream costs. On the other hand, natural extracts also generate high production costs and face a significantly

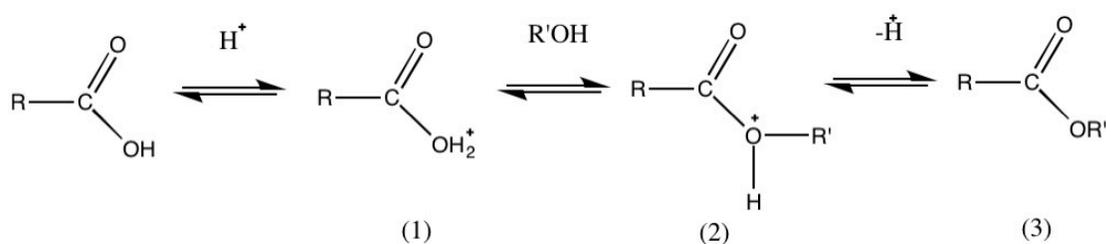
limited natural supply from biomass. Raw materials from plants contain low amounts of desired compounds (*i.e.* on the order of mg/kg of biomass), making the extraction processes expensive and time consuming. Additionally, production yields are highly dependent on various extraneous factors like climate and temperature and plant diseases, which are difficult to control [8].

Biocatalysis is the use of biological catalysts such as enzymes to perform chemical transformations on organic compounds. Enzymes used in these biochemical transformations can either be purified beforehand or reside within living cells (using a whole-cell culture) [9]. Although well over 3,000 enzymes have been identified so far, only 150-170 have been exploited for industrial purposes in different types of catalyzed reactions [10]. Interestingly, increased public awareness about natural products and environmentally friendly processes generate significant interest towards the use of fragrances and flavours from natural sources. As a result, a number of approaches have proposed using enzymes as a greener alternative for the production of natural aromas. The benefits offered by enzyme processes include high specificity, mild reaction conditions and reduced waste products. A plant exploiting engineered enzymatic reactions is more efficient than a traditional one, in addition to operating at a much lower energy capital [11]. The use of enzymes also represents a more sustainable alternative for producing short-chain bulky esters, which are commercially relevant fragrance compounds.

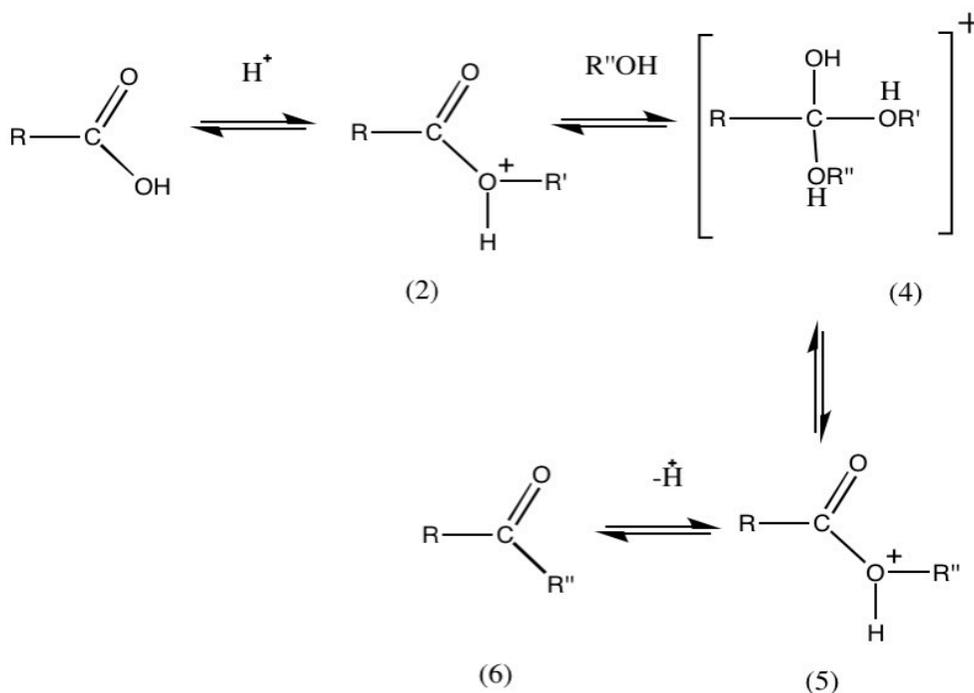
## **1.2. Chemistry of esterification**

Carboxylic acids can be esterified by alcohols in the presence of a proper enzymatic biocatalyst, as illustrated in Figure 1.2. The catalytic reaction starts with the protonation of the acid to produce an oxonium ion (1), which then is subjected to an exchange reaction with an alcohol to yield an intermediate (2). This intermediate can further lose a proton to become an ester (3). Any of these steps are chemically reversible; however, when an excess of alcohol is present in solution, reaction equilibrium moves towards complete esterification. In the presence of water (a stronger electron donor than aliphatic alcohols), no favorable intermediate is formed (2) and esterification cannot fully

proceed. Transesterification (or ester exchange) happens under similar conditions (Figure 1.3). In this case, the catalytic reaction starts with the initial protonation of the ester, followed by addition of the exchanging alcohol to produce the intermediate (4). This intermediate further proceeds through a transition state (5) to generate the ester (6). As in the esterification reaction, each step is chemically reversible and the reaction equilibrium will favor ester synthesis (6) in presence of excess alcohol. Again, if water is not excluded from the reaction, hydrolysis is favored by dissociation of an intermediate (4) analogous to a free acid ( $R'' = H$ ) [12].



**Figure 1.2** Acid-catalyzed esterification of fatty acids. Taken from ref.[12].



**Figure 1.3** Acid-catalysed transesterification of lipids. Taken from ref. [12].

Considering the requirements of the esterification reaction, three critical elements need to be considered for the development of a biotechnological program aimed at the production of aroma-related compounds: A) the type of enzyme that catalyzes the reaction (e.g. lipases, proteases, or glycosidases, etc.); B) the type of reaction medium (e.g. aqueous, solvent-free, organic, or SC-CO<sub>2</sub>, etc.); and C) the precursor molecules generating the final ester compounds.

### 1.3. Enzyme classification and reaction media

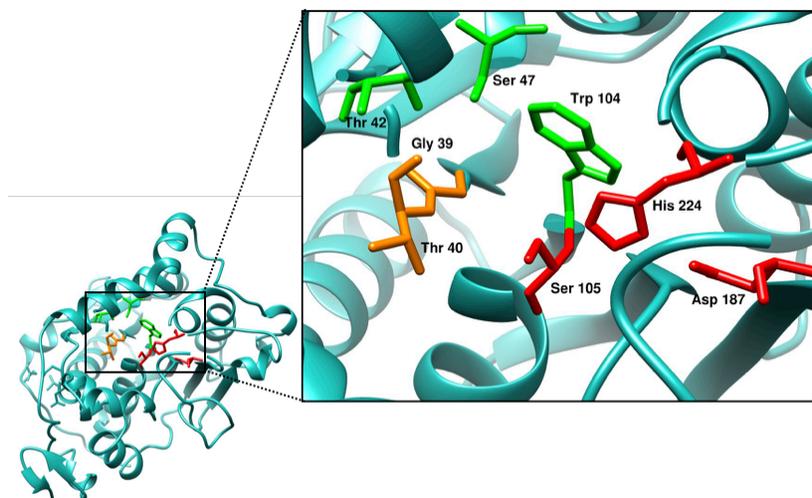
Lipases (EC 3.1.1.3) are extremely powerful biocatalysts in the synthesis of ester aroma compounds [11]. Indeed, one of the most promising enzyme technology applications in the food-aroma field is the use of reversed lipolysis in low water-content systems [13]. Lipases carry out esterification or transesterification reactions for the production of esters from inexpensive raw materials (i.e. fatty acids and alcohols). A number of lipases have been tested for their ability to promote ester synthesis in low-water content media, such as those from *Candida cylindracea*, *Pseudomonas fluorescens*, *Mucor miehei*, *Aspergillus sp.*, *Rhizopus arrhizus*, *Candida rugosa*, and *Pseudozyma antarctica*, amongst others [6]. The influence of variables such as reagent or enzyme concentrations, temperature, water content, or solvent properties have been thoroughly studied [6]. Still, exploring structural properties of the substrates (acids and alcohols) that define the potential of any defined lipase to synthesize a variety of short-chain carboxylic esters is critical. In particular, the lipase from *Pseudozyma antarctica* fraction B (CalB) is a robust enzyme used in organic synthesis, showing high catalytic efficiency for the resolution of chiral esters and amines via esterification and transesterification reactions [14].

### 1.4. CalB: Structural framework and catalysis

Among lipase family members, Lipase B from *Pseudozyma antarctica* (CalB) is one of the most widely used enzymes in industrial biocatalysis [15]. CalB is a member of the  $\alpha/\beta$  hydrolase family, displaying an active site with a catalytic triad: a nucleophilic residue (Ser105), a catalytic acid (Asp187), and a general base (His224). The enzyme

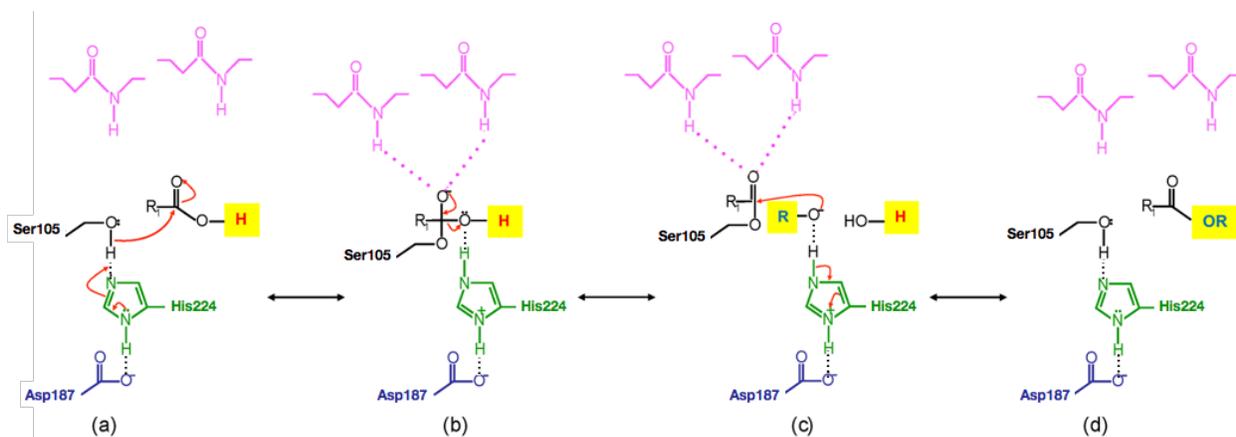
also displays an oxyanion hole (Gly39-Thr40) that promotes the stabilization of the negatively charged intermediate generated during ester bond hydrolysis, and a stereospecificity pocket (Thr42-Ser47-Trp104). The reaction mechanism mirrors that of the serine hydrolase family, with the presence of a nucleophilic serine in their active site, which is used for the hydrolysis or esterification of substrates. For proper catalysis to occur, the fast-reacting enantiomer of secondary alcohols positions its medium-sized substituent in the stereospecificity pocket and the large substituent toward the active site entrance (Figure 1.4). Experimental data have shown preference for small and medium-sized substituents for the secondary alcohol in the stereospecificity pocket (methyl = ethyl > isopropyl >> methylmethoxy). This effectively confirms that the largest alcohol substituent fitting this active-site subpocket as substrate is an ethyl group. Thus, limited tolerance for side-chain amino acid replacements is observed in the active site. In contrast, the enzyme has relaxed specificity towards larger alcohol substituents and carboxylate moieties on substrates. Apart from the size limitation of the alcohol moiety in the binding pocket, the enantiomeric preference of the enzyme is controlled through the geometric orientation of His224 and the oxygen of the alcohol leaving group [16].

Even though a large substituent (*eg.* methylmethoxy) does not fit perfectly in the binding pocket, the configuration allows for hydrogen bonding patterns essential for proper catalysis. Therefore, the limited size of the stereospecificity pocket restricts catalysis for a broader selection of substrates. As a result, we postulate that optimizing amino acid configuration at this site for substrate recognition with the simplest primary alcohol (methanol) could allow proper hydrogen bonding patterns at the transition state, resulting in enzyme activity.



**Figure 1.4** Active-site view of the CalB lipase from *Pseudozyma antarctica* (PDB entry 1TCA). Side chains for catalytic residues are shown in red (Ser105-His224-Asp187), while the stereospecificity pocket (Thr42-Ser47-Trp104) and the oxyanion hole (Gly39-Thr40) are shown in green and orange, respectively.

The catalytic mechanism for the synthesis of ester bonds catalyzed by CalB starts with the nucleophilic attack by the serine hydroxyl on the carbon of the susceptible acyl bond shown in Figure 1.5, giving rise to a tetrahedral intermediate and the formation of a negative charge on the carbonyl oxygen atom. Residues Gly39 and Thr40 of the oxyanion hole form hydrogen bonds with the negatively charged carbonyl oxygen atom and their main chain -NH groups. The nucleophilicity of the Ser105 is enhanced by transferring a proton to the catalytic His224; this transfer is facilitated by the catalytic acid residue Asp187, which orients the imidazole ring of His224 to neutralize the charge generated on it. Then, the proton is donated to the oxygen of the susceptible acyl bond, which results in cleaving and release of an ester product, while the acyl chain is esterified to the nucleophilic serine. Lastly, the alcohol molecule esterifies the covalent bond and releases the ester product [17].



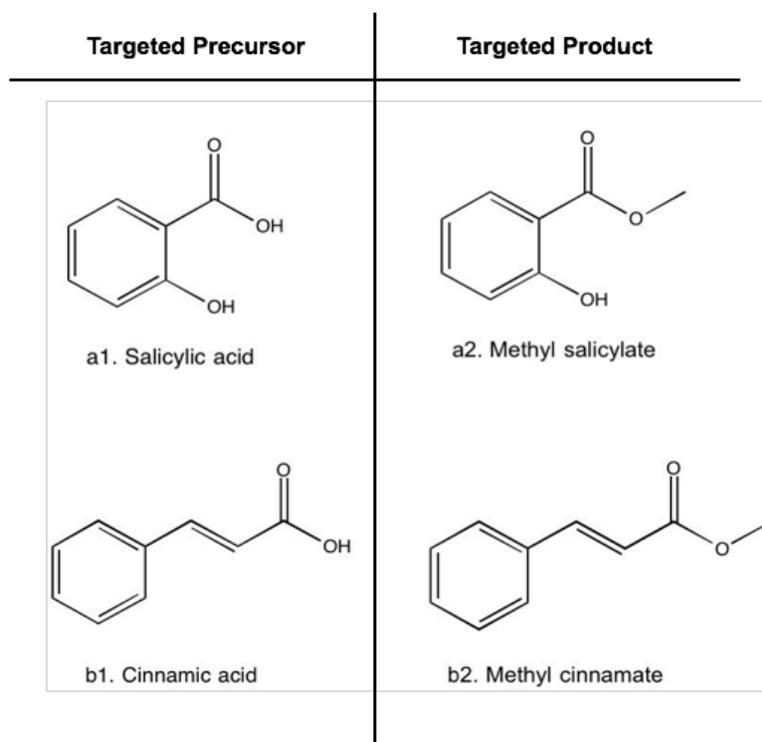
**Figure 1.5** Catalytic mechanism of CalB from *Pseudozyma Antarctica*. Active-site residues Asp187 and His224 are shown in blue and green, respectively; Ser105, substrate and water are shown in black; oxyanion hole residues are shown in magenta. (a) Nucleophilic attack of the serine hydroxyl on the carbonyl carbon of the susceptible acyl bond; (b) tetrahedral intermediate; (c) acyl-enzyme intermediate, released ester and nucleophilic attack by alcohol; (d) free enzyme and released ester product. *Modified from ref. [18].*

## 1.5. Precursor substrate molecules

Since most enzymes show high ligand specificity and selectivity, the possibility of working with industrially relevant precursor molecules remains somewhat limited. The extent of substrate specificity (*i.e.* enzyme plasticity) can determine whether a given enzyme will have general synthetic utility. Although enzymes with narrow substrate specificity are often efficient for catalyzing reactions using their natural substrate, this property becomes a limitation for the development of general purpose biocatalysts [11]. Luckily, individual enzyme properties can often be improved from their original wild-type form using directed evolution and rational design methodologies [19]. Among lipases that have shown great response to directed evolution is that of the ubiquitous environmental bacterium *Pseudomonas aeruginosa*. This bacterial lipase was evolved to catalyse the hydrolysis of 4-nitrophenyl 2-methyldecanoate, a model ester with >90% enantiomeric excess, compared with 2% enantiomeric excess for the wild-type enzyme [20]. Additionally, several studies have shown the usefulness of lipase usage for synthesizing flavour esters such as ethyl valerate, hexyl acetate [21], isoamyl acetate [22], and benzyl acetate [23].

