Université de Montréal

Towards higher predictability in enzyme engineering

Investigation of protein epistasis in dynamic ß-lactamases and Cal-A lipase

Par

Lorea Alejaldre Ripalda

Département de biochimie et médecine moléculaire

Faculté de Médecine

Thèse présentée en vue de l'obtention du grade de *Philosophae Doctor*

en Biochimie

Décembre 2020

Université de Montréal

Unité académique: Département de Biochimie et Médecine Moléculaire, Faculté de Médecine

---

Cette thèse intitulée

Towards higher predictability in enzyme engineering

Investigation of protein epistasis in dynamic ß-lactamases and Cal-A lipase

Présenté par

Lorea Alejaldre Ripalda

A été évalué(e) par un jury composé des personnes suivantes

Adrian Serohijos
Président-rapporteur

Joelle Pelletier
Directeur de recherche

Yves Brun
Membre du jury

Shelley Copley
Examinateur externe

Résumé

L'ingénierie enzymatique est un outil très avantageux dans l'industrie biotechnologique. Elle permet d'adapter les enzymes à une activité ou à une condition de réaction spécifique. En outre, elle peut permettre de déchiffrer les éléments clés qui ont facilité leur modification. Bien que l'ingénierie enzymatique soit largement pratiquée, elle comporte encore plusieurs goulets d'étranglement. Certains de ces goulets d'étranglement sont techniques, comme le développement de méthodologies pour la création de banques de mutations ciblées ou la réalisation de criblages à haut débit, et d'autres sont conceptuels, comme le déchiffrage des caractéristiques clés pertinentes d'une protéine cible pour la réussite d'un projet d'ingénierie. Parmi ces défis, l'épistasie intra-génique, ou la non-additivité des effets phénotypiques des mutations, est une caractéristique qui entrave grandement la prévisibilité. L'amélioration de l'ingénierie enzymatique nécessite une approche multidisciplinaire qui inclut une meilleure compréhension des relations structure-fonction-évolution.

Cette thèse vise à contribuer à l'avancement de l'ingénierie enzymatique en étudiant deux systèmes modèles. Premièrement, des variantes dynamiques de la ß-lactamase TEM-1 ont été choisies pour étudier le lien entre la dynamique des protéines et l'évolution. La ß-lactamase TEM-1 a été largement caractérisée dans la littérature, ce qui s'est traduit par des connaissances approfondies sur son mécanisme de réaction, ses caractéristiques structurelles et son évolution. Les variantes de la ß-lactamase TEM-1 utilisées comme système modèle dans cette thèse ont été largement caractérisées, montrant une dynamique accrue à l'échelle temporelle pertinente pour la catalyse (µs à ms) mais maintenant la reconnaissance du substrat. Dans cette thèse, l'évolution *in vitro* de ces variantes dynamiques a été réalisée par des cycles itératifs de mutagenèse et de sélection aléatoires pour permettre une exploration impartiale du paysage de 'fitness'. Nous démontrons que la présence de ces mouvements particuliers au début de l'évolution a permis d'accéder à des voies de mutations connues. De plus, des interactions épistatiques connues ont été introduites dans les variantes dynamiques. Leur caractérisation *in silico* et cinétique a révélé que les mouvements supplémentaires sur l'échelle de temps de la catalyse ont permis d'accéder

à des conformations conduisant à une fonction améliorée, comme dans le TEM-1 natif. Dans l'ensemble, nous démontrons que l'évolution de la b-lactamase TEM-1 vers une nouvelle fonction est compatible avec divers mouvements à l'échelle de temps µs à ms. Il reste à savoir si cela peut se traduire par d'autres enzymes ayant un potentiel biotechnologique.

Deuxièmement, la lipase Cal-A, pertinente sur le plan industriel, a été choisie pour identifier les caractéristiques qui pourraient faciliter son ingénierie. La lipase Cal-A présente des caractéristiques telles que la polyvalence du substrat et une grande stabilité thermique et réactivité qui la rendent attrayante pour la modification des triglycérides ou la synthèse de molécules pertinentes dans les industries alimentaire et pharmaceutique. Contrairement à TEM-1, la plupart des études d'évolution *in vitro* de la lipase Cal-A ont été réalisées dans un but industriel, avec une exploration limitée de l'espace de mutation. Par conséquent, les caractéristiques qui définissent la fonction de la lipase Cal-A restent insaisissables. Dans cette thèse, nous faisons état de la mutagenèse ciblée de la lipase Cal-A, confirmant l'existence d'une région clé pour la reconnaissance du substrat. Cela a été fait en combinant une nouvelle méthodologie de création de bibliothèque basée sur l'assemblage Golden-gate avec une visualisation structurelle basée sur des scripts pour identifier et cartographier les mutations sélectionnées dans la structure 3D. La caractérisation et la déconvolution de deux des plus aptes ont révélé l'existence d'une épistasie dans l'évolution de la lipase Cal-A vers une nouvelle fonction. Dans l'ensemble, nous démontrons que l'identification d'une variété de propriétés suite à la mutagenèse ciblée peut grandement améliorer la connaissance d'une enzyme. Cette information peut être appliquée pour améliorer l'efficacité de l'ingénierie dirigée.

Mots-clés : ingénierie enzymatique, évolution des protéines, épistasie intragénique, dynamique des protéines, TEM-1 ß-lactamase, Cal-A lipase, test d'activité *in vitro*, cinétique enzymatique, test d'activité *in vivo*, arrimage moléculaire flexible

Abstract

Enzyme engineering is a tool with great utility in the biotechnological industry. It allows to tailor enzymes to a specific activity or reaction condition. In addition, it can allow to decipher key elements that facilitated their modification. While enzyme engineering is extensively practised, it still entails several bottlenecks. Some of these bottlenecks are technical such as the development of methodologies for creating targeted mutational libraries or performing high-throughput screening and some are conceptual such as deciphering the key relevant features in a target protein for a successful engineering project. Among these challenges, intragenic epistasis, or the non-additivity of the phenotypic effects of mutations, is a feature that greatly hinders predictability. Improving enzyme engineering needs a multidisciplinary approach that includes gaining a better understanding of structure-function-evolution relations.

This thesis seeks to contribute in the advancement of enzyme engineering by investigating two model systems. First, dynamic variants of TEM-1 ß-lactamase were chosen to investigate the link between protein dynamics and evolution. TEM-1 ß-lactamase has been extensively characterized in the literature, which has translated into extensive knowledge on its reaction mechanism, structural features and evolution. The variants of TEM-1 ß-lactamase used as model system in this thesis had been extensively characterized, showing increased dynamics at the timescale relevant to catalysis (µs to ms) but maintaining substrate recognition. In this thesis, *in vitro* evolution of these dynamic variants was done by iterative rounds of random mutagenesis and selection to allow an unbiased exploration of the fitness landscape. We demonstrate that the presence of these particular motions at the outset of evolution allowed access to known mutational pathways. In addition, known epistatic interactions were introduced in the dynamic variants. Their *in silico* and kinetic characterization revealed that the additional motions on the timescale of catalysis allowed access to conformations leading to enhanced function, as in native TEM-1. Overall, we demonstrate that the evolution of TEM-1 b-lactamase toward new function is compatible with diverse motions at the µs to ms timescale. Whether this can be translated to other enzymes with biotechnological potential remains to be explored.

Secondly, the industrially relevant Cal-A lipase was chosen to identify features that could facilitate its engineering. Cal-A lipase presents characteristics such as substrate versatility and high thermal stability and reactivity that make it attractive for modification of triglycerides or synthesis of relevant molecules in the food and pharmaceutical industries. Contrary to TEM-1, most *in vitro* evolution studies of Cal-A lipase have been done towards an industrially-specified goal, with limited exploration of mutational space. As a result, features that define function in Cal-A lipase remain elusive. In this thesis, we report on focused mutagenesis of Cal-A lipase, confirming the existence of a key region for substrate recognition. This was done by combining a novel library creation methodology based on Golden-gate assembly with script-based structural visualization to identify and map the selected mutations into the 3D structure. The characterization and deconvolution of two of the fittest revealed the existence of epistasis in the evolution of Cal-A lipase towards new function. Overall, we demonstrate that mapping a variety of properties following mutagenesis targeted to specific regions can greatly improve knowledge of an enzyme that can be applied to improve the efficiency of directed engineering.

Keywords: enzyme engineering, protein evolution, intragenic epistasis, protein dynamics, TEM-1 ß-lactamase, Cal-A lipase, *in vitro* activity assays, enzyme kinetics*, in vivo* activity assays, flexible protein docking

Index

Table list

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Annex 1

Annex 2

Annex 3

Figure list

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Annex 2

Annex 3

Liste des sigles et abréviations

CTX: Cefotaxime

E. coli: *Escherichia coli*

IPTG: Isopropyl 1-thio-ß-D-galactopyranoside

Kan: Kanamycin

k_{cat}: Turnover number

k_{cat}/K_M: Catalytic efficiency

K_M: Michaelis-Menten constant

MD: Molecular Dynamics

MIC: Minimal Inhibitory Concentration

NMR: Nuclear Magnetic Resonance

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PEG: Polyethylene glycol

PELE: Protein Energy Landscape Exploration

PUFA: Polyunsaturated fatty acids

P. *pastoris*: *Pichia pastoris*

RMSD: Root Mean Square Deviation

SDS-PAGE: Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate

S. cerevisiae: *Saccharomyces cerevisiase*

WT: Wild-type

To my parents Carlos and María Jesus

Acknowledgements

My main goal in this thesis was to evolve enzymes but at the end, it's all these years of studying enzyme evolution that have helped me evolve scientifically and personally. Like in the directed evolution of enzymes, there are several features that have facilitated my personal evolution throughout the PhD journey. And unlike enzymes that never thanked me for evolving them (also never asked for it), I would like to take a moment to show my gratitude to the people that have helped during the completion of this thesis either directly or indirectly.

An essential component of directed evolution is to have a laboratory setting to do the appropriate experiments. Therefore, I would like to thank first Dr. Joelle Pelletier for accepting me in her research group and supporting me during this thesis. Joelle easily welcomed me in her lab after some email exchanges and Skype interviews, believing in my potential and always encouraging me to pursue my interests. Under her supervision I have been able to explore different research lines, collaborate, attend international conferences, guide undergraduate students and obtain several scholarships and presentation prizes. But her support has expanded the lab, always with her door open, she has also helped me improve my teaching skills and set up scientific outreach activities. I also appreciate her patience while I improved my French, doing as many corrections to the texts as necessary. Thanks for everything Joelle, your contributions have been so broad that I am sure I will continue to appreciate them more with time!

The environment in the lab is another important factor for the scientific development. In Joelle's lab we have held countless round table meetings and journal clubs. But we have also held beer and chip gatherings and ate lots of cake. It is the place where I mastered speaking *frenchglish* and first learned how to survive Quebec winters by eating poutine and learning to recognize the signs of frostbite. Throughout the years we have formed a diverse community that has evolved as people have come and go, full of interesting discussions, lessons and fun memories. All of you have contributed in some aspect of my journey! (Since we have been so many, I will put you In alphabetical order) Thanks Sarah Melissa A., Jonathan B., Christophe B., Stella C., Dr. David C., Tiffany C., Antoine D., Sarah Slim D., Abdelhadi D., Simon D., Max E., Zineb E., Ali F., Charlotte-Skye F., Sophie G., Joaquin G., Adem H., Katia H., Saathanan I., Kiana L., Claudèle L., Sara O., Benajmin O., Dr. Daniela Q., Natalie R., Léa R., Olivier R., Jacynthe T., and Dr. Donya V.

Being a large group, there is always people that you work more closely with that I would like to highlight. First, Dr. Daniela Quaglia who came to Joelle's lab as a postdoctoral fellow at the same time I started my PhD and adapted with me to life in Québec. I have learned so much from you and we have formed a great team in the lab that has translated to a great friendship. You have been a source of support inside and outside of the lab and I will always be grateful for you.

Muchas gracias por todo Daniela, eres una amiga muy 'cara'! Sophie Gobeil, my predecessor in the ß-lactamase project and Jacynthe Toulouse who had started her PhD just a year before me. You both have helped me to navigate the biochemistry department, with Québec winters and proposed fun activities like Zumba. Special thanks to my interns who have trusted me to guide them and have taught me as much as I have taught them, it has been a pleasure to work with you. Antoine Daigneault-Lemers, my first intern who despite being interested in bioinformatics was keen to do experimental work; Christope Lastrade-Brais whose motivation I admire in trying to develop a new assay; Claudèle Lemay-St-Denis an example of motivated hard worker, always looking to improve your research; and Adem H-Parisien, my last intern, an example of perseverance. I would also like to mention Lukas Deweid, Ingrid Pulido and Ferran Sancho that came for an internship as PhD students, and although they were not my interns, I enjoyed very much introducing you to the lab and life in Québec. Special mention to Lukas and Ingrid that apart from sharing protocols and lab tricks, you also dared to come to hot yoga and explore Montréal with me. I hope that we continue with our intercontinental friendships!

To continue with the laboratory setting, I would like to thank Dr. Manuel Ferrer in the Instituto de Catalisis y Petroleoquímica in Madrid, Spain where I did a short stay to complete part of this thesis. I am very grateful for your guidance and how you welcomed me in your lab. I learned a lot under your supervision and I keep very fond memories of my stay in your lab. *Gracias por confiar en mí!* The environment in Dr. Manuel Ferrer's lab during my stay was also very welcoming and stimulating. Thanks for everything you showed me in the lab and the coffee breaks with *tortilla de patatas*, it was a great experience working with you: Sergio (Bartolo), Sandra, Laura, David and Cristina. Special thanks to Laura and David, you were always keen to offer a helping hand. And to Cristina you have not only helped me with what I thought were impossible protein purifications and have transferred as much of your lab and computational knowledge to me as possible, but you have been a continuous source of support (and laughs) ever since we met (*Siempre con la Phe*). I don't think I can summarize all of your contributions inside and outside the lab. *Muchas gracias por tu amistad inter-oceánica!* Being a small research center, I also exchanged conversations with other people that made my stay even more welcoming: Thanks Patricia (Patch), Fadia, David (el niño) and Noa (*tercer miembro de la triada catalítica y postdocker original*).

Other important structural components have been the staff of the chemistry and biochemistry departments at Université de Montréal. Thanks to the "guys from the atelier" that although we often go see you in a rush and may not feel our appreciation, you are authentic lab saviors and I am very grateful for all the time you have saved my experimental work. In this category, I would like to include the administrative staff from the departments, making sure we stay on track with all of our paperwork and responding to our thousand questions. Thanks for all your patience and work! The department of biochemistry overall has been a great community to

belong, with official activities for scientific exchange like la “fête du papiers” and unofficial like our running club. Special thanks to Lionel, for all the activities you’ve organized and Sam for all the times you’ve led us to the Mont Royal for runs longer than what we expected. I would also like to give a special mention to Marie Pageau, head of the bioinformatics labs, for your continuous support and encouragement in teaching bioinformatics. *Merci toujours par votre confiance en moi (et les chocolats)!*

I would also like to thank my thesis committee Dr. Nicolas Doucet and Dr. Jurgen Sygusch for our yearly discussions on my project. And my predoctoral committee Dr. Adrian Serohijos, Dr. Serguei Chteinberg and Dr. David Kwan for interesting discussions when my project was just starting. Within the biochemistry department, I would like to thank Dr. Pouria Dasmeh for our discussions on protein epistasis.

During my studies I have had the chance to participate in multiple collaborations from which I have learned a lot. I would like to first thank Dr. Victor Guallar and his PhD student Carles Perez at the Barcelona Supercomputing Center for their computational work in my first, first author publication. Pierre-Alexandre Paquet-Cote and Normand Voyer at Université Laval, it was a pleasure working with you to test your ß-lactamase inhibitors. I am also thankful to what I have learned from testing R67 DHFR inhibitors selected by Dr. David Koes at the University of Pittsburgh and searching for the origin of R67 DHFR in collaboration with Dr. Janine Copp at University of British Columbia.

As part of the evolution that led me to do a PhD thesis, I would like to briefly thank previous project advisors and colleagues: Dr. Susana Camarero, my master thesis advisor, and in my undergraduate studies Dr. Jonathon Stone, Dr. Yingfu Li, Dr. Juan Perez-Mercader and Dr. Elena Marin. In addition, I would like to mention Dr. Isabel Pardo for her guidance during my master thesis, I transferred to my interns many of the knowledge you transferred me when I was your intern. Dr. Ivan Ayuso-Fernández for being a long-distance lab mate and my friend to be a geek about protein evolution. And more people at CIB where I did my master thesis: Laura (Lidem) providing FPLC support in the distance. Fellow biochemistry class-mates for all our conversations and PhD comics we have shared to rise our spirits: María José Queipo, Mateo Paz-Cabezas, Alicia Mata and Jose Manuel Sánchez-Zapardiel. And lastly Sandra Sánchez-Jaut, for all the songs, support and great visit to Montréal.

Finally, I would like to thank my close friends and family that have been there for me when I thought it was never going to stop snowing or my experiments were never going to work. You have all been a continuous source of support that has undoubtedly made this thesis possible. Within the department of biochemistry, I am eternally grateful to my friends Juliana, Lidice, Neylen and Yulemi. We have shared our struggles with our experiments and adapting to Québec and have been a great source of support to each other. I already miss not seeing you every day

mis Kendys. In Montréal, my “spanish” family has been another great source of support, always helping me to relativize and see that life is not only science. Thank you Tamara, Gorka, David, Kat, Fatou, Guille, Andy, Rodolfo, Alejandra, Emilio, Alejandra, Diego, Natalia, Hector, Martí, Ángela, Estefanía. I can’t summarize your many contributions and how much I will miss you but I hope we get to cross paths again somewhere! I am also very grateful to Mario for all his love, support and helping me see everything with perspective while finishing my PhD during a global pandemic. Lastly, but not least, I would like to thank my family. My parents have always showed me the importance of education and encouraged me to follow my interests. My parents and my sisters have cheered with me for every accomplishment and listened to me through every hardship and I am so grateful for the family I have.

Chapter 1 – Introduction

1.1 The dawn of evolution

The most popular definition of life was given by a committee gathered by NASA in 1994, who proposed that life is a "self-sustaining chemical system capable of Darwinian evolution" (1). It has been proposed that the dawn of evolution on our planet happened around 4 billion years ago when a pool of molecules started to self-replicate on an earth with high temperatures, an atmosphere rich in ammonia and poor in oxygen that resembles in nothing what we know today (2).

These first self-replicators were some of the earliest enzymes, molecules capable of accelerating chemical reactions by lowering the activation energy in converting a substrate to product. The integration of self-replication, compartmentalization and metabolism in a system allowed chemistry to become biology (3). The evolution of such a rudimentary compartmentalized chemical system, through billions of years, has led to all of the complex biodiversity that we know today. Currently, as the evolutionary biologist Richard Dawkins has said in The Selfish Gene, *we are survival machines—robot vehicles blindly programmed to preserve selfish molecules known as genes*. Indeed, if we paint a simplified picture of the main components of life today, genes formed by nucleic acids would be the storage units that hold and transmit the information necessary to synthesize the main molecular machinery, proteins. A crucial task of proteins is to provide the bulk of enzymatic machinery that sustains metabolic networks. Although other molecules like RNA and DNA have been shown to be able to act as enzymes, in this thesis I will focus on protein enzymes.

1.2 Protein world

Proteins are formed of the three-dimensional arrangement of a polypeptide composed of ordered combinations of the 20 different amino acids in the genetic code. In the case of enzymatic proteins, these have an active site that is often protected from the solvent where the substrate(s) is converted to product. Protein evolution occurs through mutations, insertions and

deletions in the coding DNA. These can lead to changes in the structure of the protein that ultimately affect its function. If the mutated protein has qualities that are advantageous to the fitness of the organism and increase its survival, it will more likely be propagated. Whether a mutation is advantageous or deleterious is dependent upon the environmental factors surrounding the organism, which act as selective pressures guiding evolution. A repetition of this process over billions of years has led to the evolution of specialized enzymes that can accelerate reaction rates up to 10^{26} times (4). Some enzymes are considered to have reached catalytic perfection, as their efficiency has reached values close to the rate at which substrate and enzyme can encounter in aqueous solution (5). However, evolution is an on-going process dependent upon the particular conditions that the organism encounters and the average enzyme still has room for improvement (6).