However, because of active-site constraints and prohibitive precursor molecule size, properties or structure, using enzymes to synthesize useful flavoring esters is often limited to natural substrate specificity and diversity. As a result, the challenge of the present work was to successfully catalyze the synthesis of ice-mint flavor (methyl salicylate) and strawberry flavor (methyl cinnamate), two esters that require precursor molecules (salicylic and cinnamic acid, accordingly) for which lipase has low specificity (Figure 1.6).

CalB shows high substrate specificity for both short-chained fatty acids with linear and branched chain structure, in addition to unsaturated fatty acids coupled with different alcohols (n-butyl, isopentyl, 2-phenylethyl and geraniol) [6]. As shown in Figure 1.6, methyl salicylate and cinnamate are desired products, while the precursors salicylic acid and cinnamic acid, respectively, are bulky compounds that naturally would not fit into the precluded active site of CalB.



**Figure 1.6** Precursors and product molecules for flavoring esters synthesized in this study. Compounds a1) salicylic acid and b1) cinnamic acid are precursor substrates, while a2) methyl salicylate and b2) methyl cinnamate are aromatic esters with high commercial interest.

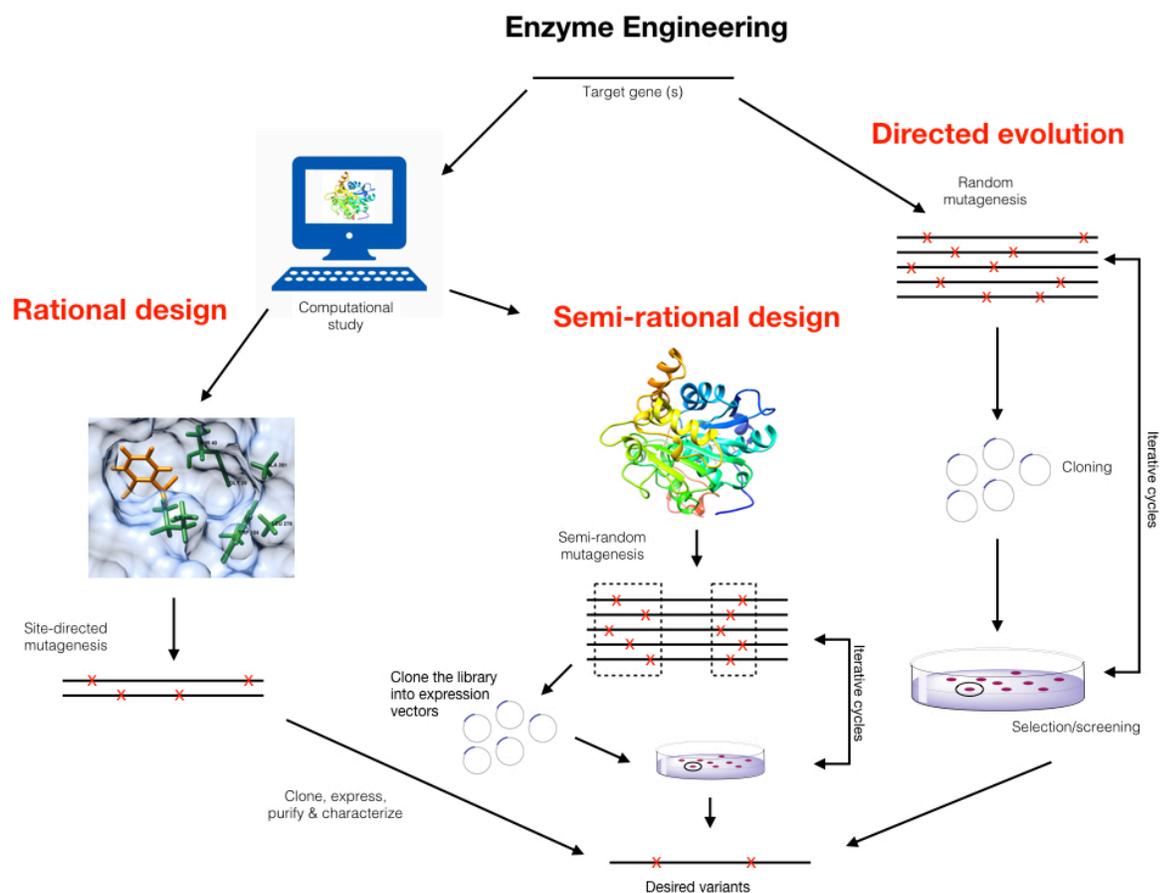
However, the high biotechnological potential of CalB makes it an ideal target for enzyme engineering efforts to achieve further functional improvements. This is also supported by an extensive collection of results concerning the engineering of the  $\alpha/\beta$ -hydrolase framework in general, and lipase in particular [14,15,16].

## 1.6. Enzyme engineering

To this day, one of the most useful enzyme engineering approaches to generate improved enzyme activity is directed evolution. Directed evolution is an effective molecular biology method to improve or alter enzyme activity and specificity for a broad variety of applications. While natural evolution generates iterated mutations providing functional and biological advantages to evolving organisms, natural selection processes seldom offer solutions adapted to the interest of human application. Additionally, spontaneous mutations occur at very low time scales and are generally insufficient for practical and commercial laboratory evolution. Directed evolution overcomes these limitations and provides means to access useful phenotypes on an accelerated time scale. Several studies have highlighted the power of the technique, including the improvement of natural enzyme activities, even using substrates unknown to exist in nature [24]. Directed evolution introduces randomly distributed, low frequency mutations in a gene of interest, followed by selection (or screening) of the desired properties among all protein variants of the mutated library generated (Figure 1.7).

This approach enables the relatively fast engineering of enzymes without in depth knowledge of the structure-function relationship. However, the large number of mutants that must be screened to obtain a desired effect on enzyme activity requires high-throughput screening capabilities. This methodology also precludes complete randomization of all positions of the protein at once. Indeed, complete randomization of a simple decapeptide would yield  $10^{13}$  unique combinations of amino acids, which exceeds an achievable library size to screen. Another weakness of this procedure is the random nature of the mutations introduced. In cases where functional and structural

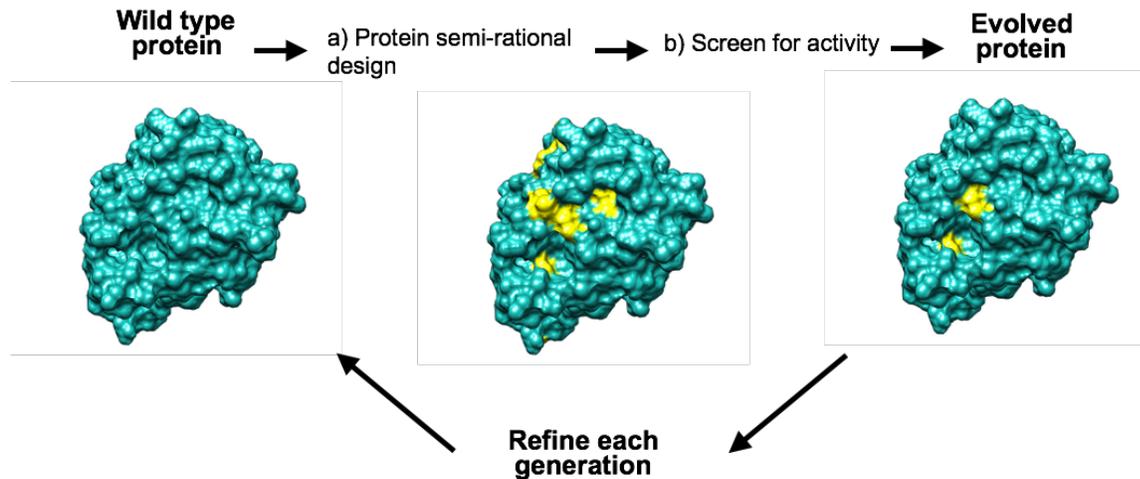
information are available, it is significantly more advantageous to concentrate mutations where they will be most effective [25]. For an enzyme with three-dimensional structural information available (such as CalB), one could also justify the use of rational design as a proper means to improve substrate specificity and/or catalytic activity. However, this approach requires extensive knowledge on how the structural features of the enzyme and active site precisely contribute to substrate recognition, stabilization, discrimination, and function. This is not easily predictable. Many proteins are structurally characterized at sufficient resolution to implicate specific residues in substrate binding or catalysis. As a result, combining the randomized nature of directed evolution with the positional information of rational design is a very powerful approach to evolve new enzyme properties [25]. Considering the vast structural and functional information available for CalB, the use of a semi-rational approach to increase catalytic efficiency is particularly suitable.



**Figure 1.7** Semi-rational design is an enzyme engineering approach to access useful phenotypes on an accelerated time scale. Semi-rational design is the combination of the randomized nature of directed evolution with the positional information of rational design to evolve new enzyme properties such as synthesis of industrially relevant fragrance compounds. Modified from ref. [26].

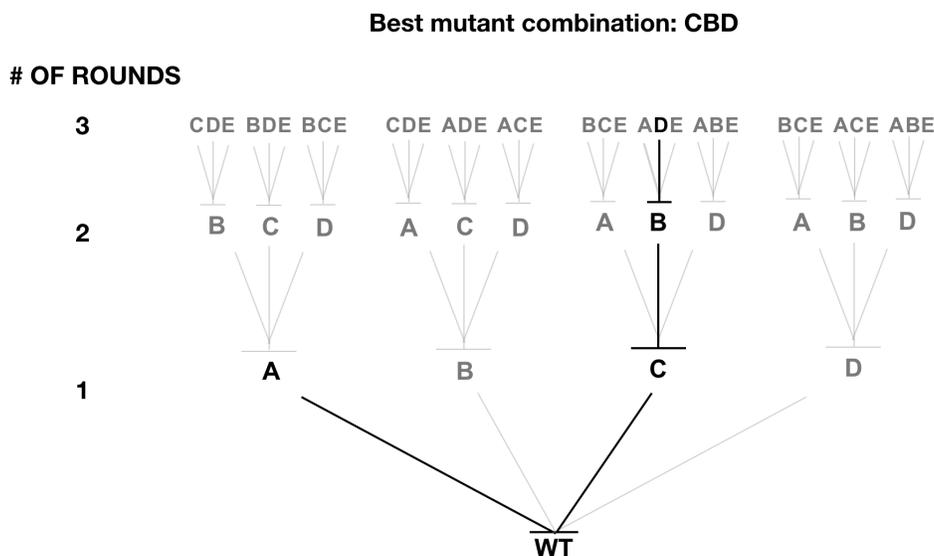
### 1.7. Semi-rational design of CalB to optimize synthetic activity

Several studies have shown the power of structure-based semi-rational approaches targeting active-site residues in enzyme engineering, namely by enhancing our capacity to perform rational design [27-32]. In this approach, the choice of the positions to mutate remains rational, but the choice of amino acid replacements is random (Figure 1.8).



**Figure 1.8** Semi-rational evolution strategy to increase CalB fitness for the synthesis of methyl salicylate and methyl cinnamate. a) Protein semi-rational design involves a computationally-guided approach to generate “smart libraries” prior to the creation of experimentally built libraries at optimized hot spots (in yellow) with saturated mutagenesis, b) Enzyme activity in organic solvent conditions must be tested on all library mutants to screen for improved esterification reaction.

The various rounds of evolution are planned to generate a combinatorial site-saturation mutagenesis biased towards residues located at or close to the active site. This approach can be achieved by an Iterative Saturation Mutagenesis strategy (ISM), leading to the incorporation of favorable mutations in sequential fashion (Figure 1.9). Each mutational replacement will further increase chances of selecting improved variants, resulting in the reduction of screening effort [33]. In this work, selection of the experimental structure and function of CalB, in addition to the proper methods available to screen the mutant libraries, was based on prior knowledge. This rational engineering approach, in combination with combinatorial methods to fine-tune the functional effects of neighboring residues, could provide further insights into the structural changes required to extend CalB substrate specificity in ester synthesis.



**Figure 1.9** Example of Iterative Saturation Mutagenesis strategy (ISM). Four positions (A,B,C and D) of a wild type protein (WT) were mutated. Two variants show increased desired activity in the first round, and both improvement trajectories (in black) are followed until they stop showing increased activity (A, first round) or achieve desired efficiency (CBD, third round).

## 1.8. Computational approach to optimize targeted positions for mutation

In addition to experimental methodologies, computational methods have been developed to uncover target positions involved with the stabilization of the enzyme-substrate complex, identifying “hot spots” that are most likely to yield active improvements for desired ligands upon mutation [34]. The optimal docking structure is chosen based on its lowest conformational energy, and used to identify near-optimal hot spots through simulated docking. This usually defines the library to be designed and experimentally screened. This protein evolution approach is an iterative exercise, as shown in Figure 1.7. Both the computational protein design (further explained in methodology section) and experimental screening efforts are important points to be developed in the present project.

## 2. CHAPTER 2. HYPOTHESIS AND RESEARCH OBJECTIVES

Our main hypothesis is that the use of a semi-rational and computationally-guided protein engineering approach on the CalB lipase can alter substrate specificity and catalytic activity to achieve efficient synthesis of the industrially relevant fragrance compounds methyl salicylate (ice-mint flavor) and methyl cinnamate (strawberry flavor).

To answer this hypothesis, two main objectives were fixed:

- 1) Perform *in silico* molecular docking predictions on CalB to identify amino acid positions that are most involved in substrate recognition and stabilization at the active site.
- 2) Based on our computational predictions, build iterative site-directed mutagenesis libraries and screen for improved CalB synthetic activity.

### 2.1. Methodology

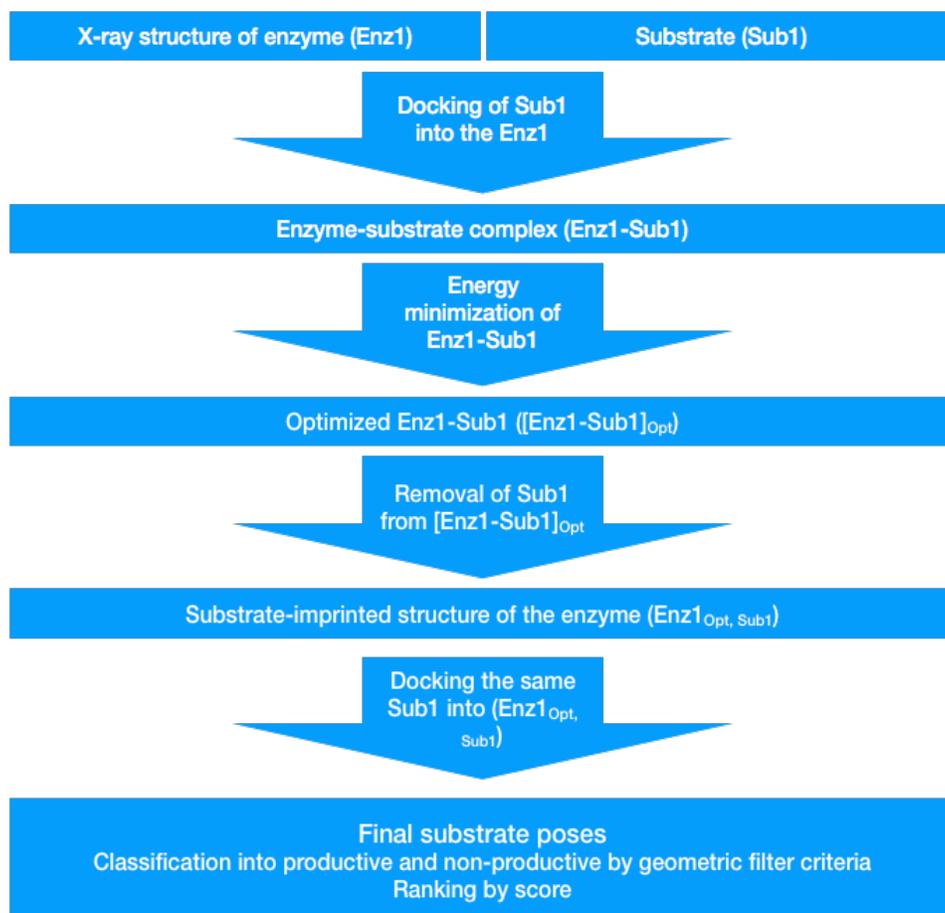
#### 2.1.1. Objective 1. Perform *in silico* predictions to identify amino acid positions most involved in substrate recognition through molecular docking

Molecular docking is commonly used to predict affinity, activity, specificity, and protein selectivity based on structural information [34]. In this case, recognition of both salicylic and cinnamic acid substrates can be explored by virtual docking, which uses a scoring function algorithm that makes possible the prediction of relative binding free energies for proteins and their substrates. Lipase structures are adequate for substrate-imprinted docking. Several lipases were subjected to a strategy of substrate-imprinted docking, using the docking program FlexX [35], adding a geometric filter criterium, which consists in pondering angles and distances of the hydrogen bonds in the substrate-enzyme configuration [34], and structure optimization by molecular mechanics to account for full protein flexibility. This approach is based on the simulation of an induced fit mechanism

between the enzyme and its substrates upon ligand binding. The process involves two rounds of virtual docking optimization, where the most energetically favorable structure of the first round is used as a template for side-chain energy minimization in the catalytic pocket of the enzyme [36] (see chapter 3). Study cases to model enantioselectivity and substrate specificity were previously performed on: a) the wild type of CalB compared to a mutant (W104A) with altered enantioselectivity [34] by docking the two enantiomers of 1-phenylethyl butyrate to model enantioselectivity, b) the enantiomers of 2- to 8-methyldecanoic acid butyl esters docked into *Candida rugosa* lipase to assess the capabilities of modelling lower enantioselectivities, and c) *Burkholderia cepacia* lipase compared by docking the enantiomers of 2-hydroxyoctanoic acid butyl ester and 2- to 4- methylpentanoic acid pentyl esters in order to model enantioselectivity and substrate specificity [15].