1.3 Laboratory evolution of enzymes

Until recently, we have been mere observers of the evolutionary processes around us. However, the same evolutionary processes that have taken place through billions of years can be mimicked in the lab to obtain protein variants with specific qualities within weeks. This process has been termed *directed molecular evolution*, as evolution is guided by applying user-defined selective pressures to a target protein. In directed molecular evolution, mutations are introduced in the gene coding for the target protein and the proteins screened for the desired quality; the fittest variant is selected (Figure 1.1). This process can be repeated for several rounds of mutation and selection until the specified goal is reached. The implementation of this concept was first done to modify the enzyme subtilisin E in 1993 and merited Dr. Frances Arnold the Nobel Prize in Chemistry in 2018 (7). The main advantage of directed molecular evolution is that, as in natural evolution, it is unbiased. Prior structural information of the protein is not necessary (8, 9).

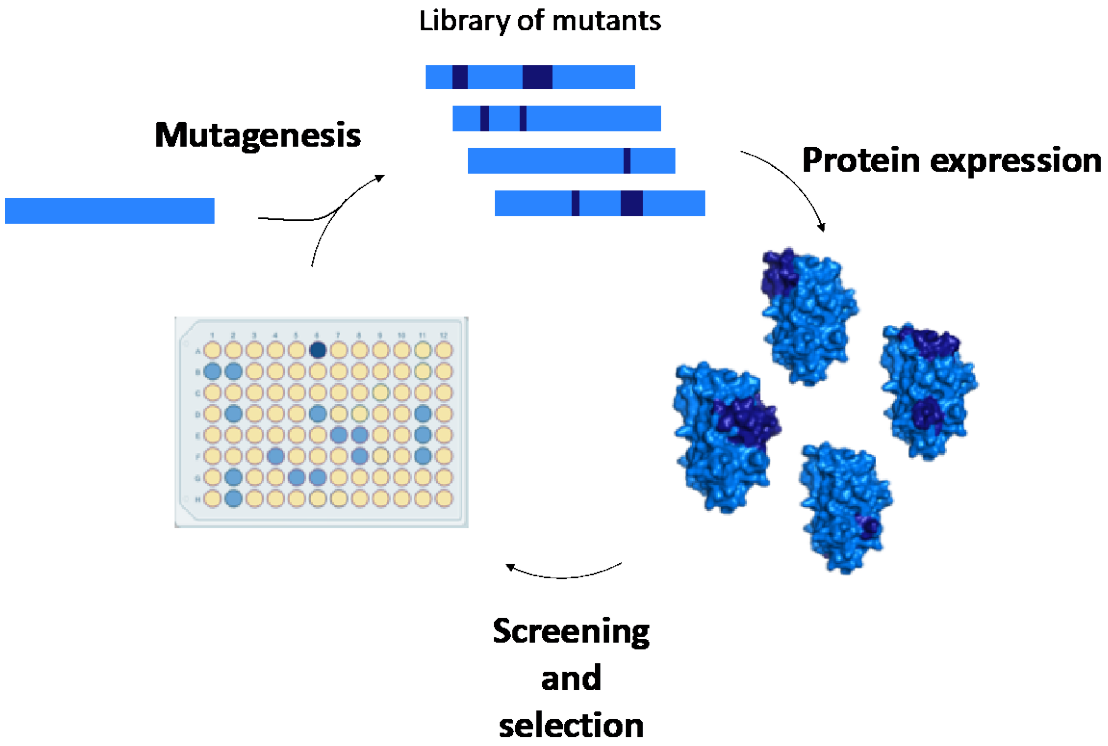


Figure 1.1. – Schematic representation of directed evolution.

A round of directed evolution starts by creating a library of mutants through mutagenesis of the gene encoding the target protein. Subsequently, the mutated proteins encoded in the library of mutants will be expressed and subjected to screening and selection. The fittest variant(s) selected can then be subjected to a subsequent round of evolution.

Many enzymes that are currently commercialized have been subjected to directed molecular evolution (10). Enzymes are a sustainable alternative to conventional chemical catalysts (11). They find use in the food and pharmaceutical industries, polymer synthesis, organic synthesis and biodegradation. The main advantages of enzyme-catalyzed reactions are their regiospecificity, stereospecificity and ability to function in mild reaction conditions of temperature, pressure and pH, reducing energy costs and need for specialized infrastructure. In addition, enzymes can be modified to improve their productivity, catalytic efficiency, specificity and tolerance to harsher conditions (8, 12-14). Enzymes can even be combined in cascade reactions to generate complex molecules (15).

The ability to modify enzymes to enhance or provide a new desired function has revolutionized biotechnological development. The term *protein engineering* was first suggested when it was thought that amino acid changes towards a new function could be easily rationalized. Indeed, *engineering* is defined in the dictionary as "the art or science of making practical application of the knowledge of pure sciences, as physics or chemistry". However, enzymes are inherently complex and whereas rational modifications are possible, protein engineers often rely on random mutagenesis or recombination of gene fragments combined with directed evolution (16, 17). Currently, it is more common to use 'semi-rational' methods where mutations are introduced randomly or specifically within a localized region of the protein (13, 16, 18-20).

A basic understanding of enzyme mechanisms and evolution has served to evolve enzymes to a new or improved function by mutating specific residues(8). In turn, laboratory evolution of enzymes has served as a tool to further understand enzyme mechanism and evolution (21-23). Our increasing understanding of the biophysical features of enzymes is making possible to make more rational choices in protein engineering. Computational efforts have improved predictability by considering biophysical features of enzymes and training algorithms with data derived from experimental evolution experiments (24-26). Some research groups are even designing complete enzymes *de novo* or incorporating alternative active sites to current enzymes (17, 27, 28). However, there is still room for improvement in the rational design of enzymes. In the words of members of Dr. Frances Arnold's lab, "What we've learned is that we know less than we thought we did. People learned the hard way that you couldn't just go in and make a mutation where you thought and expect it to improve an enzyme" (16).

1.4 Bottlenecks in laboratory evolution of enzymes

In vitro evolution of enzymes requires the creation of a library of variants and its screening for the desired function. Iterative rounds of this process are done to select the variant(s) with improved function or highest fitness. The main bottlenecks to effective exploration of all the amino acid combinations that will lead to the desired function are the creation of libraries of mutated variants, availability and efficacy of screening methods and the ability to predict successful mutations.

1.4.1.1 Library creation

The choice of methods for mutagenesis and cloning are two components of library creation that will determine the regions mutated, mutation rate and library size.

In the standard workflow of enzyme engineering, if the structure of the target enzyme or a close homolog is not available, random mutagenesis is the method of choice. Random mutagenesis can be done by using mutator strains, error-prone PCR or DNA shuffling (20, 29). In this scenario, the mutation(s) that will lead to an improved or new function can appear anywhere in the protein and the probability of success is low because most mutations are neutral or deleterious. Therefore, cloning methods that allow to consistently obtain library sizes of at least 10^5-10^6 variants are crucial (30). Another factor when choosing random mutagenesis is the mutation rate, which has to be adjusted to each system and will depend on the protein's tolerance to new mutations (31).

When structural information of the target enzyme is available, a rational or semi-rational approach can be chosen, in which 'hot-spots' are targeted (18). They may be based on proximity to the substrate binding site or to prior knowledge of the effect of mutations. This approach is preferable as it increases the chances of identifying a variant with improved function, while reducing the need for large library sizes or high-throughput screening methods. A series of mutagenesis and cloning methods used in rational and semi-rational design will be presented in Chapter 2 of this thesis.

1.4.1.2 Screening

Following the creation of libraries of variants of the target enzyme, it is crucial to have a method to screen for variants with improved activity. Library creation and screening methods are interconnected, hence the diversity created using a particular mutagenesis approach should consider the diversity that can be screened with methods available for the target enzyme (8, 30, 32). For example, a random mutagenesis approach may require screening of thousands to millions of variants and will thus require a high-throughput screening method to determine improved function. If this is not available, random mutagenesis can be reduced to specific regions of the protein to reduce library size. In fact, rational and semi-rational approaches are often

sought to reduce library size and thus reduce screening efforts (32) (Figure 1.2). A screening method is considered medium throughput when it can screen 10^3-10^5 variants per day and ultra-high throughput when it can screen 10^6-10^9 variants per day (32).

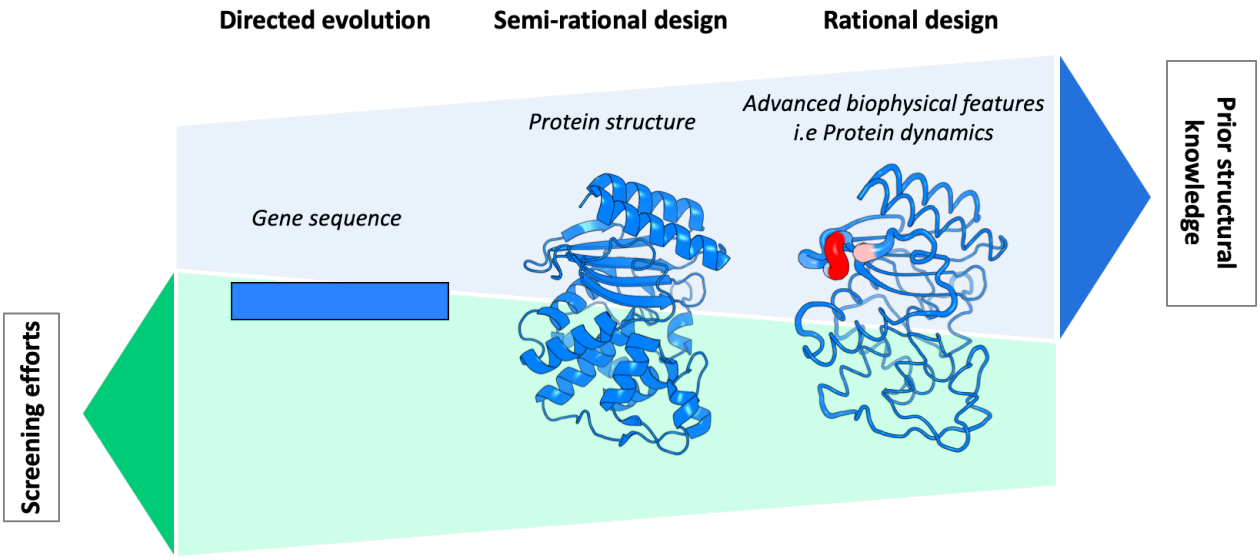


Figure 1.2. – Considerations to choose an enzyme engineering strategy based on prior structural knowledge and screening effort.