Substrate-imprinted docking showed more than 80% accuracy for the three structures modelled, further demonstrating that the procedure can be applied to docking of new substrates. The strategy can improve binding pockets that are partially formed by backbone atoms, such as the oxyanion hole of lipases and esterases. Computational work was performed in Molegro Virtual Docker 6.0 and run under “organic conditions”, *i.e.* subtraction of water molecules in the framework to prevent hydrolysis conditions. This project will use the same strategy for docking salicylic and cinnamic acid, to obtain the desired products: methyl salicylate and methyl cinnamate, accordingly, and is described in the six steps illustrated in Figure 2.1. Selection criteria of best residue orientations will be performed according to the docking score function of MolDock. This search algorithm is used in combination with a cavity prediction algorithm [36]. Best poses of the first calculation round will be used as a template for side-chain energy minimization in the catalytic pocket of CalB, with the use of the algorithm Nelder-Mead simplex iteration [37]. This algorithm independently minimizes a specified number of steps for each residue, and then performs a global energy minimization of all residues simultaneously.

From the productive amino acid poses and active-site cavity estimation, amino acid replacements that are the most energetically favorable for the binding recognition of these compounds will be selected to build experimental combinatorial libraries and to screen for enhanced synthetic activity.



**Figure 2.1** Flowchart of substrate-imprinted docking strategy. Starting from one enzyme structure (CalB) and one substrate (Sub1, that are salicylic or cinnamic acid), the substrate is docked into the structure, where the side chains of the catalytic serine and histidine are in a functional orientation, in a first round of docking. The best pose from the first docking is used to construct an enzyme-substrate complex (Enz1-Sub1), whose lateral chains are then energy minimized, giving an optimised enzyme-substrate complex ([Enz1-Sub1]<sub>Opt</sub>). The substrate is removed from this optimised complex, yielding a substrate-imprinted enzyme structure (Enz1<sub>Opt, Sub1</sub>). This structure-imprinted structure is used in a second round of docking of the same substrate (Sub1). The resulting substrate poses are classified according to geometric filter criteria into productive and non-productive, and ranked by docking score. Modified from ref. [34].

### 2.1.2. Objective 2. Build and screen CaIB libraries for synthetic activity

There are random and focused approaches to creating library diversification. Random approaches allow for high mutational rates, but also create unrealistically achievable library size to screen. Site-directed mutagenesis provides the advantage of focusing attention on functionally relevant residues that significantly increase library quality and the likelihood of screening for improved enzyme variants. Focused mutagenesis strategies can include the use of restriction enzyme cloning or gene assembly protocols. For focused approaches, the site-directed saturation mutagenesis NNK and NNS codons can be used on mutagenic primers to fully explore protein sequence space for selected residue sites. NNK and NNS codons encode all 20 amino acid possibilities, where N can be any of the four nucleotides, K can be G or T, and S can be G or C. However, because the NNK/S degeneracy still contains three codons for arginine, leucine, and serine and two codons for the five amino acids alanine, glycine, proline, threonine, and valine, redundancy is not completely eliminated and a certain amino acid bias is still present [40]. Consequently, along with the assembly protocol, a scheme to reduce the overall number of variants to eliminate genetic code redundancy can be undertaken. Twenty canonical amino acids are disproportionately coded by 61 sense codons, *i.e.* some amino acids are encoded only by a single codon (*e.g.* methionine and tryptophan), while others are encoded by up to six different codons (*e.g.* arginine, serine, and leucine) [38]. The 2-fold reduction from the original 64 codons can be further reduced to 22 codons by mixing a total of three oligonucleotides: two degeneracy carrying primers, *i.e.* one with NDT (12 unique codons) and the other with VHG (9 unique codons), and a third TGG-containing primer (Trp). Therefore, the genetic code redundancy is reduced up to a codon to amino acid ratio of 22:20. This mixture contains no stop codons and only two redundant sets for valine (GTT, GTG) and leucine (CTT, CTG) [38].

Once the genetic constructions are designed, two additional factors need to be considered to reduce library screening effort: library coverage and full coverage probability. Library coverage is the total number of bacterial colonies obtained after transformation, *i.e.* the proportion of the probed variant space (total amount of

theoretical possible variants) of a generated library and the proportion of the variant space covered by picking a certain number of samples. The second factor is the full coverage probability, which calculates the likelihood of sampling the complete variant space [38]. These factors can be mathematically considered by an approximation of the Poisson distribution (Equation 1), resulting in an estimation of completeness and equiprobable diversity in protein-encoding libraries [39]. In the present work (see chapter 3), our estimation resulted in sixty-six colonies randomly selected to test for every library (22 variants). That proportion should cover 95% of the total mutants in every bank so that each variant is represented at least once.

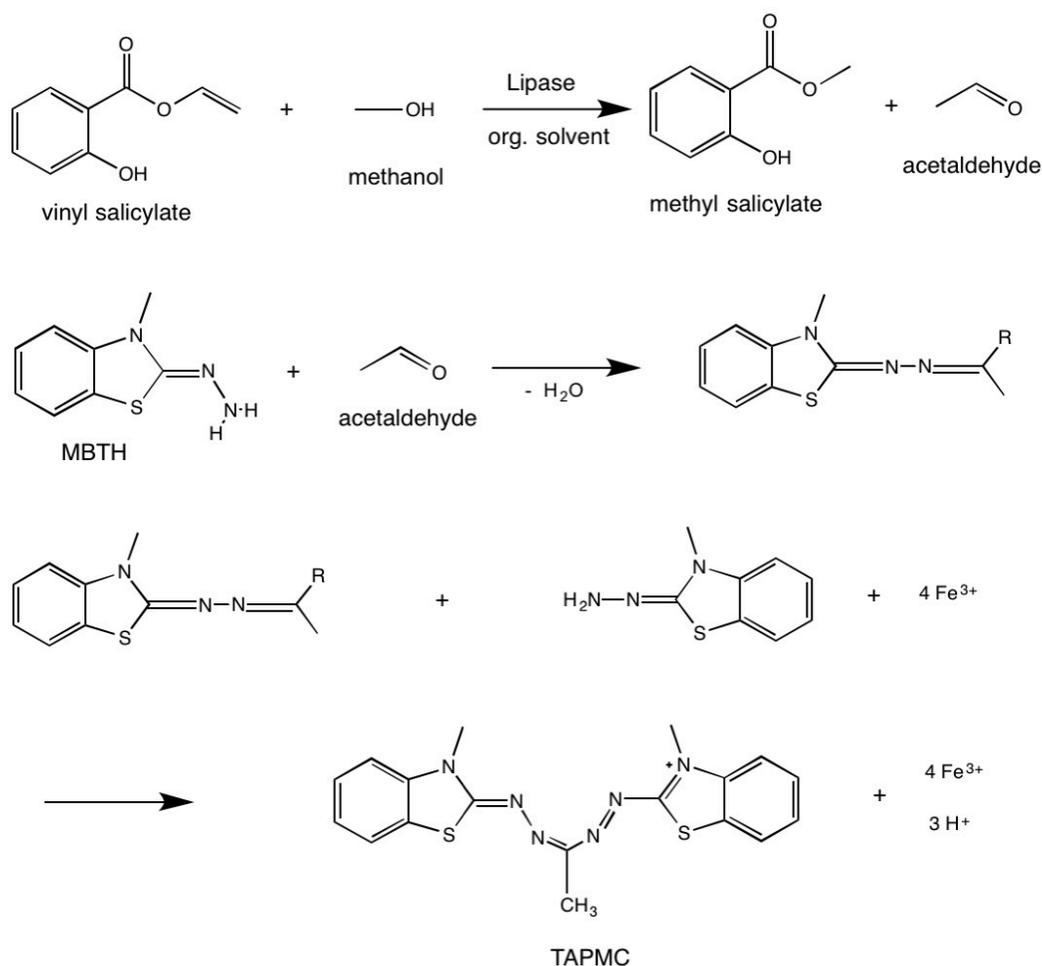
$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots$$

**Equation 1. Poisson distribution for the estimation of equiprobable variant diversity in protein-encoding libraries.**

where  $P(x)$  denotes the probability that the possible variant occurs;  
 $x$  times in the library and  $\lambda$  denotes the mean number of occurrences for possible variants.

Due to library size, high-throughput screening and mutant selection rely on the rapid detection of the accumulation (or disappearance) of a chromogenic compound by optical means such as colour, fluorescence, luminescence or turbidity. However, neither methyl salicylate nor methyl cinnamate production by CalB is associated with direct observable phenotypes, and therefore requires a fluorescent, colorimetric or other readily detectable reporter. As a result, a surrogate substrate screening approach needed to be adapted to detect improved CalB activity against these two flavoring compounds. Upon catalysis, an appropriate analogue needs to exhibit chemical rearrangements that lead to altered optical properties, including visible colour in this case. As a crucial active ester, vinyl acetate is widely used as acylating agent in lipase-catalyzed reactions, especially in chiral resolution of alcohol or amine mixtures [40]. Acetaldehyde and vinyl alcohol are stoichiometrically released from lipase-catalyzed ketoenol tautomerization between vinyl acetate and alcohol [41]. This particular chemical reaction using vinyl salicylate and vinyl cinnamate as surrogate substrates is well suited to the current CalB evolution project due to aromatic substrate similarity

between these compounds and the desired flavoring agents methyl salicylate and methyl cinnamate (Figures 1.6 and 2.2). Selecting the appropriate screening method is a critical design strategy, as evolved CalB variants will only effectively recognize proper methyl salicylate and methyl cinnamate precursors if the surrogate substrates used in the screening process are chemically similar (“*you get what you screen for*”) [42].



**Figure 2.2** Model reaction example on vinyl salicylate for the evaluation of lipase synthetic activity assay. *Modified from ref. [43].*

We used a colorimetric high-throughput screening method for determining lipase synthetic activity using the vinyl salicylate analogue (Figure 2.2) that had previously been successfully applied for assaying the transesterification between vinyl acetate and *n*-butanol [43]. The determination of released acetaldehyde was based on the 3-methyl-

2-benzothialinone (MBTH) derivatization method. MBTH reacts with acetaldehyde to the corresponding aldazine, which is converted to a blue-colored tetraaza-pentamethincyanine (TAPMC) dye under oxidative coupling with another MBTH molecule [44]. The content of acetaldehyde can be quantified based on the absorbance of TAPMC detected in a 96-well microplate using a Multi-Mode microplate reader [43]. To investigate the optimal absorption wavelength for acetaldehyde detection by the MBTH derivatization method, 400 to 800 nm wavelength scanning of MBTH and the corresponding derivative of acetaldehyde can be carried out in a Multi-Mode microplate reader. Afterwards, the standard curve to correlate absorbance intensity with the concentration of acetaldehyde can be used to quantify the conversion for each reaction. After library construction, an enzymatic assay can be performed under organic conditions to favor synthesis over hydrolysis activity (Figure 2.3). The surrogate analogues for cinnamic acid and salicylic acid used in the present study are vinyl cinnamate and vinyl salicylate, respectively.



**Figure 2.3** General screening strategy employed in the current study. Colonies of isolated variants (step 1) are individually picked and transferred to liquid culture (step 2). After growth, cultures are lyophilized (step 3) to allow resuspension in organic medium required for lipase esterification conditions (step 4). Measurements are performed using a colorimetric reporter substrate and an automated plate reader (step 5).

### **3. CHAPTER 3. MANUSCRIPT SECTIONS**

In addition to specific sections of the introduction and methodology, the following chapter will be used in its entirety as part of a manuscript in preparation to be submitted for publication in the first few months of 2018. The complete and final version of this manuscript will be included in the final version of the present M.Sc. thesis and prior to submission for publication, taking into account comments of M.Sc. jury members.

#### **3.1. Author contributions**

Ying Lian Chew Fajardo constructed gene libraries and performed activity screens for the mutants encoded in six libraries corresponding to single point mutations in the first generation (G1) of salicylic acid substrate screening: T138X, I189X, D134X, T40X, V154X and Q157X. She also executed molecular docking experiments to design the second generation of libraries (G2) and *in vitro* screening of the ISM framework: T40X, I189X, W104X, G39X, A281X, and L278X. Finally, she pursued the evolution of the CalB lipase for enhanced recognition of salicylic acid by designing, constructing, and screening a final third generation of mutant libraries (G3). Postdoctoral fellow Yossef Lopez de Los Santos developed and standardized the directed evolution program, from the design of the libraries to the virtual and *in vitro* screenings for cinnamic acid. He also supervised Ying Lian Chew Fajardo in the execution and improvement of CalB toward salicylic acid synthesis. Guillaume Brault and Nicolas Doucet initiated the CalB project idea aimed at improving an existing lipase to synthesize flavoring agents for commercial purposes. Nicolas Doucet and Guillaume Brault also secured financing of the project through a Mitacs collaboration with an industrial partner. Nicolas Doucet supervised the overall project for the evolution of CalB towards both flavoring compounds: cinnamic and salicylic acid. All authors participated in writing and revision of the manuscript.

## **3.2. Materials and methods**

### **3.2.1. DNA constructs and expression systems**

A codon-optimized gene of wild-type CalB was synthesized for optimal expression in *E. coli* (GenScript) and cloned into expression vector pET22b+ (MilliporeSigma). A C-terminal hexa-histidine Tag was designed to facilitate the purification process of all CalB variants. Using the same codon-optimized wild type (WT) CalB sequence, three previously reported mutants of CalB displaying increased activity to bulky substrates in non-aqueous conditions were also synthesized: S47L, I189A, and double mutant I189A-L278P [45, 46, 47]. Overexpression of the CalB enzyme from IPTG-inducible vector pET22b+ was tested using five different *E. coli* strains: Rosetta-Gami (DE3), Rosetta (DE3), Rosetta 2 (DE3) PLacl, Origami 2 (DE3), and BL21 (DE3) (MilliporeSigma). The selection markers employed were a combination of kanamycin and chloramphenicol (Rosetta-Gami (DE3), chloramphenicol (Rosetta (DE3) and Rosetta 2 (DE3)-PLacl), and streptomycin (Origami 2 (DE3)). Carbenicillin was used as resistance marker for expression vector pET22b+, which was electro-transformed into each strain using 0.1 cm Gene Pulser/MicroPulser electroporation cuvettes (Bio-Rad) and an ECM 630 electroporator (BTX Harvard Apparatus). Cells were recovered in 1 mL of SOB medium [48] supplemented with 20  $\mu$ L of 2 M glucose (filter-sterilized) and 5  $\mu$ L of 2 M magnesium (1 M MgSO<sub>4</sub> and 1 M MgCl<sub>2</sub>, autoclave-sterilized)..

### **3.2.2. Wild type and mutant CalB overexpression**

All CalB constructs were initially pre-cultured overnight in Lysogeny Broth (LB) medium at 37°C and 250 rpm [48]. Expression cultures of 5 mL were then performed in two different liquid growth media (LB and Super Broth, or SB) incubated at 16°C and 37°C for 16 hours (250 rpm). Protein expression was induced at OD<sub>600</sub>=0.5-0.6 upon addition of varying concentrations of IPTG (0.1 to 1 mM). SB medium was prepared using 32 g of polypeptone (Bio Basic), 20 g yeast extract (ThermoFisher), 5 g NaCl (Bio Basic), and 5 mL 1 M NaOH in a total of 1 L of water (Larsen et al., 2008). After incubation, cells were collected by centrifugation (21,000g, 5 minutes, 4°C) and resuspended in 200  $\mu$ L of denaturing buffer for sonication (750  $\mu$ L of double-distilled water, 200  $\mu$ L of 10 %

sodium dodecyl sulfate (SDS), and 50  $\mu$ L of 1 M dithiothreitol (DTT). After sonication, the culture supernatant was separated by centrifugation (21,000g, 20 minutes, 4°C) and cell pellets were resuspended in 200  $\mu$ L of the same denaturing buffer. All protein supernatant and pellet fractions were quantified and analyzed using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Mini-PROTEAN TGX Precast Gels, Bio-Rad) and Western-Blot analyses using an anti-His-Tag HRP conjugate mouse monoclonal antibody (Qiagen) and a TMB substrate (G-Biosciences).

### **3.2.3. Wild type and mutant CalB esterification assays**

To confirm proper refolding and enzymatic activity of WT CalB and mutant forms, esterification assays were performed using a preculture of *E. coli* Rosetta (DE3) incubated overnight at 37°C. LB agar plates were inoculated with 5, 20 and 100  $\mu$ L of bacterial culture to obtain a bacterial lawn and evenly spread with crystal pearls to obtain individual colonies using carbenicillin as resistance marker and a combination of oleic acid and 1-decanol (Sigma) as substrates for the enzymatic reaction to occur (15.62 g of each substrate per liter of medium). A concentration of 0.001% (w/v) rhodamine B (R-6626 rhodamine O) was used as reporter of esterification activity, as previously reported [49]. Agar plate incubation was performed for 72 hours at 30°C.

### **3.2.4. Enzyme extraction and lyophilization**

Selected colonies were picked and grown overnight in liquid LB medium at 37°C (250 rpm) with carbenicillin as resistance marker. Precultures were used to inoculate a final volume of 5 mL SB culture in presence of carbenicillin, using the overexpression procedure described above (0.1 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) induction, 48 h incubation, 16°C, 250 rpm). After incubation, cells were centrifuged at 2,400g (4°C, 15 minutes) and then resuspended in phosphate buffer (50 mg  $K_2HPO_4$ , 4.9 g KCl dissolved MilliQ  $H_2O$  in a final volume of 200 mL. [50]. Buffer pH was adjusted to pH = 7.0 using 100 mM HCl. Samples were sonicated on ice for 15 seconds and centrifuged at 2,400g (4°C, 20 minutes) to extract cell supernatant. Lyophilization of the supernatant was performed for 72 hours at -56°C and the final lyophilized samples were stored at 4°C in a desiccator with drierite for future enzymatic assays.

### **3.2.5. Virtual docking of substrates in the CalB active site**

To identify which amino acid residues are interacting with the substrate in the active site binding cleft of CalB, a substrate-imprinted docking approach was employed [34]. This approach is based on the simulation of an induced fit mechanism between the enzyme and its substrates upon ligand binding. The process involves two rounds of virtual docking optimization, where the most energetically favorable structure of the first round is used as a template for side-chain energy minimization in the catalytic pocket of the enzyme. Calculations employed the Piecewise Linear Potentials for steric and hydrogen bonding interactions, and the Coulomb potentials for electrostatic forces [36]. The Nelder-Mead simplex iteration algorithm was used for all steps of energy minimization [37]. This algorithm independently minimizes a specified number of steps for each residue, and then performs a global energy minimization of all residues simultaneously. The second round of virtual docking was performed using the newly calculated configuration of the CalB active-site binding pocket. The five most energetically favorable active-site configurations from the second round of virtual docking were sampled and compared to obtain the best residue targets guiding the site-directed saturation mutagenesis and experimental library design. The MolDock docking scoring function was used for all virtual docking simulations. Ligands cinnamic acid (PubChem CID 444539), methyl cinnamate (PubChem CID 637520), salicylic acid (PubChem CID 338), and methyl salicylate (zinc\_490 mol file) were used against the crystallographically-resolved CalB template (PDB entry 1TCA) [51]. The scoring function is based on the guided Differential Evolution (DE) hybrid search algorithm, in combination with a cavity prediction algorithm [36]. All computational work was realized with the Molegro Virtual Docker 6.0 (MVD) suite without incorporation of water molecules. To maintain searching robustness, twenty rounds of iterations were performed for each docking procedure.

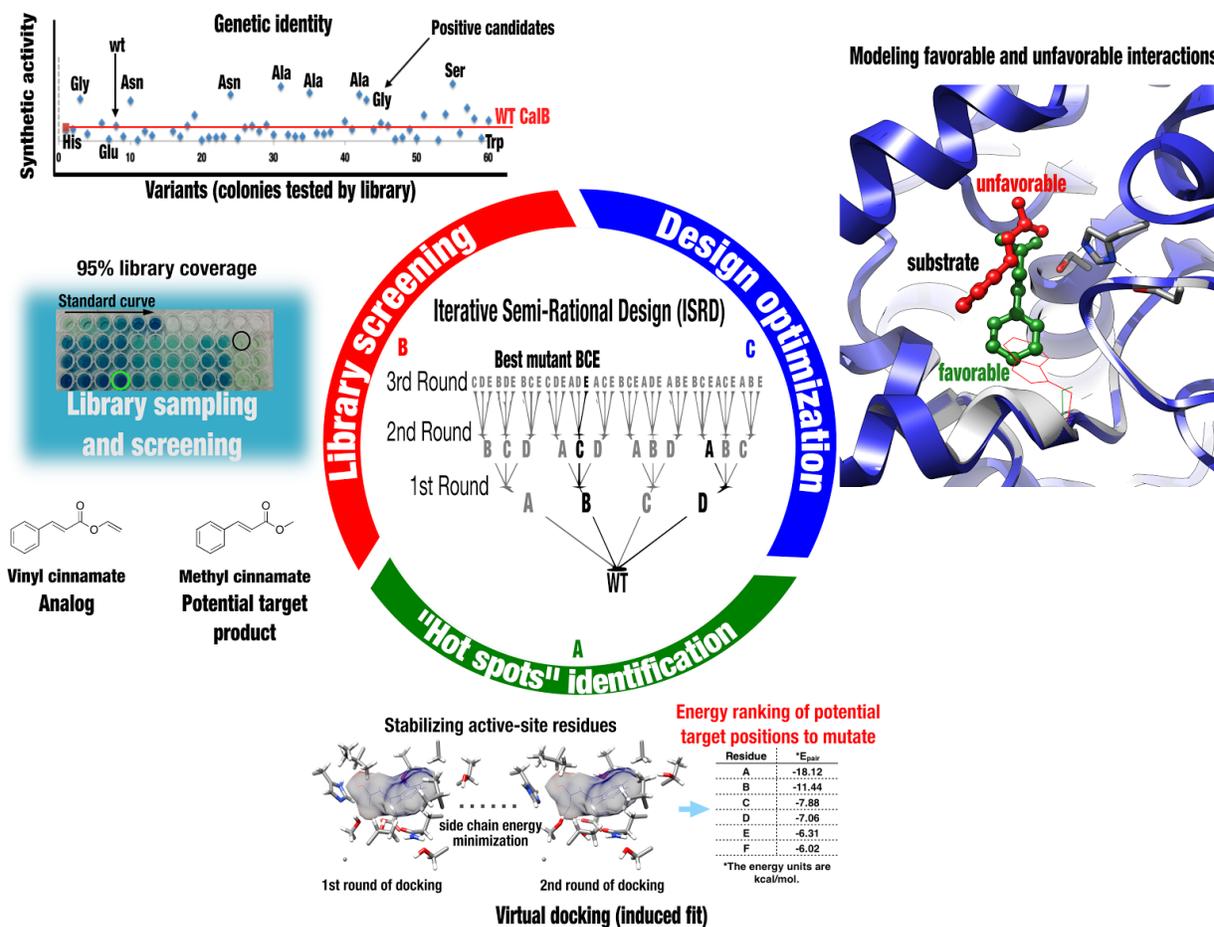
### **3.2.6. Iterative semi-rational evolution of CalB towards bulky substrates**

To overcome a number of limitations inherent to typical directed evolution strategies in enzyme design, we used an iterative improvement strategy combining a series of

computational and experimental approaches to modify active-site architecture, geometry, and affinity in CalB, providing means to alter its synthetic capabilities. This Iterative Semi-Rational Design (ISRD) strategy was performed in three complementary steps (Figure 3): (1) identification of potential 'hot spots' for substrate recognition and stabilization in the CalB active site using a combination of virtual docking and Residue Interaction Network (RIN) analyses [59]; (2) experimental design and screening of mutant libraries, providing means to select variants displaying improved transesterification; (3) *in silico* analysis of the most active variant(s) by molecular modeling to provide structural and energetic feedback for the design of subsequent generations of mutant libraries.

### **3.2.7. Library Design Strategy**

Based on virtual docking predictions, an individual saturation mutagenesis library was designed for each selected position in the active site of CalB. For every position identified, a specific saturation mutagenesis library derived from the WT CalB sequence was acquired by gene synthesis. Libraries for the first two-generation screening of both substrates (cinnamic and salicylic substrate analogs) were obtained through the gene synthesis service of Life Technologies (Thermo Fisher Scientific). The third generation library for both evolutionary pathways was obtained by the Gibson method [52], in combination with a combinatorial oligonucleotide method using degenerate codons NDT, VHG and TGG. A codon bias reduced the number of necessary codons from 64 to 22, thus balancing the proportion of mutations and avoiding over-representation of certain amino acids within the libraries [38]. This strategy was coherent with the Iterative Saturation Mutagenesis (ISM) approach (Figure 3, center), where the best mutations selected in one round are fixed and used as starting template for subsequent library design. This approach aims to reduce screening efforts and maximize the selection of beneficial mutations [33].



**Figure 3** Schematic representation of the Iterative Semi-Rational Design (ISRD) strategy employed in the present work. A) Identification of active-site mutational 'hot spots' by mimicking the induced fit mechanism of the enzyme using virtual docking. Targeted active-site positions are selected based on ranking of energetic contributions for each docking complex and residue interaction likelihood with the desired substrates. B) Building and screening of the mutant libraries using a vinyl analog to measure the synthetic activity of CalB in organic solvent conditions (see Materials & Methods for details). Genetic identification of the most active mutants allows selection of the best candidates as template(s) for the next round of mutational selection. C) Modeling of experimentally selected variants with favorable active-site replacements as template for the next round of virtual docking (A). This complementary procedure combining experimental and computational approaches is iteratively repeated to improve esterification activity catalyzed by CalB. In the present work, three generations of CalB variant libraries were designed.

### 3.2.8. Screening of mutant libraries and transesterification assays

Screening efforts required to statistically explore the diversity of the different designer libraries was estimated using the following equation:

$$F = 1 - e^{-\frac{L}{V}}$$

where F is the fractional completeness;  
V the number of possible sequence variants;  
and L the number of clones required to reach fractional completeness [39].

We used the ISM strategy, whereby independent single positions are mutated to saturation [33]. Using the degenerate codon strategy explained above, 22 possible codon combinations (V) were obtained, with a 95% fractional completeness (F). Consequently, following the previous equation our screening effort required sequencing of 66 clones per library to explore 95% diversity:

Sixty-six colonies were randomly picked and sequenced in each library, thus ensuring 95 % coverage where each mutant is represented at least once [39]. Primers T7 (5-TAATACGACTCACTATAGGG-3) and T7<sub>terminator</sub> (5-GCTAGTTATTGCTCAGCGG-3) were used to verify the genetic identity of all CalB variants. DNA sequencing of twenty independent colonies was used to confirm sequence coverage of each library. Each picked colony (*i.e.* each CalB variant) was expressed as described above. After lyophilization, enzymatic screening was performed on cell extracts (5 mg per sample) using 2-methyl-2-propanol (tert-butanol) (Sigma) as organic co-solvent. Substrates were used on a 3:1 ratio of methanol (Sigma) and vinyl cinnamate (or vinyl salicylate) (Aldrich), two analogs of the target substrates cinnamic acid and salicylic acid. Enzyme assay conditions were fixed at 100 mM vinyl cinnamate or 50mM vinyl salicylate, and 300 mM or 150 mM methanol for *v.* cinnamate or salicylate respectively. Samples were incubated at 30°C for 2 hours and 1,000 rpm agitation in a final volume of 0.5 ml. Deionized water content was fixed at 3% v/v (see Figure S3). After reaction completion, 3-methyl-2-benzothianoline (MBTH) was used as reporter to detect the stoichiometric ratio (1:1) between released acetaldehyde and the reduction of one molecule of vinyl

analog [43]. Enzymatic samples were first diluted in deionized water (1/20 to 1/250) to allow proper absorbance reading of the standard curve in presence of acetaldehyde. 200  $\mu$ l of the diluted solution was further mixed with 200  $\mu$ l MBTH solution (0.1% m/v in deionized water) and incubated for 10 minutes at 30°C and 200 rpm. After incubation, 80  $\mu$ l of the derivatization solution was added (1% m/v  $\text{H}_4\text{FeNO}_4\text{S}_2 \cdot \text{H}_2\text{O}$  in a solution of 0.1 M HCl prepared with deionized water) and incubated for an additional 30 minutes at 30°C (200 rpm). Acetaldehyde concentration was obtained by monitoring absorbance at 600 nm with a Tecan Infinite M1000 PRO microplate reader. A standard curve was calibrated using 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5 mM acetaldehyde.

### **3.2.9. Bioinformatics**

UCSF Chimera (version 1.1) was used for visualization and molecular analysis of all structures [53]. The RosettaBackrub server was used to predict mutational effects on the protein backbone, surface, and volume of the catalytic pocket in CalB after site-saturation mutagenesis [54]. A new round of virtual docking was performed between each generation of ISRD library design (Figure 3). The CASTp service was employed to analyze topographic surface changes in the catalytic cavity upon mutation [55]. The Evolutionary Trace Analysis was used to obtain conservation scores for residues of the  $\alpha/\beta$  hydrolase superfamily with the ConSurf server [56]. The WebLogo server was employed to generate sequence images of the evolutionary conserved residues from the Consurf server and the best mutations selected in this project [57].

## **3.3. Results**

### **3.3.1. Optimization of CalB overexpression**

To develop a reliable expression system aimed at improving esterification of bulky substrates in CalB, we tested five *E. coli* strains in combination with a codon-optimized CalB gene expressed on IPTG-inducible T7 vector pET22b(+) (see Materials & Methods). After careful optimization of strain expression, growth temperature, culture

media, and IPTG concentration, the most efficient CalB producer was observed for *E. coli* Rosetta (DE3) grown on SB medium (0.1 mM IPTG induction, 16°C) (Figures S1-S2). Using these experimental conditions, we also demonstrated that this expression system could reliably produce an active recombinant form of CalB that efficiently catalyzes esterification of 1-decanol and oleic acid (Figure S2B), further providing an efficient solid-state screening medium for activity.

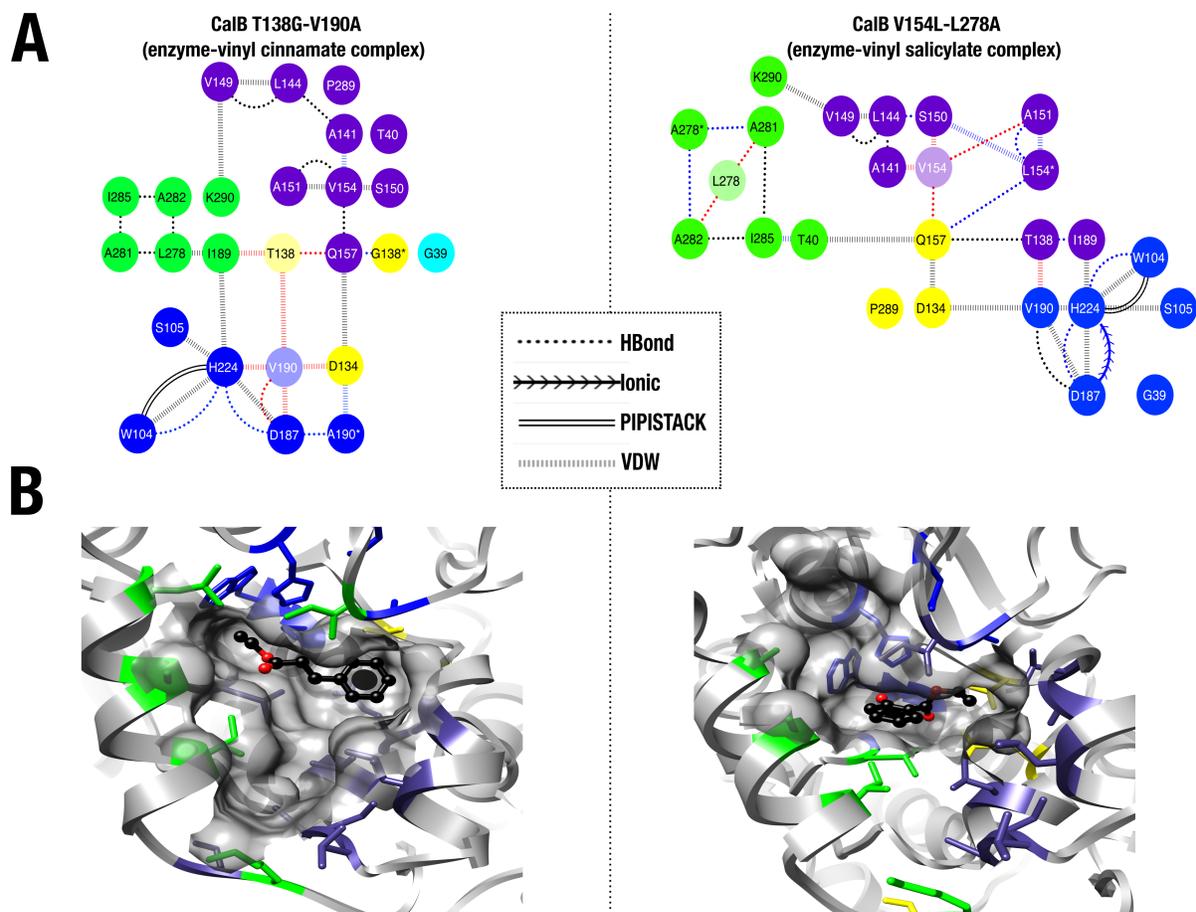
### 3.3.2. High-throughput screening of lipase transesterification