This figure was inspired by GenScript's Synthetic Precision Libraries brochure

Screening is an important bottleneck in enzyme engineering as the number of theoretical protein variants is practically unlimited. The availability of a fast high-throughput screening method allows to speed-up the exploration of the sequence space and increase success in enzyme engineering. However, high-throughput methods for the desired activity oftentimes cannot be developed or need specialized equipment, due to the nature of the chemical transformation (32). The most efficient high-throughput methods available often use specialized instruments such as droplet-based screening platforms combined with fluorescence-assisted cell sorting (FACS) (30, 32). These approaches employ fluorophores as markers for cell sorting: they are either quenched or activated after the target reaction takes place (32). Using conventional instruments, screening of 10^4-10^5 variants can be achieved in microtiter plate-based screens (30, 33). To use this approach, individual variants are grown in 96 or 384-well plates and assayed for specific

enzymatic activities using a colorimetric or fluorometric assay, mass spectrometry or chromatography (33). Throughput of this approach can be increased using robotic automation (30, 32).

An exception in high-throughput screening that does not require special equipment is when the target activity is linked to survival. This is only available when selecting enzymes that are essential for survival under conditions such as antibiotic resistance, metabolism of xenobiotic compounds or enzymes that synthesize components that are essential for cell growth (33). In this case libraries of ~10^9 variants can be easily selected for on selective agar plates: depending on the target activity, plates may contain and antibiotic, lack an essential nutrient or contain an otherwise toxic compound.

1.4.1.3 Predictability - protein epistasis

Predicting which mutations will be successful in laboratory evolution of enzymes would be straightforward if the effects of multiple mutations on function were always additive. In such cases, the combined phenotypic effect (i.e. catalytic activity, stability or other) of a set of mutations is the result of the addition of the individual effect of each mutation. If that evolutionary path to the highest activity was represented in a fitness landscape, the climb up to the highest peak would be smooth and follow what has been named the Mt. Fuji model (34, 35) (Figure 1.3).However, non-additivity in mutations is prevalent in enzyme engineering, preventing predictability of the combined effects of mutations (36-38). In this scenario, the climb up to the highest activity is not straightforward as the fitness landscape is rugged and might contain several peaks and valleys (Figure 1.3)(35).

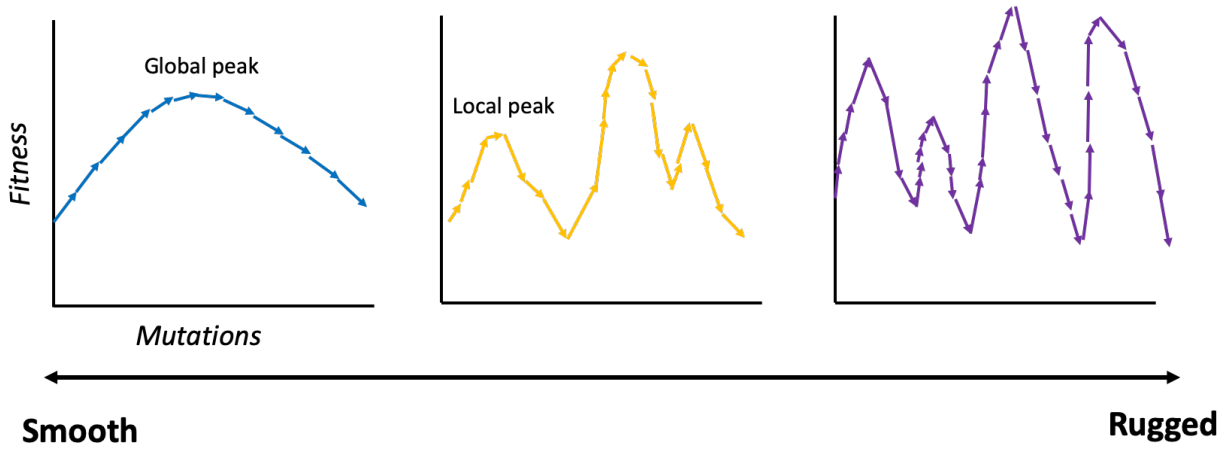


Figure 1.3. – Representation different topologies of fitness landscapes.

The evolutionary trajectories observed are the result of acquiring mutations that alter fitness. The more rugged the fitness landscape, the more local peaks and valleys that exist. Reaching a global peak is easier when the topology of the fitness landscape is smoother.

Non-additivity of the combined effects on activity of mutations in enzyme engineering is defined as *protein epistasis* (35, 39). Protein epistasis can be classified as magnitude epistasis when the combined phenotypic effect of the single mutations is greater (positive or synergistic) or weaker (negative or antagonistic) than the addition of the individual effect of the single mutants (35, 39). Sign epistasis occurs when the combined phenotypic effect is the inverse of that of the single mutants. For example, the combined benefit of mutations can be greater even though one or more of the single mutants is deleterious (positive sign epistasis) or it can be lower even though one or more of the single mutants was beneficial (negative sign epistasis). Sign epistasis can also be reciprocal if there is a change in sign for both of the single mutants when combined, that is, where the combination of favorable mutations results in a deleterious effect, or vice-versa (35) (Figure 1.4). Most studies on protein epistasis to date focus on the characterization of the interaction of between two mutations, or pairwise epistasis (40-42). Recently, the study of more complex interactions between three or more mutations, termed high-order epistasis, has gained in popularity (43-46).

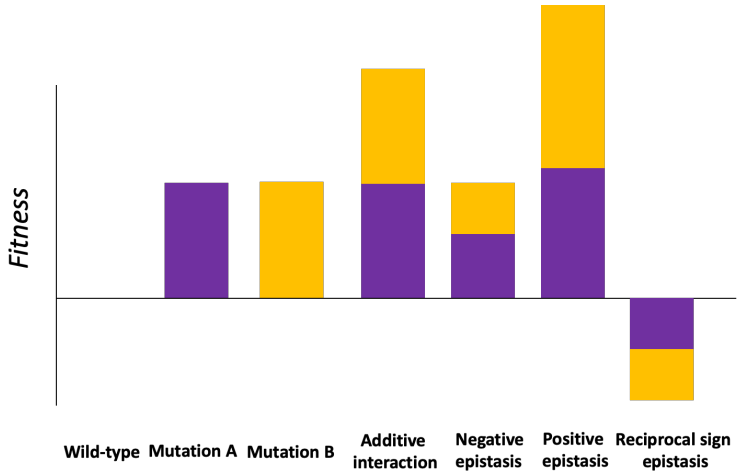


Figure 1.4. – Types of pairwise epistasis.

The additive effect of the simultaneous incorporation of beneficial mutations A and B is shown relative to epistatic interactions. In negative epistasis, the fitness of the double mutant is lowe than the sum of the two beneficial effects on fitness, in positive epistasis it is higher and in reciprocal sign epistasis the effect of adding both mutations is opposite – therefore negative in the case of individually beneficial mutations.

Protein epistasis will constrain protein evolution by shaping the evolutionary paths available in the evolution towards a new function (35, 39, 47). For example, the presence of multiple peaks in the evolution of a given enzyme suggests the presence of reciprocal sign epistasis (47).

Indeed, through studies using the gold-standard enzyme in directed molecular evolution, TEM-1 ß-lactamase, it has been shown that sign epistasis limits the theoretically accessible paths towards a new function whereas positive epistasis guides its evolution (42, 48, 49). Furthermore, analysis of the evolutionary paths of nine different enzymes towards new function has revealed that positive epistasis is prevalent in protein evolution (37). More recently, studies on several fitness landscapes have revealed that the influence of epistasis on protein evolution is reduced when the epistatic order is increased (number of interacting residues) (43). In fact, it has been suggested that the magnitude of epistasis has a larger influence on evolutionary paths than the epistatic order (44).

Although efforts are being made to understand protein epistasis in order to incorporate it rationally into enzyme engineering workflows, to date, epistatic interactions are mostly revealed *a posteriori* through directed evolution or combinatorial evolution experiments (36, 39, 50). Advances in library generation and sequencing technology have made it feasible to assess the individual effect of mutations in a given protein through deep mutational scanning (51, 52). This approach has even been used to study epistatic interactions of all possible double mutants in the 55 residue protein GB1 (40). However, this approach has its limitations due to exponential library sizes needed as protein size increases and/or number of interacting residues increases. In addition, pairwise analysis will not reveal more complex interactions, such as high-order epistasis that happens in natural and directed evolution. Recently, a method to study high-order epistasis by reducing library size was proposed (53). This method was applied to analyze all possible epistatic interactions that exist between 6 positions that changed in the evolution of a phosphotriesterase enzyme into an arylesterase (54, 55), thoroughly analyzing a library of 2^6=64 variants. Overall, developing models to understand epistasis will eventually procure a global understanding of epistatic effects and enable their successful incorporation into protein design.

1.5 Epistasis, evolution and inherent biophysical properties of proteins

Protein epistasis results from the specific biophysical context of each protein, making it particular to each enzyme. Direct interactions between residues can be easy to predict, yet physical interactions between mutations or nearby residues can affect global properties of the enzyme such as stability, affinity, catalysis or protein dynamics (39). Naturally, epistasis resulting from long range-interactions is more complex and harder to incorporate into protein engineering (37, 56-60). Whereas the impact that protein epistasis has on predictability in enzyme engineering is evident, the interplay between molecular evolution, epistasis and biophysical properties such as stability and dynamics is not (39).

1.5.1 Protein stability

Protein stability shapes the evolutionary paths available in the evolution towards a new protein function by modulating the quantity of functional protein available (61-63). It encompasses thermodynamic stability as well as the time the protein remains in its native state

without unfolding, aggregating or degrading, defining its kinetic stability. Thermodynamic stability is more often considered as it can be predicted *in silico* (64) or determined *in vitro* (65), procuring a measure of the quantity of folded protein available. It has been shown that most mutations are destabilizing (66-68). Although gain-of-function mutations in key catalytic residues have been shown to be statistically more destabilizing (68).

Proteins have been shown to tolerate destabilizing mutations up to a certain stability limit beyond which the effect of such mutations is apparent, resulting in a fitness landscape dominated by negative epistasis (42, 62). Thus, mutations that increase protein stability promote the evolution towards new functions by buffering the effect of destabilizing mutations (69). (Figure 1.5). Indeed, the appearance of such compensatory stabilizing mutations in directed molecular evolution experiments has been shown to be adaptive (70). That protein stability promotes evolvability agrees with reports of higher evolvability in highly thermostable, ancestrally reconstructed proteins and in thermophilic proteins (71-74).