Based on a previous report [43], we adapted and optimized a liquid medium CalB synthetic activity screening assay using vinyl analogues to mimic cinnamic and salicylic acid compounds [43]. Since both the cinnamic and salicylic lipase-catalyzed products of interest are devoid of distinctive spectroscopic properties relative to their respective substrates, this high-throughput screening approach exploits the release of acetaldehyde and its reactivity with 3-methyl-2-benzothialinone (MBTH) to generate the corresponding aldazine moiety (Figure S3A). This aldazine is further converted to a blue-colored tetraaza-pentamethincyanine (TAPMC) dye under oxidative coupling with another MBTH molecule [44]. Effectively, an acetaldehyde molecule is released from each vinyl substrate analogue of the lipase-catalyzed transesterification reaction, affording colorimetric detection due to derivatization with reporter compound MBTH. Absorbance measurements therefore allow proper quantification of lipase transesterification. To demonstrate the efficiency of this approach, we tested four lipase variants that mimic potential CalB mutants to be engineered in the present work, *i.e.* WT CalB and three previously reported variants displaying modest synthetic activity improvements against bulky substrates (S47L, L278P, and the double mutant I189A-L278P) [46, 47, 58]. Figure S3 shows that WT CalB performs better in salt-activating organic conditions and low water concentration. Due to similar synthetic performance relative to the previously reported variant CalB-L278P, we selected WT CalB as a proper starting template for all future protein engineering steps (Figure S3C). To optimize screening conditions and to prepare the enzymatic extracts for proper organic solvent conditions. Prior to lyophilization, the samples were prepared in 3% water content (v/v) and a salt-activating potassium phosphate buffer (50mg of  $K_2HPO_4$  and

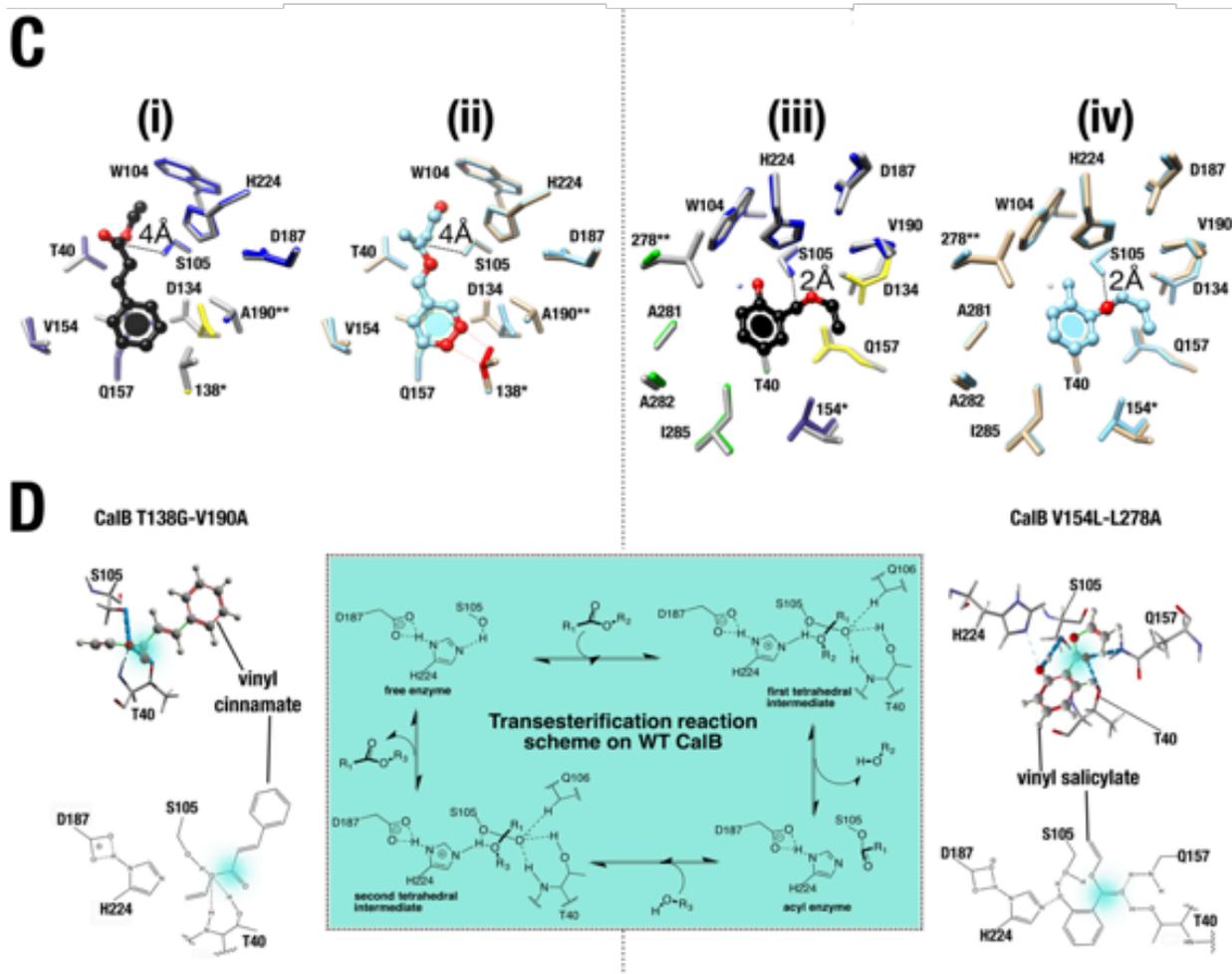
4.9g of KCL dissolved in 200ml of H<sub>2</sub>O MilliQ to reach to final concentration of 98% w/w). (Figure S3B-C).

### **3.3.3. Using ISRD to uncover and target active-site 'hot spots' for CalB activity improvement**

We first used a combination of virtual docking and Residue Interaction Network (RIN) analyses to guide the semi-rational design of CalB for transesterification improvement. Docking simulations were first performed to uncover which residues are directly involved with substrate positioning and stabilization prior to reaching the transition state of the catalytic reaction. The RIN analysis was implemented to identify those residues exhibiting a higher number of connection nodes throughout the network, further offering additional control over cavity network modules [59]. Since the scoring function (Moldock) and the substrate-imprinted docking procedure both consider conformational sampling of residue side-chains involved in the docking complex, only the theoretically active configuration of the catalytic triad residues (D187, H224 and S105) was fixed during docking, thus mimicking transition state orientation [36]. The energy cut-off to differentiate the importance of individual residue contributions was set to  $-1.0 \text{ k}_{\text{cal}}/\text{mol}$  (Table 1). Our RIN analysis shows that the catalytic cavity of CalB is divided into three main residue modules, which describe clustering interactions defined by the Linkland algorithm [60, 61) (Figure 3.1.1-3.1.2). Over the course of all three generations presented in this work, systematic mutational exploration of these CalB active-site modules was performed.



**Figure 3.1.1** Structural changes in the active-site cavity of CalB upon mutagenesis and selection. **A)** Changes in the RINs under selective pressure shows that each evolutionary trajectory (vinyl cinnamate or vinyl salicylate) is preceded by module reorganization of the residues forming the catalytic pocket. Both pathways exhibit 4 main residue modules in the CalB cavity identified as blue, purple, green and yellow residue clusters sharing hydrogen bonding, electrostatic,  $\pi$ - $\pi$  stacking, and/or van der Waals interactions. Lost (gained) residue interactions upon mutation are labeled as red (blue) line symbols and the asterisks illustrate mutated residue positions. Each RIN illustrates the comparison between WT and variant enzyme-substrate complexes. **B)** Modeled enzyme-substrate complexes showing how vinyl cinnamate interacts with a larger active-site cavity in mutant T138G-V190A relative to vinyl salicylate in mutant V154L-L278A. Both variants stabilize substrate molecules by exploiting distinct physicochemical properties in different active-site areas.



**Figure 3.1.2** C) Both double mutants T138G-V190A and V154L-L278A favour stabilization of substrate aromatic rings, allowing proximal alignment of the carbonyl moiety 2-4 Å away from the catalytic S105 (dashed line). Sub-panels (i) and (iii) show WT side-chains in gray and mutant side-chains color-coded according to their respective module shown in panel A. Sub-panels (ii) and (iv) illustrate how active-site mutations allow for proximity with the substrates by preventing atomic clashes (red dashed lines and red colored atoms). Mutant and WT proteins are presented in blue and gray, respectively. D) Formation of a transition state-like pre-intermediate between the best CalB variants and their respective substrates. A schematic view of the transesterification reaction is provided (center), with variants T138G-V190A and V154L-L278A (left and right, respectively) depicted in 3D (top) and in schematic view (bottom). H-bonding interactions are illustrated by dashed lines and the carbonyl moiety is highlighted with a cyan dot.

Virtual docking of both cinnamic and salicylic acid analogs generated a number of energetically favorable interactions between the enzyme and substrates, highlighting individual ranking energy contributions for each residue position (Table 1). For the first

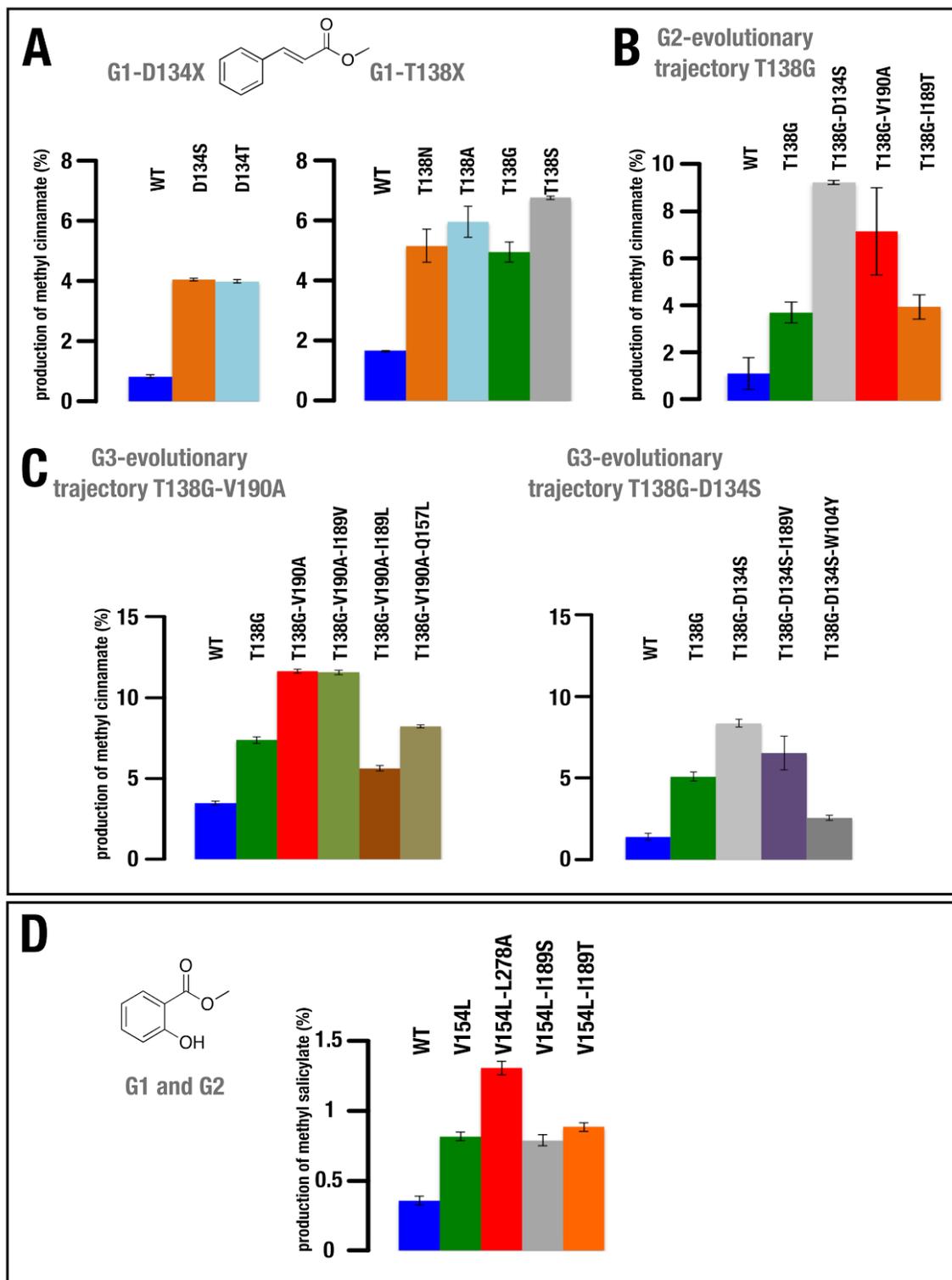
generation of mutants (G1), we designed six distinct site-saturation mutagenesis libraries at CalB positions 40, 134, 138, 154, 157, and 189, which were then individually screened against both substrate analogs. These positions were selected according to their favorable energy contributions to the stabilization of the enzyme-substrate complexes, increasing the likelihood of obtaining positive variants [25]. Residues H224 and S105 were excluded from this randomization because of their functional importance in the catalytic triad of CalB. For subsequent library generations (G2 and G3), mutational exploration of the active-site cavity was expanded to CalB positions 39, 104, 190, 278, and 285. For G1 and G2, single-site saturation libraries were individually screened against each substrate analog. For G3, we reduced the number of positions explored to investigate three individual single-site saturation libraries for two distinct templates obtained from previous generations. G3 was also different in that we explored the combination of successful mutations that were independently selected from G1 and G2.

Our screening efforts on G1, G2, and G3 libraries illustrate that the evolutionary trajectories of both vinyl analogs show significant differences with respect to favorable mutational solutions that benefit methyl cinnamate and methyl salicylate transesterification (Figure 3.2).

**Table 1. Binding energy contributions of enzyme-substrate complexes for WT CalB and the most efficient CalB variants selected after three rounds of Iterative Semi-Rational Design (ISRD). Note: Asterisks specify mutations that were selected for additional rounds.**

Ranking	Vinyl cinnamate				Vinyl salicylate			
	WT CalB		T138G/V190A		WT CalB		V154L/L278A	
	aa	k <sub>cal</sub> /mol	aa	k <sub>cal</sub> /mol	aa	k <sub>cal</sub> /mol	aa	k <sub>cal</sub> /mol
1	T40	-18.31	T40	-17.49	T40	-16.96	T40	-19.70
2	Q157	-14.85	Q157	-15.82	Q157	-11.18	I189	-16
3	I189	-8.71	I189	-11.31	S105	-9.76	H224	-9.79
4	S105	-6.77	H224	-6.81	I189	-8.92	Q157	-9.66
5	H224	-6.73	S105	-5.24	H224	-7.71	D134	-5.18
6	V154	-5.52	D134	-4.13	D134	-5.51	V190	-4
7	T138	-5.08	G138*	-3.61	W104	-3.61	W104	-3.6
8	D134	-4.36	G39	-3.45	T138	-3.37	I285	-2.28
9	G39	-3.65	W104	-2.85	V154	-3.05	A281	-2.24
10	W104	-3.60	V154	-2.62	V190	-2.99	G39	-2.2
11	A141	-2.22	A190*	-2.32	G39	-2.95	L154*	-2.05
12	V190	-1.39	A141	-1.9	I285	-0.90	T138	-1.76

Ranking	Vinyl cinnamate				Vinyl salicylate			
	WT CalB		T138G/V190A		WT CalB		V154L/L278A	
	aa	k <sub>cal</sub> /mol						
13	I285	-1.22	I285	-1.01	Q106	-0.6	Q106	-1.48
14	T158	-1.04	T158	-0.87	A141	-0.41	G41	-0.46
15	G137	-1.02	A281	-0.79	A281	-0.38		
16	A281	-0.93	L278	-0.66				
17	L278	-0.81	Q106	-0.3				
18	T42	-0.54						
	S153	-036						
<b>Total energy of the enzyme-substrate complex</b>	<b>-82.74 k<sub>cal</sub>/mol</b>		<b>-83.27 k<sub>cal</sub>/mol</b>		<b>-67.15 k<sub>cal</sub>/mol</b>		<b>-76.31 k<sub>cal</sub>/mol</b>	



**Figure 3.2** Methyl cinnamate and methyl salicylate production by the CaIB lipase subjected to 3 rounds of Iterative Semi-Rational Design (ISRD). **A**) First-generation (G1) selection of 6 improved mutants towards methyl cinnamate production at positions 134 and 138. **B**) Additive effects of second-generation (G2) mutations at positions D134S, V190A and

**I189T over the G1-selected T138G template. C) Exploration of two evolutionary trajectories from starting templates T138G-V190A and T138G- D134S as part of the third generation (G3). D) Summary of the best variants selected from two generations (G1-G2) of CalB ISRD performed with vinyl salicylate as substrate. Each experiment was performed in triplicate.**

In accordance with our computational energy calculations (Table 1), library design for the methyl cinnamate trajectory was more successful at predicting positive variants, generating fourteen CalB variants with increased synthetic activity relative to WT CalB (Figure 3.2A-C). In contrast, the methyl salicylate trajectory generated three variants with increased activity relative to WT CalB (Figure 3.2D). Interestingly, the most active methyl cinnamate G1 variants were observed for randomized positions T138 and D134 (Figure 3.2A), where six variants showed increased transesterification relative to WT CalB. The most efficient G1 variants are T138G, T138A, T138S, and T138N, displaying up to 4-fold increases in methyl cinnamate production relative to WT CalB (Table S1). The addition of G2-selected mutations (V190A and D134S) to the G1-T138G template allowed for selection of double mutants displaying production yields nearly six-fold that of the WT conversion rate for the same compound (Figure 3.1.1B). The tendency exhibited by the mutants at the position T138 (at least for G1) was the reduction of the side-chain, which provided more room in the cavity. It is worth to mention that the T138G is the smallest substitutions possible. The potential of a short side chain replacing T138 to be compatible with more mutations (or more substrates for future work with this enzyme) led us to choose T138G, the difference with T138A and T138S being considered marginal. In contrast, the evolutionary trajectory of vinyl salicylate yielded mutants with lower synthetic activity enhancements (Figure 3.2D). With only two designed library generations, we selected a double mutant (V154L-L278A) showing 3.6-fold increase in catalytic activity relative to WT CalB. Noteworthy is the fact that WT CalB shows negligible synthetic activity against vinyl salicylate (< 0.5% conversion). These modest enhancements thus remain promising for additional rounds of design, allowing selection of active variants with respectable activities for desired substrates. We found similar improvement patterns with both substrates, whereby two rounds of ISRD with two single mutations increased transesterification against aromatic substrates unrecognized by WT.