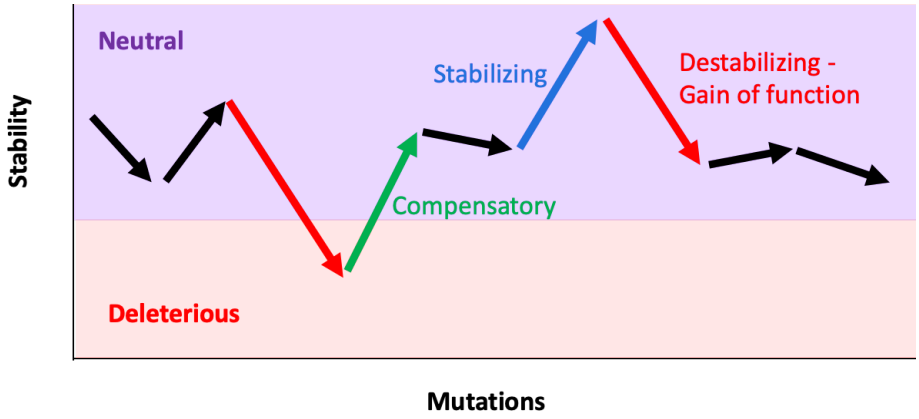


Figure 1.5. – Protein stability and evolvability.

The neutral range of stability determines which mutations can be acquired in the course of evolution. Inspired by Tokuriki and Tawfik (62).

The interplay between protein epistasis and stability has been shown to be prevalent in long-term evolution(75, 76). A mutation can act epistatically through direct interactions that increase protein stability. Alternatively, it can act through long-range interactions, such as when a destabilizing mutation is evolutionarily 'fixed' in an enzyme, complemented with a distal stabilizing residue (39, 77). Long-range interactions can be mediated by protein dynamics, such

as when a mutation exerts its stabilizing effect by increasing the rigidity of the protein scaffold (77, 78).

1.5.2 Protein dynamics

Enzymes have been traditionally described as static structures as a result of their observation under crystalline form. However, the past two decades abound with evidence showing that proteins are inherently dynamic (57, 79-81). Moreover, the relationship between protein dynamics and function has been demonstrated for several protein systems (82-84); interestingly, proteins with similar function have been shown to exhibit similar dynamic patterns (85). The functional importance of protein motions is such that they have been shown to be selected for in natural and directed evolution (57, 58, 80, 83, 86-88). That protein dynamics are important for function is so accepted in the scientific community now, that even in the short amount of time that we have lived the COVID-19 pandemic, we find studies on the flexibility of the SARS-CoV-2 spike protein (89, 90) .

Protein dynamics is a complex trait that encompasses motions on different timescales and of differing amplitude. Bond vibration, side-chain rotation, hydrogen bonding and proton transfers occur at fast timescales (ps to ns) whereas loop rearrangements to accommodate the ligand or substrate turnover occur at slow timescales (μs to ms) (79, 91, 92) (Figure 1.6).

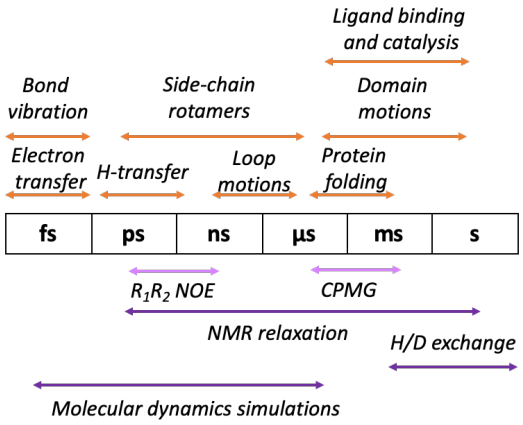


Figure 1.6. – Motions associated to each timescale in protein dynamics.

The different protein motions are linked to their relevant timescale (orange arrows) and the methods that can be used to monitor the protein dynamics at specific timescale are depicted in purple arrows. Figure inspired by Henzler-Wildman and Kern (79) and Ortega, Pons (92).

All of the conformations that a protein can adopt due to its inherent flexibility form a heterogenous population in solution. Certain conformers may be more abundant due to their greater thermodynamic stability. As a general rule, the most stable conformers are those crystallized; occasionally we are fortunate to observe different conformers in x-ray crystallography (80). Enzymes that can accept multiple substrates must adopt different conformations to accommodate and turn over the different ligands (93-95). This conformational flexibility allows enzyme promiscuity and can be the basis for the evolution towards new function even in an enzyme that natively displays low promiscuity. Indeed, it has been hypothesized that protein motions can modulate evolution by favoring certain conformations that increase activity towards a new function upon the introduction of mutations (96) (Figure 1.7). This increase in the breadth of activity can be the consequence of reshaping the active site cavity to accommodate a new ligand or by an indirect mechanism such as increasing protein stability (96).

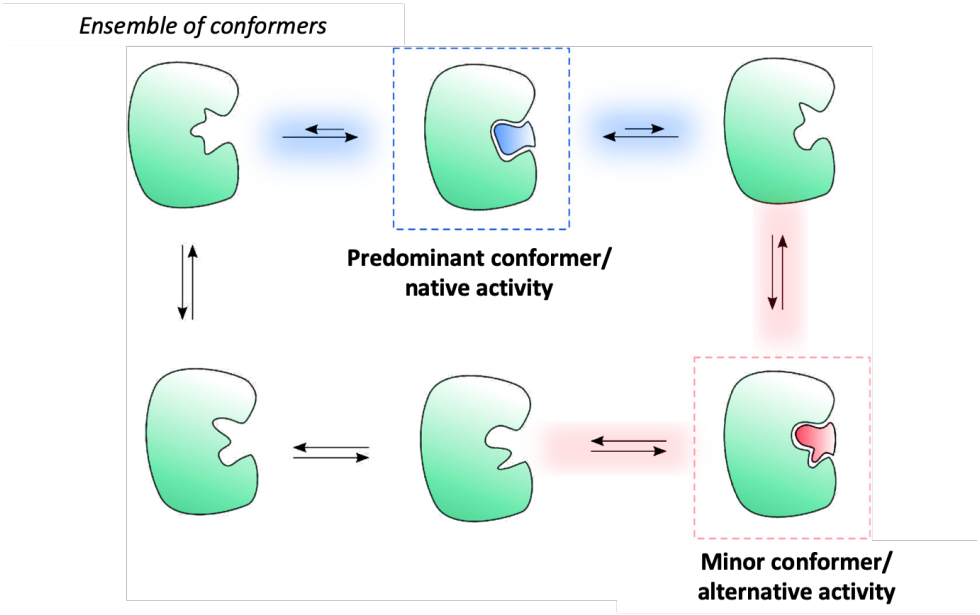


Figure 1.7. – Protein dynamics and evolvability.

An initial ensemble of conformers (top left) has a predominant conformer (top middle) that confers the native function and alternative conformers that may provide alternative functions. Evolution of the new function can happen through the enrichment of conformers that confer alternative functions (i.e mutation that stabilizes alternative conformer; bottom right). Adapted from Petrovic, Risso (57).

The link between protein epistasis and protein dynamics has not been thoroughly explored. However, there are reports of negative epistasis, positive epistasis and sign epistasis occurring due to modifications in protein dynamics in different enzymatic systems. In a recent study using transketolase, it was demonstrated through molecular dynamics (MD) simulations that mutations as far away as 33 Å interacted epistatically via a network of interactions that could modify the flexibility in distal parts of the protein (77). In another study, an increase in protein dynamics in a key region in the active site has been linked to negative epistasis in TEM-1 ß-lactamase (78). In the BcII metallo ß-lactamase, sign epistasis was observed between two mutations that increase slow dynamics close to the active site region (97). Interestingly, the deleterious mutation alone eliminates slow dynamics but increases dynamics in the presence of the advantageous mutations, giving rise to the sign epistatic effect. Sign epistasis has also been correlated to an increase in protein dynamics at slow timescales in the enzyme Cyclophilin A (98).

Incorporating knowledge of protein dynamics to predict epistatic interactions is complex due to the inherent complexity of protein dynamics. Currently no single methodology can be used to fully characterize the dynamic landscape of a protein (Figure 1.6) (88, 91, 99-101). Few examples in which NMR and MD methodologies are combined to characterize the dynamic landscape of enzymes are available (80, 81). However, there are recent examples in which protein dynamics has been incorporated into protein engineering (60, 102, 103).

1.6 Research objectives

In the previous sections, I highlighted the importance of enzyme engineering as well as the main bottlenecks that need to be addressed to better direct engineering efforts. In particular, I focused on protein epistasis and its interplay with inherent biophysical properties of enzymes such as protein dynamics. Improving predictability in enzyme engineering is a key question in the field and protein epistasis has been postulated to be one of the main mechanisms hindering predictability. To that end, this thesis examines mechanisms by which protein epistasis acts on the evolution of new protein function.

Objective 1: Interplay of protein dynamics and epistasis on the evolution of new protein function

Protein dynamics is an intrinsic trait of proteins that is selected for during evolution due to its structural importance. However, whether a highly dynamic enzyme would be more attractive as a start-point in an enzyme engineering project remains unanswered. Moreover, it is unknown whether protein motions at specific locations or timescales have a positive, neutral or negative effect on evolution towards a new protein function. Differences in evolution are due to variations in the accessibility of evolutionary trajectories. As protein epistasis is a major mechanism by which accessibility to evolutionary trajectories is determined, another question that arises is whether protein dynamics affect protein epistasis. The first objective of this thesis is to analyze the interplay between specific aspects of protein dynamics, epistasis and evolution.

Objective 2: Characterization of protein epistasis in a biotechnologically relevant enzyme

Protein epistasis is a prevalent mechanism in enzyme evolution. However, its importance in the evolution towards new functions is specific to each enzyme. Enzyme engineering in industrially relevant enzymes is often targeted towards a specific goal. Therefore, investigation of basic mechanistic aspects, such as the relevance of protein epistasis, is often lacking. However, characterization of protein epistasis in relevant enzymes with potential biotechnological interest could aid to accelerate their engineering. The second objective of this thesis is to examine the prevalence of non-additive in the directed evolution of an industrially relevant enzyme.