For G3 library design in the vinyl cinnamate trajectory, we opted to combine positive mutations selected in G1 and G2 to investigate additive or synergistic behavior on catalytic activity. As a result, two initial G1-G2 templates were used for further active-site exploration: T138G-V190A and T138G-D134S (Figure 3.2C). Contrary to expectations, merging the two most active G2 substitutions (T138G-D134S-V190A) in the context of the favorable G1 single mutation T138G had deleterious effect on the performance of CalB towards methyl cinnamate synthesis (Figure S4). Similarly, amino acid replacements I189V, I189L, and Q157L were the only G3 substitutions that provided increased synthetic activity relative to WT CalB (Figure 3.2C), while any other substitution in the T138G-V190A evolutionary trajectory decreased activity below WT CalB. Because our G3 screening effort was reduced to only three libraries for two templates (T138G-V190A and T138G-D134S), we found that none of the G3 variants outperformed the best G2 candidates.

#### **3.3.4. Effect of mutations on the catalytic pocket of CalB**

Since mutations in the active-site cavity may disrupt CalB conformation and potentially alter stabilizing enzyme-substrate interactions, we used the RosettaBackrub energy function to identify putative conformational changes in CalB [54]. We also performed a Residue Interaction Network (RIN) analysis with the most active variants of each evolutionary trajectory (T138G-V190A and V154L-L278A) for both methyl cinnamate and methyl salicylate. The Linkland algorithm [60, 61] was used to identify modules (i.e. clusters of interacting residues) that are formed and/or altered upon mutation and ligand binding (Figure S5). Our analysis yielded eleven residue modules distributed throughout the CalB structure. Focusing on functional residue networks, we observed that the active-site cavity of CalB is primarily comprised of three modules, one of them integrating residues of the catalytic triad and the other two dominating the rest of the substrate pocket. The most active CalB variant towards vinyl cinnamate (T138G-V190A) showed almost the same number of new interactions than WT (Figure 3.2A), establishing six new interactions but replacing seven WT links (blue and red lines, respectively). On the other hand, the most active vinyl salicylate variant (V154L-L278A) increased the number of residue interactions by four. In both substrate cases, the three-

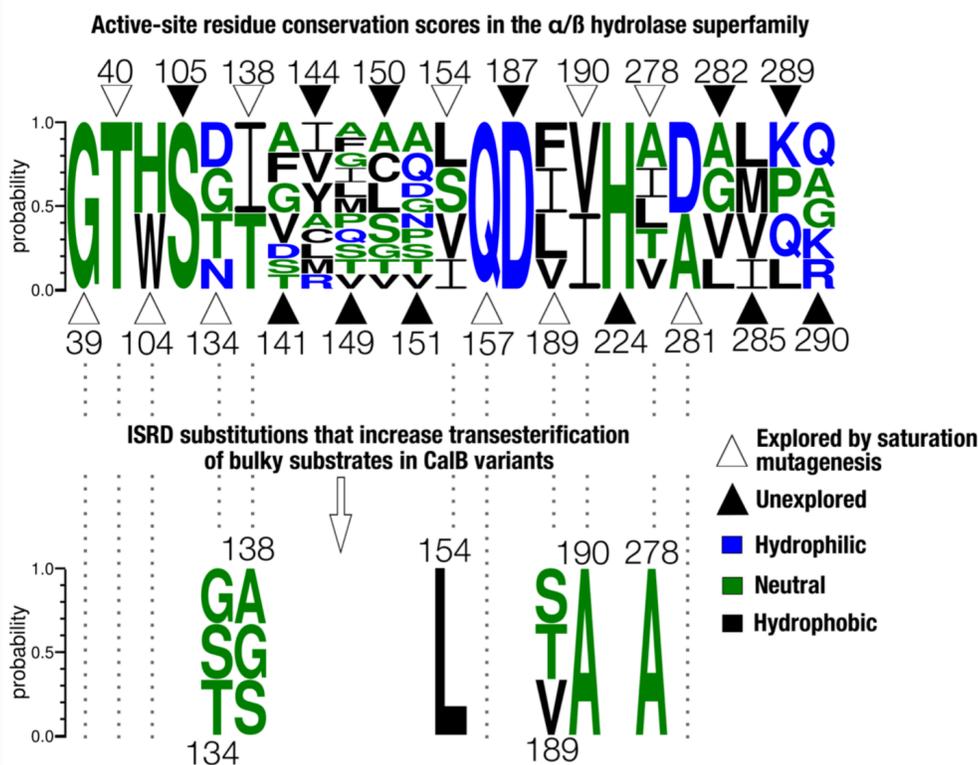
residue modules remain practically unchanged relative to WT. Nevertheless, surface cavity architecture was significantly altered in both variants. Active-site cavity volume was reduced from 531 Å<sup>3</sup> to 385 Å<sup>3</sup> in the V154L-L278A vinyl salicylate variant (Table SI), while that of the T138G-V190A variant against vinyl cinnamate remained similar. In both evolutionary trajectories, enzyme-substrate complexes generated by virtual docking indicated the formation of two hydrogen bonding interactions between H224, D187 and W104 (Figure 3.1.1A). According to simulations, the active-site mutations altered the shape of the cavity surface (Figure 3.1.1B), with different modules stabilizing substrate binding. Both cinnamate and salicylate analogues were stabilized by hydrogen bonding interactions established with residues T40, H224 and S105 (Figure 3.1.2D). This allows proper orientation and positioning of the substrate carbonyl group in range with the catalytic serine for optimal transesterification to occur (Figure 3.1.2C). These results also suggest that the T138G and V190A mutations prevent direct atomic clashes between these residues and the aromatic ring of the methyl cinnamate analog.

Docking simulations of the enzyme-substrate complexes suggest that both analogs form a coordinated pre-intermediate that resembles the theoretical tetrahedral intermediate of the WT reaction expected with bulkier substrates (Figure 3.1.2D). These complexes resolve several aspects of the molecular architecture and substrate orientation observed in the mutated active site of CalB. These include: 1) the molecular proximity between the catalytic serine and the ester group carbonyl of the substrate, 2) complex stabilization by the oxyanion hole (primarily formed by T40), 3) the proper conformation of amino acids forming the catalytic triad, and 4) the accommodation of the aromatic ring in the catalytic pocket, which leaves enough room for the alcohol pocket near W104 to be occupied by methanol during enzyme turnover.

### **3.3.5. Natural vs. directed evolution of the CalB active site**

We randomly explored eleven positions out of the twenty-three residues forming the active-site cavity of CalB (Figure 3.3). Despite the low water content and hydrophobic nature of the transesterification reaction performed under organic conditions, the most favorable residue replacements selected in our screening effort did not translate into strict hydrophobic amino acid replacements. For both substrates, serine and alanine

were the most common replacements selected after three rounds of ISRD. As expected, smaller amino acid replacements were found in the most active variants, a mandatory requirement to accommodate bulkier substrates in the active-site cavity of CalB. Our Evolutionary Trace Analysis (ETA) revealed that either serine at positions 134, 138 and 189 or alanine at position 190 represented two new solutions that were unexplored in the natural diversity of the  $\alpha/\beta$  hydrolase superfamily (Figure 3.3).



**Figure 3.3** Comparative analysis of  $\alpha/\beta$  hydrolase sequence diversity with ISRD-selected active-site mutations that increase CalB transesterification on bulky substrates. Relative frequency conservation scores per position is presented for active-site residues throughout the  $\alpha/\beta$  hydrolase superfamily. Open triangles represent active-site residues subjected to saturation mutagenesis in the present work, while closed triangles represent unexplored active-site residues. Hydrophilic, neutral, and hydrophobic residues are colored blue, green, and black, respectively. Figure prepared with the WebLogo server using CalB sequence numbering, ref. [57].

Consistent with the conservation score obtained for residues 39, 40 and 157, no replacements at these positions favored synthetic activity or tolerated significant sequence diversity. Those findings are also consistent with our binding predictions for

the vinyl cinnamate and vinyl salicylate analogs, where T40 acts as the stabilizing oxyanion hole over the course of the catalytic reaction with both substrates (Figure 3.1.2D). Those results support the fact that T40 is the most important energy contributor during formation of the enzyme-substrate complex (Table 1). However, selective pressure to preserve glycine at position 39 might be related with a smaller side-chain requirement rather than a direct contribution to substrate stabilization in the active-site cavity of CalB ( $\sim 3.0$  kcal/mol, Table 1).

### 3.4. Discussion

To this day, one of the few accessible methods to quantifiably measure esterification reactions of target molecules remains HPLC. While extremely precise and reliable, HPLC is hardly adaptable to High-Throughput Screening (HTS) methodologies, significantly restricting the possibility of conducting directed evolution studies. The lack of a CalB lipase selection system or any reliable HTS method to measure transesterification improvements for desired substrates inspired us to develop an alternative approach to explore the mutational landscape of CalB. This ISRD approach (Figure 3) employs vinyl substrate analogues of the highly sought flavouring agents methyl cinnamate and methyl salicylate, in addition to concentrating mutations around the active-site environment, *i.e.* where they are most likely to be effective [25]. Consequently, ISRD focuses on improving the design of smaller and smarter mutant libraries rather than blindly increasing screening capabilities. As proof of concept, we tested this approach using the CalB lipase against bulky aromatic substrates vinyl cinnamate and vinyl salicylate. On a side note, computational docking methods were performed using the “real” industrial substrates cinnamic acid and salicylic acid, a procedure that validated the mutational targeting of the same active-site positions. These computational results were expected since the “real” industrial substrates and the vinyl substrate analogs are very similar chemical compounds. However, since most of the experimental work presented in this work relates to activity against vinyl analogs, it is clearer to focus attention on the vinyl substrates here. Future work will indeed try to investigate the experimental activity of the evolved lipase variants against cinnamic acid

and salicylic acid for improved flavour synthesis relative to WT CalB. This was justified by a number of experimental requirements cherished by industrial environments. Indeed, WT CalB natively recognizes long aliphatic substrates with low commercial appeal, while exhibiting really poor synthetic activity on bulkier, branched, or aromatic substrates with significant biotechnological interest. Additionally, this iterative screening and design approach allowed us to identify synergistic, antagonistic, compensatory, and deleterious effects between active-site mutations in CalB. As part of a wider program, adding this information into a database alongside other computational tools (i.e. machine learning), it would be possible to more accurately predict the chemical replacement requirements for proteins, e.g. using members of the same family.

Our results illustrate that ISRD is easily applicable to this context, acting as an efficient approach to engineer enzymatic properties. In the present work, exploration of eleven out of the twenty-three positions inside the active-site pocket of CalB was performed, allowing us to uncover single, double, and triple mutants from the combination of individually saturated positions. Even when accounting for our limited screening ability, the ISRD approach proved significantly efficient to increase CalB activity. The total sequence space exploration of our framework would have generated 106,260 variants [39], so testing a total of 360 CalB variants per evolutionary trajectory (six libraries per generation, *i.e.* 120 CalB variants for three generations) represents only the 0.33% of the total putative screening effort needed to explore this sequence space [39]. Screening of an experimentally modest 60-66 bacterial colonies (transformants) per library over three generations yielded twelve mutants exhibiting improved CalB transesterification activity over WT (Figure 3.1.1-3.1.2). ISRD outperformed results obtained in similar CalB engineering studies, where screening effort often exceeded 1,000 explored variants to select a total of ~10-32 improved variants [16, 58, 62]. The closest screening effort comparison with our work is provided by a study carried out by Wu and coworkers [63], whereby the same ISM mutagenesis protocol was performed with a library size of 430 colonies per generation (our approach only used ~360 colonies). Because ISRD combines virtual docking with ISM, our approach also considerably helped reduce total screening effort. This advantage even holds true in

comparison with approaches that solely rely on *in silico* screening methodologies. For instance, Juhl and coworkers used a library size of 2,400 CalB variants to generate eleven improved CalB mutants [64]. Considering the low screening efforts and activity yields obtained after this modest selection, these results represent a significant percentage of enhanced activity variants relative to previous reports and approaches. The iterative and independent detection of beneficial mutations confirmed the feasibility of ISRD to quickly uncover improved enzyme variants. Additionally, this iterative screening and design approach allowed us to identify synergistic, antagonistic, compensatory, and deleterious effects between active-site mutations in CalB.

Both evolutionary trajectories (methyl cinnamate and methyl salicylate) suggest that one of the main challenges in CalB design is shaping cavity size to accommodate bulkier aromatic substrates. Indeed, redesign of the active-site cavity is critical for proper positioning and orientation of the substrates to achieve transition-state stabilization during catalysis. For the cinnamate trajectory, size reduction of specific amino acid side chains (Thr and Val replaced by Gly and Ala, respectively) provided sufficient room for proper positioning and orientation of the substrate carbonyl with the catalytic serine (Figure 3.1.2C-D). This was achieved by a rearrangement of the surface cavity in the vicinity of positions 134, 138 and 190, resulting in a significant increase in energy of the enzyme-substrate complex (Table 1). This observation suggests that synthetic activity improvements rely on proper substrate accommodation in the active-site cavity rather than important increases in enzyme-substrate affinity. Interestingly, vinyl salicylate activity correlates with a decrease in active-site cavity volume (Table SI). This reshaping still provides proper accommodation of the substrate, which translates into increased stabilization of the enzyme-substrate complex energy (Table 1).

## 4. CONCLUSION AND PERSPECTIVES

Selecting the experimental approach for CalB engineering was based on prior knowledge of structure and function, in addition to screening methods available to interrogate libraries to be generated. In nature, the utility of CalB against bulky esters as substrates is constrained by the limited size of the stereospecific pocket. Based on docking simulations and enzyme-substrate complex energy calculations, we successfully produced three rounds of targeted saturation mutagenesis and achieved significant increases in transphosphorylation using substrate analogs for methyl cinnamate and methyl salicylate. As a governing principle of semi-rational design, the choice of positions to mutate remained rational but the choice of amino acid replacements to introduce was random. In comparison with the low activity of the wild type CalB enzyme for methyl salicylate, we showed the power of a structure-based semi-rational approach targeting active-site residues in enzyme engineering. The combination of computational methods with semi-rational design further illustrated the beneficial and complementary nature of these approaches.

One of the things that could benefit from future improvements in ISRD is the docking procedure. In the case of CalB, a more complete predictive model should take into account natural diversity and active-site particularities of the carboxylic ester hydrolase superfamily. As explained before for carboxyl ester synthesis, a tetrahedral intermediate is formed during nucleophilic attack of the catalytic serine, which is considered the rate-limiting step of the reaction [34]. Binding pockets of esterases provide a pre-organized environment to stabilize this intermediate, namely by forming transient hydrogen bonding interactions. Ideally, substrates would need to be covalently docked to the enzyme in a state that mimics the tetrahedral intermediate. While docking of molecules in their ground state allows for ground-state binding predictions to the enzyme, it does not allow to draw direct conclusions on whether the molecule is converted by the enzyme or not. A docking method that aims to model enzymatic catalysis should reflect the molecular role of the enzyme in stabilizing the transition state [34]. A tetrahedral intermediate that is covalently bound to the catalytic serine is closer to the transition state formed during enzyme-catalyzed ester hydrolysis [34]. Since enzyme interactions

with the acid moiety and the alcohol are identical in both states, the tetrahedral intermediate could be considered more appropriate to predict catalytic activity towards a variety of substrates.

Finally, the coordinated investigation of selected library members through experimental techniques, computational studies, and structure determination will be necessary to rationalize the observed changes at the molecular level. Specifically, these additional efforts may help address the role of long-range mutations and their effectiveness on catalysis in relation to their distance from the active site.