1.7 Description of model systems chosen based on research objectives

In order to study the interplay of protein dynamics and epistasis on the evolution of new protein function (Objective 1), we have chosen dynamic variants of an enzyme that has been well characterized in terms of structure-function-evolution relations. These dynamic variants are derived from TEM-1 ß-lactamase (TEM-1), a well-accepted model in molecular evolution to understand the rise in antibiotic resistance, as well as protein evolution in general (31, 49, 68, 104-110). The dynamic landscape of these variants has been characterized. They serve as starting-points with known differences in dynamics for undertaking directed evolution in parallel.

Secondly, in order to study the prevalence of non-additivity in directed molecular evolution, the enzyme *Moesziomyces antarcticus* lipase A (previously *Candida antarctica* - Cal-A) was used (Objective 2). This enzyme was previously engineered in the Pelletier lab in collaboration with the industrial partner DSM and has broad applications in the biotechnological industry. In this thesis, we identify epistatic mutations in the evolution of this enzyme towards substrate selectivity. Few studies on its reaction mechanism, applications or evolution are available compared to other more widely used lipases (ie. *Candida antarctica* lipase B) and studies on its epistasis could be highly advantageous in further evolution projects.

1.7.1 Model 1: Dynamic variants derived from TEM-1 ß-lactamase

ß-Lactamases (EC 3.5.2.6) are bacterial and fungal enzymes that catalyze the hydrolysis of antibiotic compounds containing a ß-lactam ring, rendering them inactive. First described in 1940, they are estimated to have originated as long as 2 billion years ago as a defense mechanism of bacteria against organisms secreting antibiotic ß-lactam-type compounds (111). Currently, ß-lactamases are the main cause of resistance to ß-lactam antibiotics (112-114). ß-Lactamases have been classified into four groups according to their reaction mechanism and sequence similarity (115). Classes A, B and D have a catalytic serine whereas class C have a catalytic Zn^{2+}. The serine ß-lactamases are the most extensively characterized. They catalyze ß-lactam hydrolysis via formation of an acyl-enzyme intermediate between the substrate and active site serine. Among serine ß-lactamases, Class A serine ß-lactamases are the most widely characterized due to their high adaptability to new ß-lactam antibiotics (116).

The ß-lactamase variants cTEM2m and cTEM17m were created *in vitro* through recombination of two class A ß-lactamase homologs, TEM-1 and PSE-4, that share 40% sequence identity, high structural similarity (RMSD = 0.98 Å) and differ slightly in terms of ß-lactam substrate specificity(108). Compared to TEM-1 ß-lactamase, cTEM2m has two mutations close to the catalytic Ser70, whereas cTEM17m has 17 substitutions in other areas of the large active site pocket (Figure 1.8). This means that both cTEM2m and cTEM17m are more closely related to TEM-1 with sequence identities of 99.2% and 93.5% to TEM-1, respectively. The function and the

dynamic landscape of variants cTEM2m and cTEM17m from milliseconds to femtoseconds were characterized in the Pelletier lab, as detailed in the following sections.

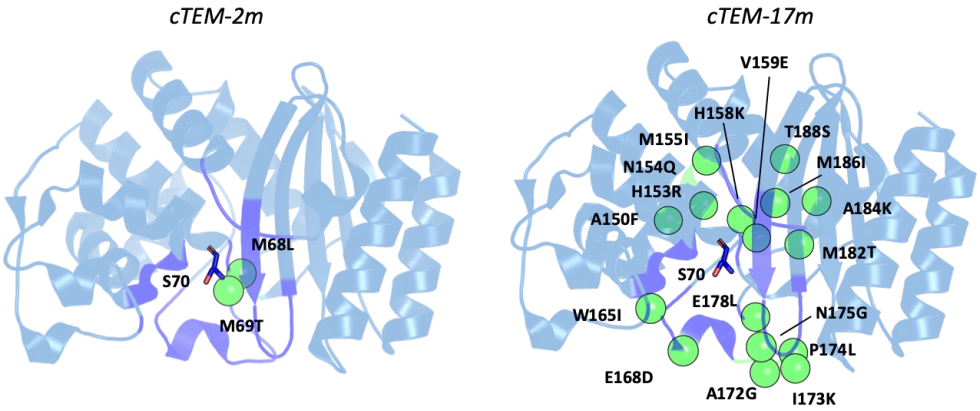


Figure 1.8. – 3D structure of cTEM-2m and cTEM-17m ß-lactamases indicating residues that are mutated relative to TEM-1.

The active site walls are highlighted in dark blue, the catalytic S70 is shown as sticks and the mutated residues as green spheres in the crystal structure of cTEM-2m (PDB ID: 4mez) and cTEM-17m (PDB ID: 4id4).

1.7.1.1 Kinetic and structural characterization of ß-lactamase variants cTEM2m and cTEM17m

Enzymatic assays of cTEM2m and cTEM17m using a range of penicillin and cephalosporin substrates showed that substrate recognition was maintained with respect to the most closely related parental enzyme, TEM-1 (80, 117). Specifically, the hydrolysis of two penicillins (carbenicillin and benzylpenicillin) and three cephalosporins (cephalotin, cefotaxime and cefazolin) by the two variants were assayed. The catalytic efficiency against penicillins was in the same order of magnitude for the variants and the parental TEM-1 (Table 1.1). This trend was maintained with cephalosporins except for cTEM2m that displayed ~10-fold lower catalytic efficiency compared to cTEM17m and TEM-1.

k_{cat}/K_M (M^{-1} s^{-1})	Benzylpenicillin x 10^7	Carbenicillin x 10^6	Cephalotin x 10^5	Cefazolin X 10^5	Cefotaxime X 10^2	T_m(°C)
TEM-1	2.3	1.9	4.7	4.2	8.3	49.9
cTEM-2m	1.2	1.5	0.6	0.4	0.8	49.6
cTEM-17m	1.7	1.7	2.4	1.2	5.4	49

Table 1.1. – **Catalytic efficiency and thermal stability of TEM-1, cTEM-2m and cTEM-17m**

Data obtained from Gobeil, Ebert (80)

The native fold of TEM-1 ß-lactamase consists of an all-α domain and an α/ß domain with the active site in the interface of both domains. The all-α domain is comprised of eight α-helices and the α/ß domain is comprised of a ß-sheet composed of five ß-strands and three surrounding α-helices. Four motifs are conserved in the active site pocket of class A ß-lactamases: S70-X-X-K73; the SDN loop formed by residues S130-D131-N132; K234-T/S235-G236 and the Ω-loop residues 163 to 178 (118, 119) (Figure 1.9). Comparison of the crystal structures of cTEM2m (PDB ID: 4mez; RMSD = 0.396) and cTEM17m (PDB ID: 4id4; RMSD = 0.381) with that of TEM-1 (PDB ID: 1xpb) confirmed the maintenance of the native fold, with some minor differences (80, 120).

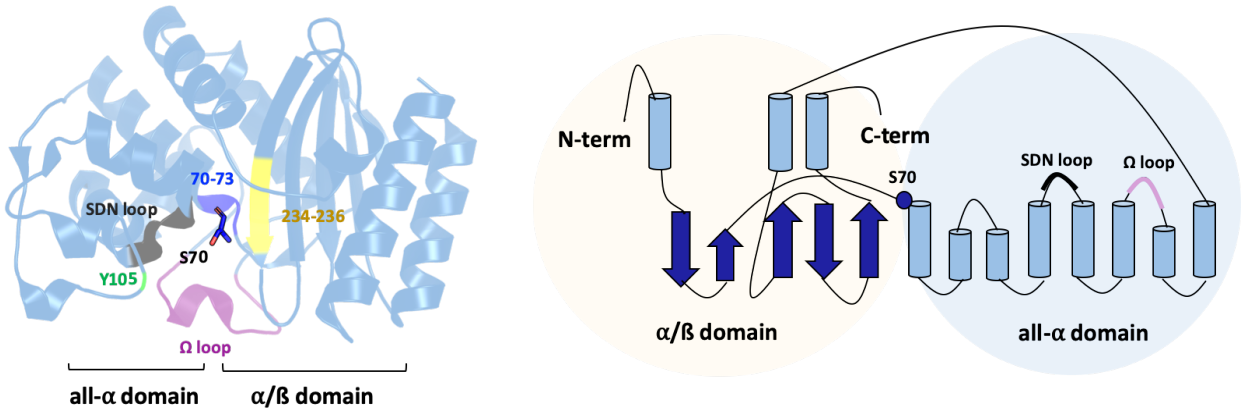


Figure 1.9. – Three-dimensional structure and diagram of TEM-1 ß-lactamase showing its key functional regions.

The conserved motifs are shown in the 3D structure (PDB ID: 1xpb): 70-73 (blue); SDN loop (black); 234-236 (yellow) and Ω loop (purple). The all-α and α/ß domains are shown in both representations. Figures created with PyMOL Molecular Graphics System, Version 2.1 Schrödinger, LLC and Microsoft Powerpoint 16.16.10.

The dynamic landscape from milliseconds to femtoseconds of variants cTEM2m and cTEM17m was determined using NMR and molecular dynamic simulations. The microsecond-to-millisecond timescale (defined as slow) was assayed using Carr-Purcell-Meiboom-Gill (CPMG) NMR, widely used to probe protein dynamics at this timescale (80, 117). The fast, picosecond-to-nanosecond timescale was determined using R_1-R_2-NOE NMR (120). Finally, the intermediate timescale from nanosecond-to-microsecond was determined computationally after benchmarking the method with experimental data obtained for the fast timescales (80). These studies revealed an overall conservation of motions at fast and intermediate timescales in the parental ß-lactamases TEM-1 and PSE-4, as well as in the cTEM2m and cTEM17m variants. In contrast, important differences in dynamics were observed in the studied variants at the slowest timescales, with respect to TEM-1 and PSE-4 (Figure 1.10).