## 5. REFERENCES

- [1] Amorim RVS, Ledingham WM, Kennedy JF and Campos-Takaki GM. 2006. Chitosan from *Syncephalastrum racemosum*: using sugar cane substrates as inexpensive carbon sources. *Food Biotechnology* 20(1): 43-53.
- [2] Ting WJ, Tung KY, Giridhar R and Wu WT. 2006. Application of binary immobilized *Candida rugosa* lipase for hydrolysis of soybean oil. *Journal of Molecular Catalysis B: Enzymatic* 42(1):32-38.
- [3] Damasceno FRC, Freire DMG and Cammarota MC. 2014. Assessing a mixture of biosurfactant and enzyme pools in the anaerobic biological treatment of water waste with a high-fat content. *Environmental Technology* 35(16):2035–2045.
- [4] Müller J, Sowa MA, Fredrich B, Brundiek H and Hornscheuer UT. 2015. Enhancing the Acyltransferase Activity of *Candida antarctica* Lipase A by Rational Design. *ChemBioChem* DOI: 10.1002/cbic.201500187.
- [5] Leblanc D, Morin A, Gu D, Zhang XM, Bisailon JG, Paquet M, Dubeau H. 1998. Short chain fatty acid esters synthesis by commercial lipases in low-water systems and by resting microbial cells in aqueous medium. *Biotechnology letters* 20:1127-1131.
- [6] Longo M. and Sanromán A. 2006. Production of Food Aroma Compounds: Microbial and Enzymatic Methodologies. *Food Technology and Biotechnology* 44 (3) 335–353.
- [7] Salah R., Ghamghui H., Miled N. , Mejdoub H. and Gargouri Y. 2007. Production of butyl acetate by lipase from novel of *Rhizopus oryzae* J. *Biosci. Bioeng.*, vol. 104, pp. 268-372.
- [8] Charu G., Dhan P. and Sneh G. 2015. A Biotechnological Approach to Microbial Based Perfumes and Flavours. *Journal of Microbiology & Experimentation* 2 (1): 00034.
- [9] Anthonsen T. 2000. Reactions Catalyzed by Enzymes. In Adlercreutz, Patrick; Straathof, Adrie J. J. *Applied Biocatalysis* (2nd ed.). Taylor & Francis. pp. 18–59. [10] Jaques P, Béchet M, Bigan M, Caly D, Chataigné G, Coutte F, Flahaut C, Heuson E, Leclère V, Lecouturier D, Phalip V, Ravallec R, Dhulster P and Froidevaux R. 2017. High-throughput strategies for the discovery and engineering of enzymes for biocatalysis, *Bioprocess and biosystems engineering* (40) 161-180.

- [11] Kathryn M. and Chi-Huey W. 2001. Enzymes for chemical synthesis. *Nature* 409 (1) 232-240.
- [12] Christie W. 1993. Preparation of ester derivatives of fatty acids for chromatographic analysis. *Advances in lipid methodology* 69-111.
- [13] Dordick JS. 1989. Enzymic catalysis in monophasic organic solvents, *Enzyme Microbial Technology* 11:194–211.
- [14] Paiva AL, Van Rossum D and Malcata FJ. 2002. Kinetics of lipase mediated synthesis of butyl butyrate in n-hexane. *Biocatalysis and Biotransformation* 20:43–51.
- [15] Magnusson A, Rotticci-Mulder J, Santagostino A, and Hult K. 2005. Creating space for large secondary alcohols by rational redesign of *Candida antarctica* Lipase B. *ChemBioChem* (6) 1051-1056.
- [16] Lutz S. 2004. Engineering lipase B from *Candida antartica*. *Tetrahedron: Asymmetry*. 15 (18): 2743-2748.
- [17] Jaeger KE, Dijkstra BW, and Reetz MT. 1993. Bacterial biocatalysts: Molecular biology, three-dimensional structures, and biotechnological applications of lipases. *Annual Review of Microbiology* 53:315–351.
- [18] Borrelli G-M. and Trono D. 2105. Recombinant Lipases and Phospholipases and Their Use as Biocatalysts for Industrial Applications. *International Journal of Molecular Sciences*. 16(9), 20774-20840.
- [19] De Carvalho CCR. 2011. Enzymatic and whole cell catalysis: finding new strategies for old processes. *Biotechnology advances*, 29(1):75–83.
- [20] Reetz M. and Jaeger K. 2000. Enantioselective enzymes for organic Synthesis created by directed evolution. *Chemistry A European journal* 6 (3): 407-412.
- [21] Karra-Chaâbouni M, Ghamgui M, Bezzine S, Rekik A, and Gargouri Y, 2006. Production of flavor esters by immobilized simulans lipase in a solvent-free system, *Process Biochemistry*, 41:1692-1698.
- [22] Krishna SH, Divakar S, Prapulla SG and Karanth NG. 2001. Enzymatic synthesis of isoamyl acetate using immobilized lipase from *Rhizomucor miehei*, *Biotechnology*, 87:193–201.

- [23] Majumder AB, Singh B Dutta D, Sadhukhan S and Gupta M.N. 2006. Lipase catalyzed synthesis of benzyl acetate in solvent-free medium using vinyl acetate as acyl donor, *Bioorganic and Medicinal Chemistry Letters*, 16:4041- 4044.
- [24] Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. 2013. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* 339, 307–310.
- [25] Chica R, Doucet N, and Pelletier J. 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current opinion in biotechnology*. 16:378-384.
- [26] DeSieno M., Du J., and Zhao H., 2009. Altering Enzyme Substrate and Cofactor Specificity via Protein Engineering. *Protein Engineering Handbook Vol. 2*. 777–796.
- [27] Rui L, Cao L, Chen W, Reardon KF, Wood TK. 2004. Active site engineering of the epoxide hydrolase from *Agrobacterium radiobacter* AD1 to enhance aerobic mineralization of cis-1,2-dichloroethylene in cells expressing an evolved toluene ortho-monooxygenase. *Journal of Biological Chemistry*, 279:46810-46817.
- [28] Hill CM, Li WS, Thoden JB, Holden HM, Raushel FM: Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active site. *Journal of the American Chemical Society*, 125:8990-8991.
- [29] Wilming M, Iffland A, Tafelmeyer P, Arrivoli C, Saudan C, Johnsson K. 2002. Examining reactivity and specificity of cytochrome c peroxidase by using combinatorial mutagenesis. *ChemBioChem*, 3:1097-1104.
- [30] Antikainen NM, Hergenrother PJ, Harris MM, Corbett W, Martin SF. 2003. Altering substrate specificity of phosphatidylcholine-preferring phospholipase C of *Bacillus cereus* by random mutagenesis of the headgroup binding site. *Biochemistry*, 42:1603-1610.
- [31] Yew WS, Akana J, Wise EL, Rayment I, Gerlt JA. 2005. Evolution of enzymatic activities in the orotidine 5(-monophosphate decarboxylase suprafamily: enhancing the promiscuous D-arabino-hex-3-ulose 6-phosphate synthase reaction catalyzed by 3-keto- L-gulonate 6-phosphate decarboxylase. *Biochemistry*. 44:1807-1815.

- [32] Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG. 2004. An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Science*, 101:7566-7571.
- [33] Reetz M.T. and Carballeira J.D. 2007. *Nature protocols*. 2(4):891-903
- [34] Juhl P, Trodler P, Tyagi S, Pless J. 2009. Modelling substrate specificity and enantioselectivity for lipases and esterases by substrate-imprinted docking. *BMC Structural Biology*, 9:39.
- [35] Rarey M, Kramer B, Lengauer T, Klebe G. 1996. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261:470-489.
- [36] Thomsen R and Christensen MH. 2006. Moldock: a new technique for high-accuracy Molecular Docking. *Journal of Medicinal Chemistry* 49: 3315-332.
- [37] CLC bio. 2013. Molegro Virtual Docker (MVD). Denmark.
- [38] Kille S, Acevedo-Rocha C, Parra L, Zhang Z, Opperman D, Reetz M, Acevedo J. 2012. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synthetic biology*, (2) 83-92.
- [39] Patrick WM, Firth AE and Blackburn JM. 2003. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Engineering* 16(6) 451-457.
- [40] Macrae A, Roehl .L, and Brand, H.M. 1990. Bio-esters in Cosmetics, *Journal of Cosmetic Science*. 147, 36–39.
- [41] Wang YF, Lalonde JJ, Momongan M, Bergbreiter DE and Wong CH. 1988. Lipasecatalyzed irreversible transesterifications using enol esters as acylating reagents: preparative enantio- and regioselective syntheses of alcohols, glycerol derivatives, sugars, and organometallics, *Journal of the American Chemical Society* 110:7200–7205
- [42] Packer M and Liu D. 2015. Methods for the Directed Evolution of Proteins. *Nature reviews genetics*. 7: 379–94.
- [43] Zheng J, Fu X, Ying X, Zhang Y and Wang Z. 2014. A sensitive colorimetric high-throughput screening method for lipase synthetic activity assay. *Analytical Biochemistry* 452:13–15.
- [44] Zurek G, Karst U. 1997. Microplate photometric determination of aldehydes in disinfectant solutions, *Analytica Chimica Acta* 351: 247–257.

- [45] Larsen MW, Bornscheuer UT and Hult K. 2008. Expression of *Candida antarctica* lipase B in *Pichia pastoris* and various *Escherichia coli* systems. *Protein Expression and Purification* 62: 90-97.
- [46] Wang F, Hou S, Wang Q, Wang P, Liu J, Yang B & Wang Y. 2015. Impact of Leucine 278 residue on fatty acid length specificity of *Candida antarctica* Lipase B. *Advances in Microbiology* 5: 493-499.
- [47] Marton Z, Leonard V, Syren PO, Bauer C, Lamare S, Hult K, Tran V. and Graber M. 2010. Mutations in the stereospecificity pocket and at the entrance of the active site of *Candida antarctica* lipase B enhancing enzyme enantioselectivity. *Journal of Molecular Catalysis B: Enzymatic, Elsevier* 65:11-17.
- [48] Sambrook J, Fritschi EF and Maniatis T. 1989. *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, New York.
- [49] Sandoval G and Marty A. 2007. Screening methods for synthetic activity of lipases. *Enzyme and Microbial Technology* 40:390-393.
- [50] Ru MT, Dordick JS, Reimer JA and Clark DS. 1999. Optimizing the salt-Induced Activation of Enzymes in Organic Solvents: Effects of Lyophilization time and water content. *Biotechnology and Bioengineering* 63(2):233-241.
- [51] Uppenberg J, Hansen MT, Patkar S and Jones A. 1994. The sequence, crystal structure determination and refinement of two crystal forms of the lipase B from *Candida antarctica*. *Current Biology* 2:293-308.
- [52] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, and Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6(5): 343-347.
- [53] Lauck F, Smith CA, Friedland GF, Humphris EF and Kortemme T. 2010. RosettaBackrub-a web server for flexible backbone protein structure modeling and design. *Nucleic Acids Research* 38(2):W569-W575.
- [54] Smith C.A. and T. Kortemme. 2011. Predicting the Tolerated Sequences for Proteins and Protein Interfaces Using Rosetta Backrub Flexible Backbone Desing. *PLoS ONE* 6(7):e20451. Doi:10.1371/journal.pone.0020451.

- [55] Dundas J, Ouyang Z, Tseng J, Bonkowiski A, Turpaz Y & Liang J. 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research* 34: 116-118.
- [56] Armon A, Graur D and Ben-Tal N. 2001. ConSurf: An algorithmic tool for the identification of Functional Regions in Proteins by Surface Mapping of Phylogenetic Information. *Journal of Molecular Biology* 307:447-463.
- [57] Crooks GE, Hon G, Chandonia JM and Brenner SE. 2004. WebLogo: A Sequence Logo Generator. *Genome Research* 14: 1188-1190.
- [58] Larsen MW, Zielinska DF, Martinelle M, Hidalgo A, Jensen LJ, Bornscheuer UT and Hult K. 2010. Suppression of water as a nucleophile in *Candida antarctica* Lipase B catalysis. *CHEMBIOCHEM* 11: 796-801.
- [59] Doncheva NT, Assenov Y, Domingues FS and Albrecht M. 2012. Topological analysis and interactive visualization of biological networks and protein structures. *Nature protocols* 7 (4): 670-685.
- [60] Kovács IA, Palotai R, Szalay MS and Csermely P. 2010. Community landscapes: An integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. *PlosOne* 5(9): e12528.
- [61] Szalay-Beko M and Palotai R. 2012. Modulan plug-in for cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics* 28: 2202-2204.
- [62] Park CG, Kwon MA, Song JK and Kim DM. 2011. Cell-free synthesis and multifold screening of *Candida antarctica* lipase B (CalB) variants after combinatorial mutagenesis of hot spots. *Biotechnology Progress* 27: 47-53.
- [63] Wu Q, Soni P and Reetz MT. 2012. Laboratory evolution of enantiocomplementary *Candida antarctica* lipase B mutants with broad substrate scope. *Journal of the American Chemical Society* 135: 1872-1881.
- [64] Juhl PB, Doderet K, Hollmann F, Thum O and Pleiss J. 2010. Engineering of *Candida antarctica* lipase B for hydrolysis of bulky carboxylic acid esters. *Journal of Biotechnology* 150: 474-480.

# ANNEXE 1. SUPPORTING FIGURES

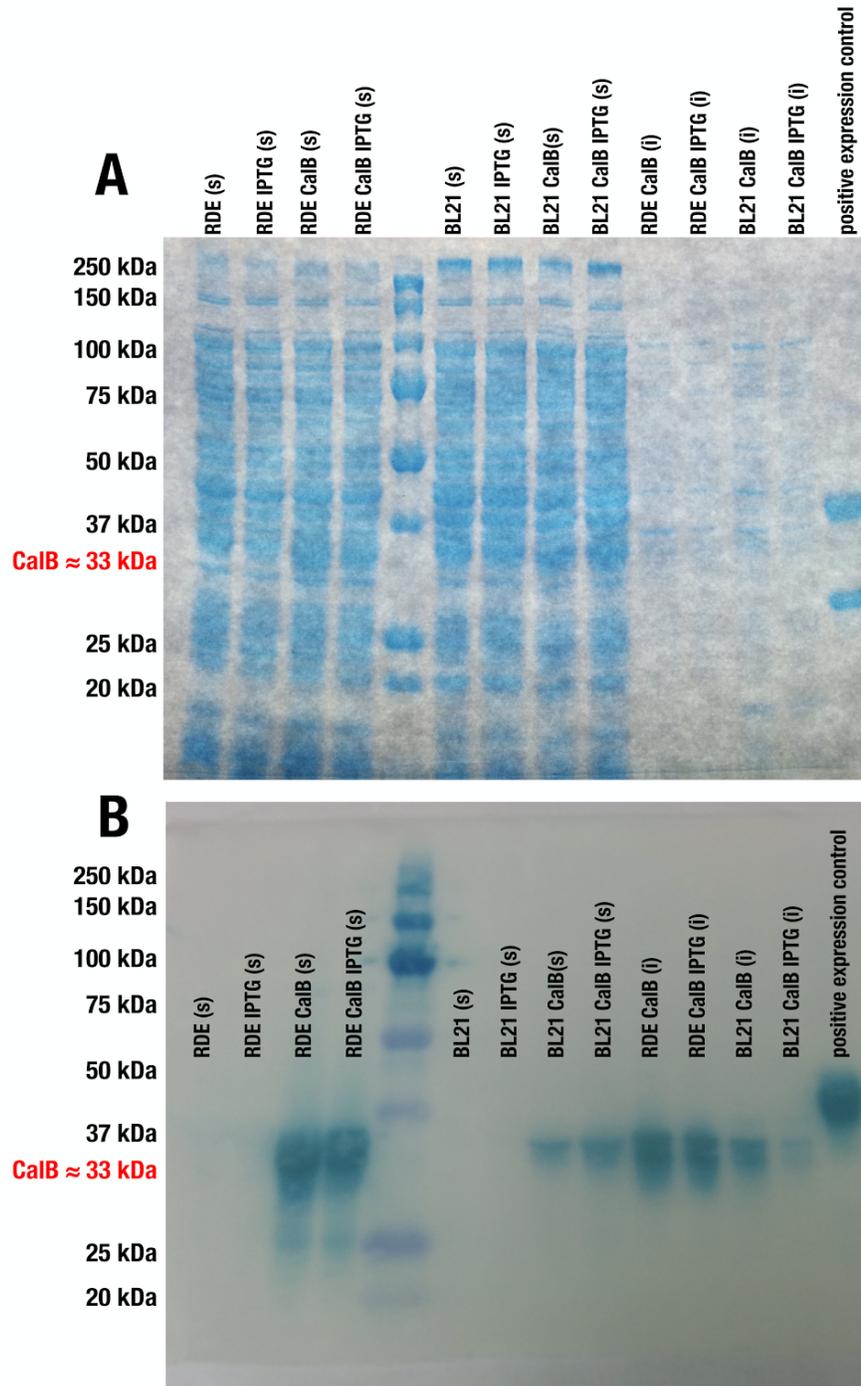
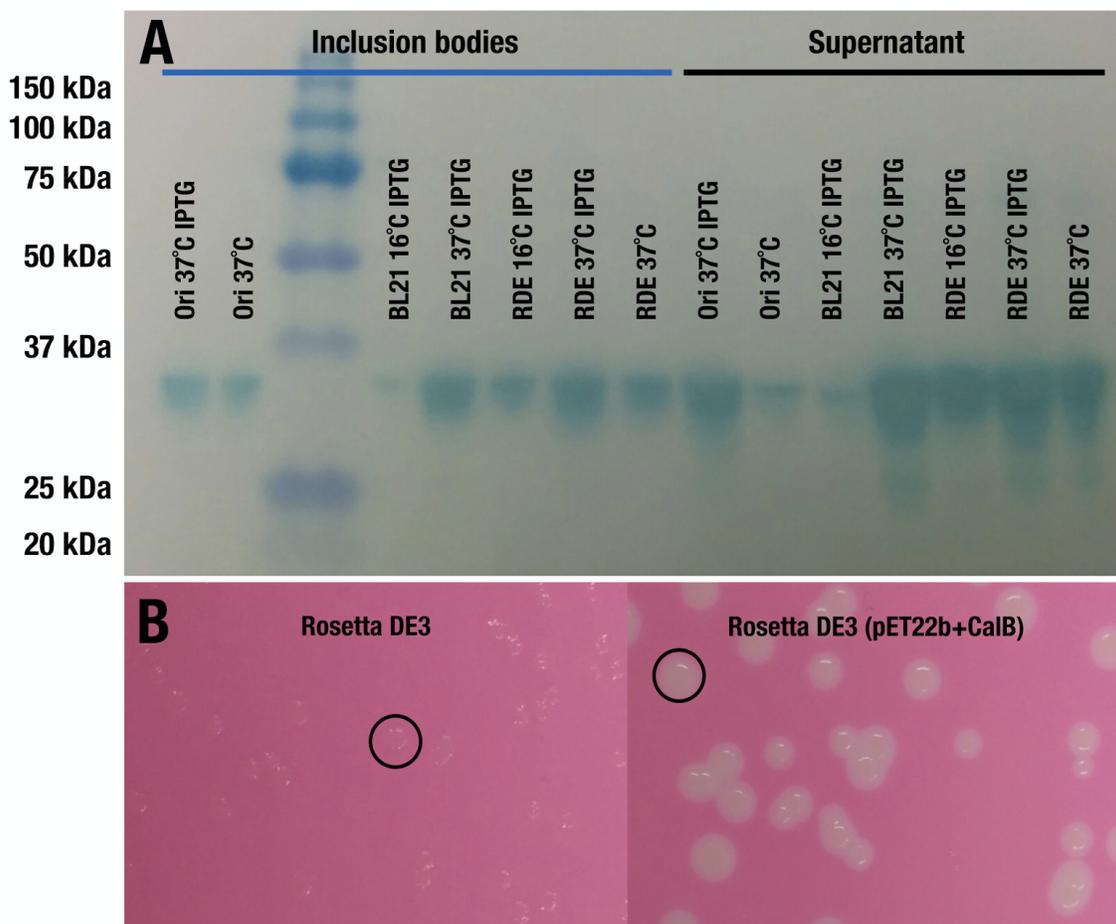
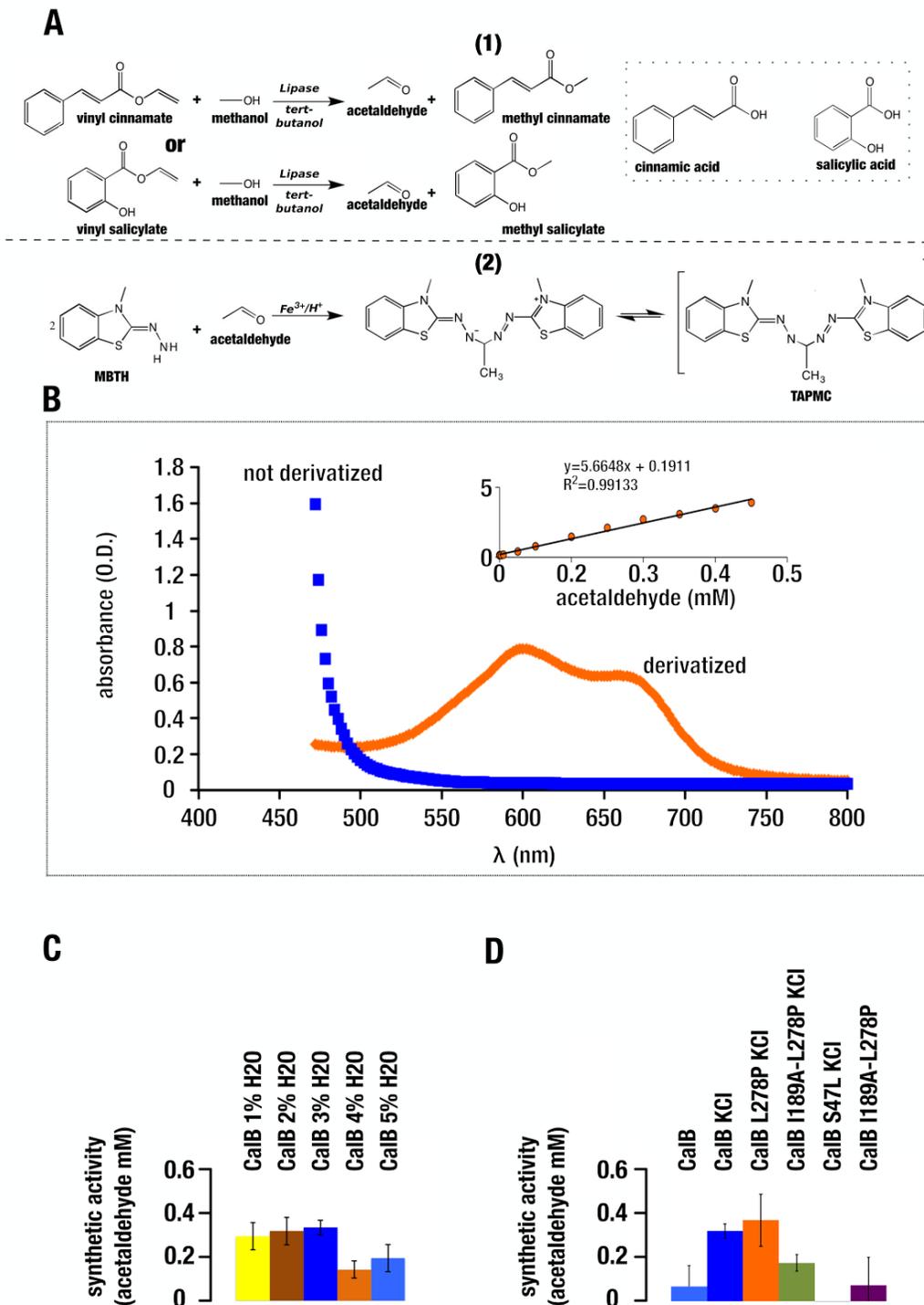


Figure S1. Wild-type CalB overexpression in *E. coli*. A) Comparative expression yields between *E. coli* Rosetta DE3 (RDE) and BL21 DE3 (BL21). Coomassie-stained SDS-PAGE illustrates two phenotypic expression backgrounds with and without IPTG induction for soluble (s) and insoluble (i) forms of CalB, which are respectively observed in the supernatant or expressed as inclusion bodies in cell pellets. B)

Western-Blot analysis showing that the WT pET22b(+)-CalB cell system is leaky, displaying significant protein expression even in the absence of IPTG induction. See Materials & Methods for details.

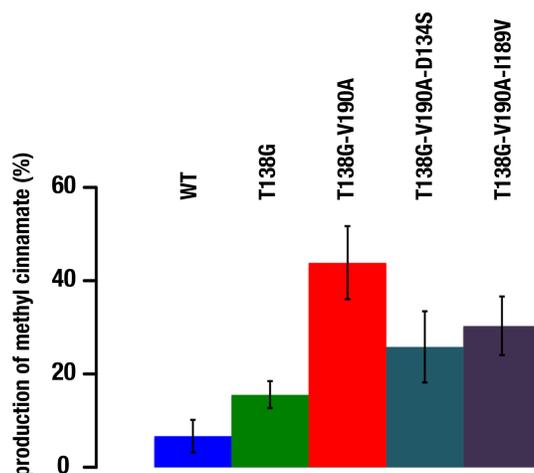


**Figure S2.** Combined effects of *E. coli* strain, IPTG induction, and incubation temperature on WT CalB production yields. **A)** Western-Blot analysis illustrates that optimal WT CalB expression is observed in the soluble extract of a Rosetta DE3 strain incubated at 16°C. Under these conditions, CalB is not exhibiting degradation. Expression levels for soluble and insoluble extracts are shown for *E. coli* strains Origami DE3 (Ori), BL21 DE3 (BL21), and Rosetta DE3 (RDE) incubated at 16°C and 37°C with and without IPTG induction. **B)** Lipase esterification assay performed on solid-state agar medium in the presence of oleic acid and decanol as substrates, in addition to rhodamine B as activity reporter [49]. The assay shows WT CalB expression in a Rosetta DE3 background relative to an empty strain control. The CalB lipase is fully functional, as revealed by the appearance of a colony halo upon enzyme overexpression, further providing a suitable agar plate screening method for enzyme activity (see Materials & Methods for details).

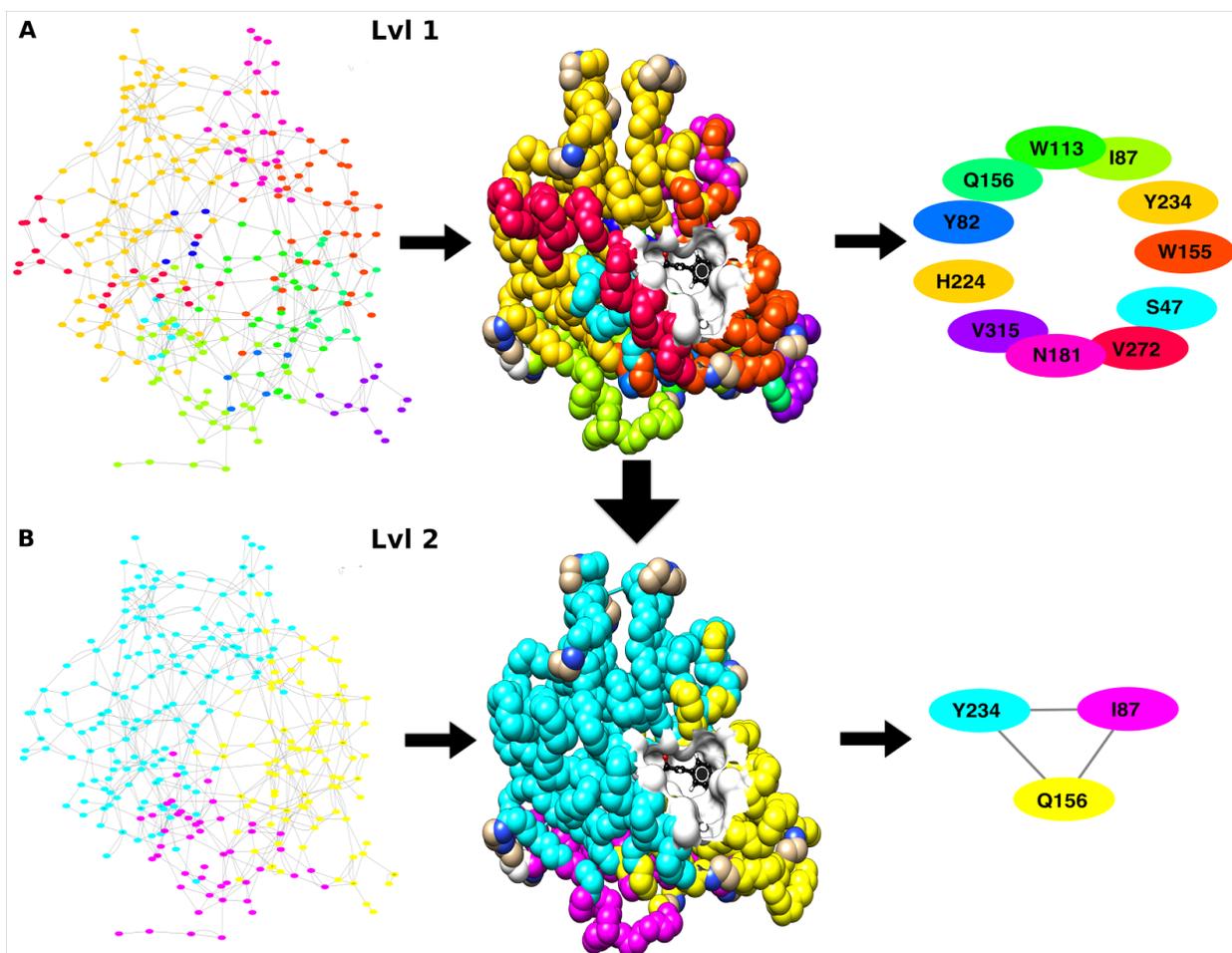


**Figure S3.** Optimization of the CalB library screening methodology for the detection of methyl cinnamate and vinyl salicylate. Standardization, buffer conditions, starting template selection, and water percentage in the reaction mixture were evaluated for proper mutant selection in the enzyme engineering approach. A) Schematic representation of the reporter reaction used for screening lipase-catalyzed transesterification. The transesterification reaction is performed using the vinyl cinnamate and vinyl salicylate substrate analogs (step 1). For each enzymatic turnover with co-substrate methanol, an acetaldehyde and flavoured ester

molecule (methyl cinnamate or methyl salicylate) are released in a 1:1 stoichiometric ratio. The synthetic activity is measured by a colorimetric method where 3-methyl-2-benzothialinone (MBTH) reacts with acetaldehyde to release an aldazine moiety that is converted to blue-colored tetraaza pentamethincyanine (TAPMC) (step 2) [43, 44]. B) UV-Visible spectral curves of the colorimetric method employed to measure the synthetic activity of CalB in *tert*-butanol. A vinyl cinnamate or vinyl salicylate analog reacts with methanol in a transesterification reaction that releases an acetaldehyde that is further derivatized for quantifiable detection at a wavelength of 598 nm. C) Water content optimization in the reaction mixture. A content of 3% H<sub>2</sub>O (v/v) was found optimal to measure the synthetic activity of WT CalB expressed from pET22b(+) in *E. coli* Rosetta (DE3). D) Prospective CalB template activity comparison under salt-induced activation (KCl) to promote lipase activity in *tert*-butanol (see Materials & Methods for details). CalB variants S47L, L278P, and I189A-L278P were used as previously reported mutations showing increased synthetic activity against bulky substrates [46, 47, 58]. In our experimental conditions, no significant activity improvement was observed relative to KCl-treated WT CalB.



**Figure S4.** Methyl cinnamate production by CalB variants obtained from G1, G2 and G3 library screening. Experimental conditions: 10 mg of cell extracts, 300 mM vinyl cinnamate substrate (1:3 analog: methanol ratio), 2 h incubation at 30°C and 250 rpm.



**Figure S5.** Hierarchical modular organization of CalB using a Residue Interacting Network (RIN) analysis [61]. A) The first hierarchical organization level of CalB exhibits eleven modules, dividing the catalytic pocket in three main components (orange, yellow and red residue clusters). B) The second hierarchical organization level clusters the previous 11 component modules of level 1 into three new meta-modules. The RIN analysis of CalB is displayed on the left side of each panel, where each circle (or node) represents a single amino acid residue. Each interaction between two nodes (hydrogen bonding, electrostatic,  $\pi$ - $\pi$  stacking, and/or van der Waals) is illustrated by a gray connecting line. Structural mapping of the RIN analysis on the structure of CalB (PDB 1TCA) is illustrated in the center, with the most interconnected residue per module shown on the right. The color palette is conserved between sections.

<b>Evolutionary trajectory</b>	<b>Identity</b>	<b>Activity fold increase relative to WT CalB</b>
<b>Methyl cinnamate production</b>	<b>D134S</b>	<b>2.39</b>
	<b>D134T</b>	<b>2.36</b>
	<b>D138N</b>	<b>3.05</b>
	<b>T138A</b>	<b>3.52</b>
	<b>T138S</b>	<b>4</b>
	<b>T138G</b>	<b>3.17</b>
	<b>T138G-D134S</b>	<b>5.21</b>
	<b>T138G-V190A</b>	<b>5.56</b>
	<b>T138G-I189T</b>	<b>2.35</b>
<b>Methyl salicylate production</b>	<b>V154L</b>	<b>2.28</b>
	<b>V154L-L278A</b>	<b>3.65</b>
	<b>V154L-I189S</b>	<b>2.2</b>
	<b>V154LI189T</b>	<b>2.47</b>

**Table S1. Synthetic activity increase of CalB variants.**

Model	Volume ( $\text{\AA}^3$ )	Area ( $\text{\AA}^2$ )	Cavity mouth area ( $\text{\AA}^2$ )
WT CalB	531	418.6	74
V154L	276.5	300.4	57.2
V154L-L278A	384	335	102.7
T138G	580	450	74.6
T138G-V190A	519.5	423.8	97.2

**Table S2.** Changes in CalB active-site volume upon mutagenesis and selection. Note: Volumes were calculated with the CASTp server. See ref. [55].