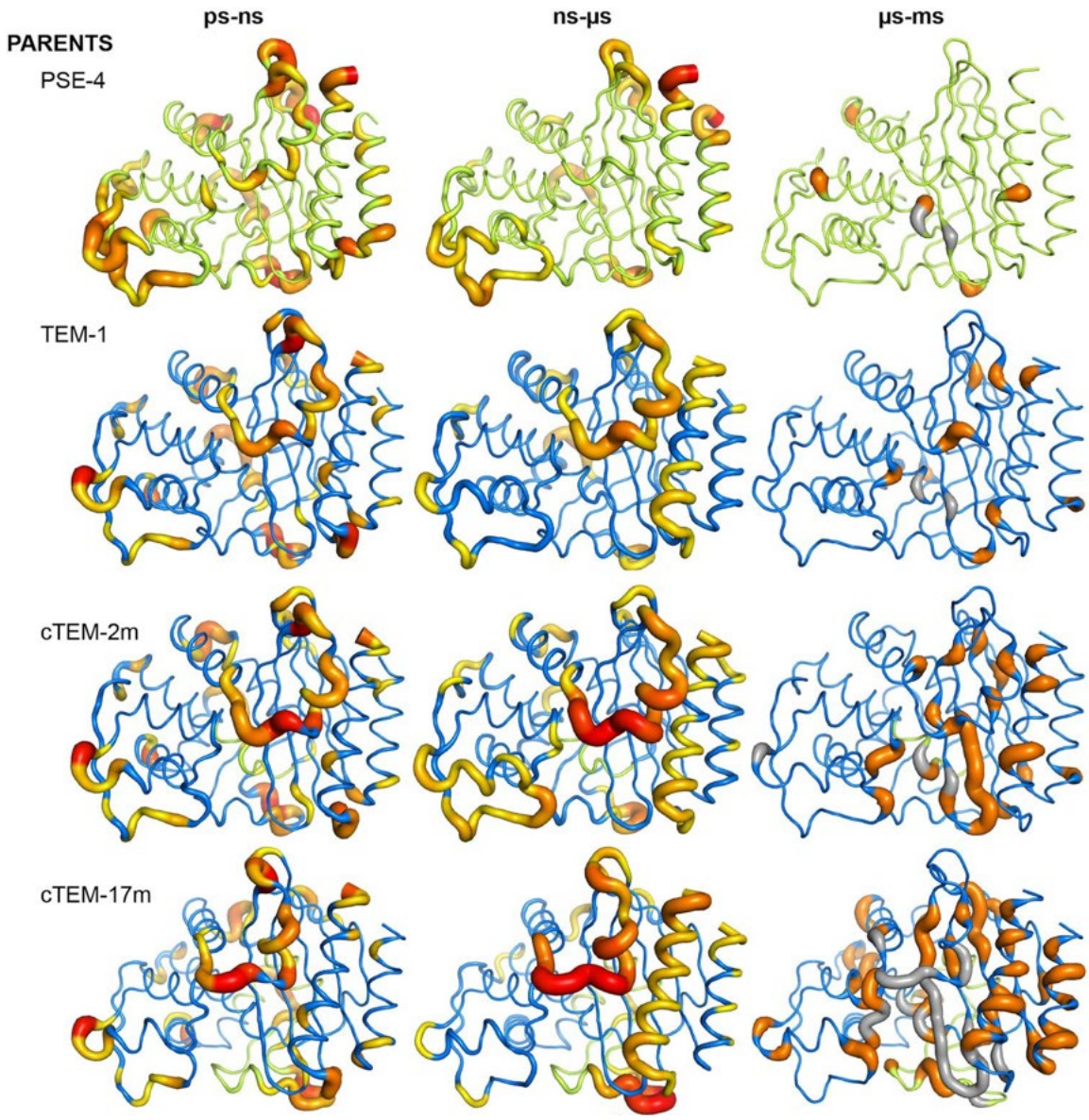


Figure 1.10. – Dynamic landscape of PSE-4, TEM-1, cTEM-2m and cTEM-17m ß-lactamases.

Protein dynamics on three timescales are shown for the parental PSE-4 (PDB ID: 1g68), and TEM-1 (PDB ID: 1xpb) ß-lactamases and their variants cTEM-2m (PDB ID: 4mez) and cTEM-17m (4id4). Dynamic residues are colored from less to more dynamic in a yellow to red gradient for each timescale (labelled above each column). Unassigned residues in the CPMG NMR determination of the µs to ms timescale are shown in grey. It should be noted that unassigned residues are proposed to be highly dynamic. Modified figure from Gobeil, Ebert (80)

Variant cTEM2m presented slow motions in 29 residues, whereas cTEM17m has 82 residues displaying dynamics at this millisecond-to-microsecond timescale. The dynamic residues in cTEM-

17m are more broadly distributed throughout the active site than those of cTEM2m and include the catalytically-relevant omega-loop region. Overall, protein dynamics at slow timescales in these variants vary in location and in magnitude incrementally from TEM-1 to cTEM2m to cTEM17m. It is remarkable that this increase in protein motions at the timescale of catalytic turnover does not importantly alter substrate reactivity in the variants (Table 1). Similarly, protein stability measured with circular dichroism and thermal scanning fluorescence revealed no significant change in the melting temperature of cTEM2m and cTEM17m despite the observed variances in slow protein motions (80, 117) (Table 1).

1.7.1.2 Current knowledge on TEM-1 ß-lactamase reaction mechanism

The reaction mechanism of serine ß-lactamases has been largely characterized. It involves the acylation of the active site Ser70 upon nucleophilic attack on the ß-lactam ring of the substrate. This is followed by deacylation mediated by an active site catalytic water activated by E166 (119, 121-125). The acylation mechanism remains controversial (119, 126). The two leading hypotheses and evidence pointing to an activation of S70 as a nucleophile, through K73 and S130 or a water molecule coordinated to E166 (119, 122, 127-130) (Figure 1.11). It should be noted that the rate-limiting step of the reaction varies with the substrate, being deacylation for penicillins and acylation for cephalosporins (126).

Other key residues relevant for substrate binding and catalysis are Y105, SDN loop, N170, K234, S235 and A237 (119, 126). Y105 has been proposed to be a gate-keep residue important for substrate stabilization (131, 132). Residues in the SDN loop are key for stabilization of the enzyme transition state, as well as maintaining the structure of the active site pocket (119, 126). More in particular, S130 in the SDN loop has been proposed to be implicated in acylation and N132 to stabilize and position the substrate (122). N170 in the Ω-loop interacts with E166 to position the residue for the activation of the active site catalytic water (126). K234 and its neighbouring S235 interact with the substrate and may be implicated in stabilization of the transition state (122). Mutations in the 234-236 regions have been shown to have a greater effect on the catalytic efficiency of cephalosporin hydrolysis than penicillin hydrolysis (119, 133, 134). Lastly, the

backbones of residues S70 and A37 form the oxyanion hole, key for stabilization of the tetrahedral acyl-enzyme intermediate (122).

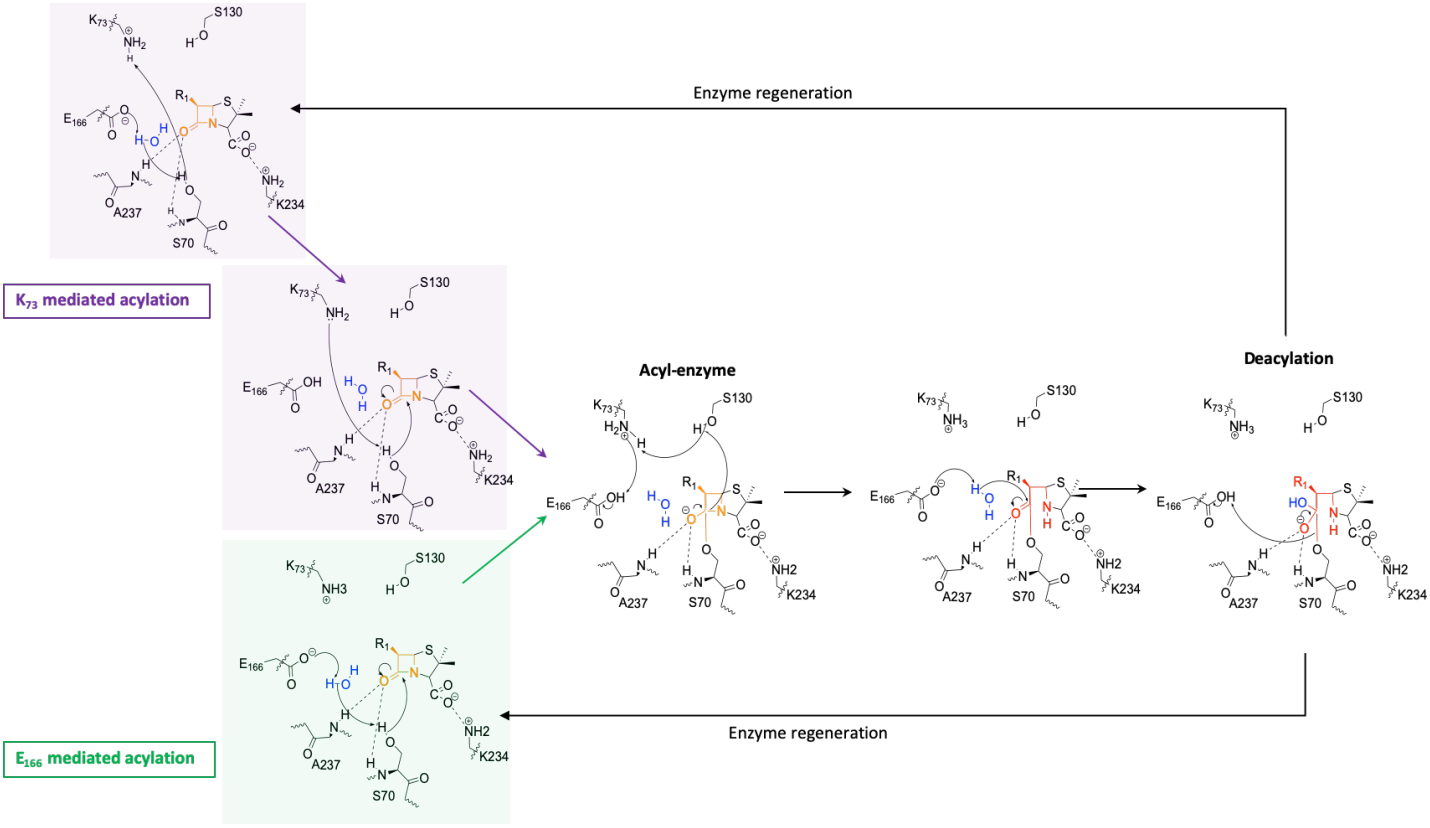


Figure 1.11. – Schematic representation of the reaction mechanism of TEM-1 ß-lactamase.

Activation of S70 to form the acyl-enzyme intermediate is shown either via E166 by activating a catalytic water (green) or via K73 (purple). Subsequently, S70 performs a nucleophilic attack on the carbonyl group of the ß-lactam ring. Deacylation then proceeds by an attack by an active site catalytic water activated by E166. Following deacylation, the substrate is released and the enzyme is regenerated and ready for another catalytic cycle. Adapted from Drawz and Bonomo (119) and Doucet (135). Figure created in ChemDraw Professional 16.0.

1.7.1.3 TEM-1 as a model in directed molecular evolution

TEM-1 has been used as a model to study molecular evolution due to the high adaptability of ß-lactamases to natural and semi-synthetic ß-lactam antibiotics (114), as well as the ease of screening for ß-lactamase variants, since this can be done by differential survival in the presence of ß-lactam antibiotics. Furthermore, the simplicity of this single-substrate reaction that requires no cofactor facilitates *in vitro* analyses. Therefore, there is a rich database of mutations found in experimental evolution, deep-mutational scanning as well as in clinical isolates. As a

consequence, there are several documented epistatic mutations that have been identified and studied (126). The substitutions identified at the highest frequency in antibiotic-resistant clinical isolates and experimental evolution are Q39K, E104K, R164C, R164H, R164S, A237T, G238S and E240K (107, 126, 136). These mutations provide the means to hydrolyze newly developed antibiotics by modifying the configuration of the active site pocket. These variants have also been commonly found to result in positive epistasis with other mutations, such as G238S/E240K, E104H/G238S, E104K/G238S, A237T/R164S, A237T/E240K, R164S/G238S and Q39K/E104K/R164S/A237T/E240K(126). Whereas most of these mutations are in the active site cavity, the majority of cefotaximase-enhancing mutations are found >10 Å away from the catalytic S70 (Figure 1.12).

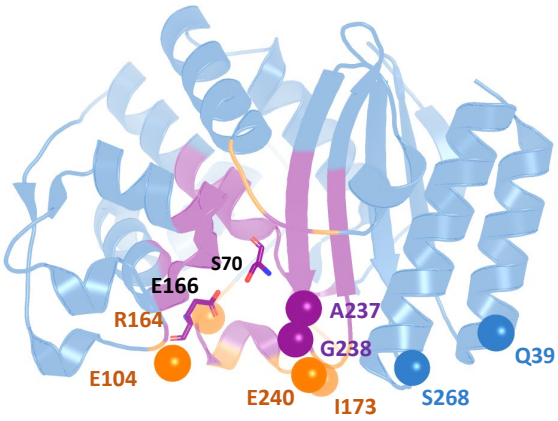


Figure 1.12. – Cefotaximase mutations in TEM-1 ß-lactamase.

The region colored in orange corresponds to the active site walls from which the 10 Å radius from the catalytic S70 is colored in purple. Catalytic E166 is shown as purple sticks. The residues were cefotaximase mutations Q39K, E104K, R164C, R164H, R164S, I173V, A237T, G238S, E240K and S268G occur are shown as spheres colored according to their location. Figure created with PyMOL Molecular Graphics System, Version 2.1 Schrödinger, LLC.

In fact, prevalent cefotaximase mutations in TEM-1 distal from the active site have been shown to be allosteric sites modulating antibiotic and inhibitor resistance (137, 138). In addition, stabilizing mutations far from the active site such as H153R, A224V and R275L (136, 139) or the well-known M182T that is often encountered in conjunction with E104K/G238S have been shown to be important in TEM-1 evolution (48, 49, 106, 107, 136).

1.7.1.4 Protein dynamics and evolution in TEM-1 ß-lactamase

Protein dynamics is an inherent feature of proteins. Characterization of the dynamic landscape of the two related class A ß-lactamases TEM-1 and PSE-4 has shown an overall, strong conservation of protein motions, consistent with protein dynamics being selected for during natural evolution (80). The implications of protein dynamics in the evolution of new protein function have been explored in previous studies using TEM-1 ß-lactamase due to its ample use as a model in molecular evolution. MD simulations of cefotaximase mutations far from the active site in TEM-1, Q39K, I173V, E240K, and S268G, have revealed that these modify dynamics of the active site residues S70 and E166 despite having no direct interaction (138). Moreover, it was shown that mutations in medium flexibility regions distal from the active site could modulate dynamics in the active site cavity (138). MD simulations of ancestral reconstructed ß-lactamases ranging from 53 to 79% sequence identity with TEM-1 have shown that higher dynamics correlated with higher substrate promiscuity. Specifically, it was suggested that a decrease in active site dynamics contributes to greater substrate specificity (93). In another study, the introduction of the synergistic mutations E104K/G238S into TEM-1 and two TEM-1 variants containing stabilizing mutations, revealed greater active site flexibility accompanied with increased rigidity in the surrounding areas, a smaller loss of thermal stability and identical kinetic profiles (110). The smaller loss of thermal stability in the stabilized TEM-1 variants (loss of 3°C compared to loss of 16°C in TEM-1) along with the altered flexibility led the authors to suggest that these stabilized variants could tolerate further destabilizing mutations and evolve further. In other words, they hypothesized that a rigid structure with a flexible active site could promote the evolution of new protein function in TEM-1 (110).

A more detailed study into the mechanism by which E104K/G238S confer positive epistasis for cefotaxime revealed that a decrease in conformational dynamics of the omega-loop and stabilization of particular conformations correlates with higher cefotaximase activity (140). Interestingly, a study of the negative sign-epistasis mutations R164S/G238S suggested that too much flexibility in the omega-loop regions could prevent evolution towards new functions (78). In particular, the introduction of mutation R164S increased flexibility in the omega-loop region leading to non-productive omega-loop conformations (78).

Overall, these reports suggest that protein dynamics play a key role in the evolution of new activities in TEM-1 ß-lactamases. However, it is relevant to note that all the work mentioned until now refers to general protein dynamics, with no timescale associated, obtained through normal-mode simulations or from normalized B-factors of X-ray crystal structures. Studies in which dynamic timescales are studied in detail are few as they require longer experimental time and expertise in NMR and MD simulations. Nevertheless, they have the potential of providing more detail into the specific mechanisms by which protein dynamics alter function, ultimately affecting protein evolution.

In TEM-1, it is known that specific protein motions are important for catalysis. For example, hydrogen-deuterium exchange in TEM-1 has revealed that protein dynamics at the millisecond timescale in the region comprised of residues 250-257 are necessary for cephalosporin turnover, but not for ampicillin turnover (141). In another study, it was revealed that changes in fast and slow timescales upon the introduction of mutation Y105D decrease catalytic efficiency (131). This data supports the idea that further work using systems whose whole dynamic landscape has been characterized is needed to dissect the evolutionary implications of each motion and timescale.

1.7.2 Model 2: Moesziomyces antarcticus lipase A (Cal-A)

Lipases (EC 3.1.1.3) are enzymes found in all kingdoms of life. They catalyze the hydrolysis of triacylglycerides into diacylglycerides. In addition, lipases can catalyze esterification, interesterification and transesterification reactions. These properties, along with their ability to catalyze reactions in water:oil emulsions, their substrate versatility and their regio/enantioselectivity have made lipases one of the most prevalent types of enzymes used in biotechnology (142-144). This class of enzymes is widely used in the food, pharmaceutical, cosmetics, biofuel and pulp-and-paper industries.

Lipases have been classified into three classes according to the residues comprising the oxyanion hole: GGGX-, GX- and Y (where X can be any amino acid) (145). In this thesis I focus on the Y-lipase from the yeast *Moesziomyces antarcticus* (formerly *Candida antarctica*), Cal-A. It is interesting to note that the Y-lipase class was created to include this particular enzyme that possesses a different 3-D structure compared to other known lipases (145). This enzyme has

attracted more attention in recent years due to its high thermal stability, tolerance to organic solvents, stability at low pH and high substrate versatility, being one of the few lipases that can accept sterically-hindered tertiary alcohols and favor esterification of trans-fatty acids (144, 146).

1.7.2.1 Structural characterization of Cal-A lipase

Cal-A is a monomeric hydrolase composed of 431 amino acid residues that fold into three domains: the conserved α/ß hydrolase-fold domain with 8 ß-strands connected by 6 α-helices, a flap formed by two ß-strands and a cap domain formed by six consecutive α-helices (Figure 1.13).

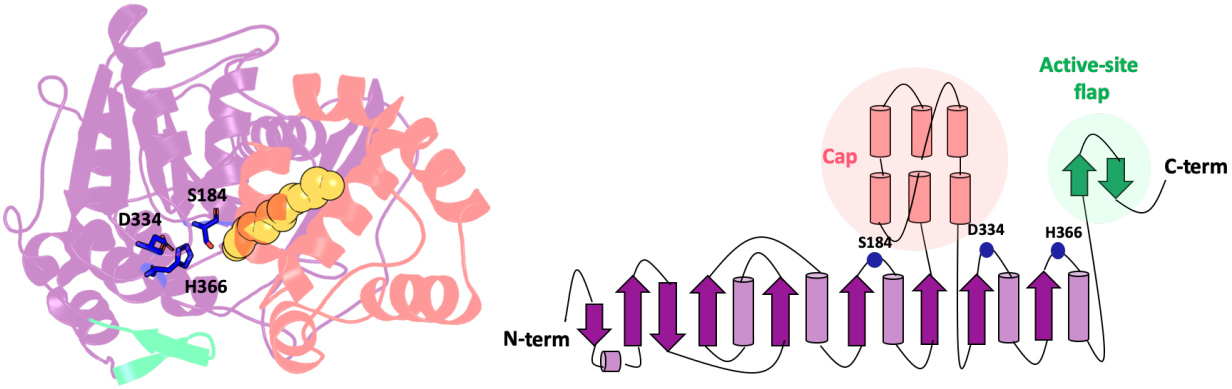


Figure 1.13. – Three-dimensional structure and diagram of Cal-A showing its main structural features.

The active site flap (green), cap (red), α/ß hydrolase-fold (purple) and the catalytic triad (blue)are highlighted in the 3D structure of Cal-A (PDB ID: 2veo). The putative acyl-binding tunnel is shown with the co-crystallized PEG molecule (yellow) (left panel). Adapted from Brundiek, Padhi (147). Figures created with PyMOL Molecular Graphics System, Version 2.1 Schrödinger, LLC and Microsoft Powerpoint 16.16.10.

Lipases belong to the α/ß hydrolase class along with esterases, cutinases and epoxide hydrolases. Enzymes of this class possess the characteristic α/ß hydrolase fold along with GXSXG or GXDXG motifs in the active site, a catalytic triad S(D or E)H and an oxyanion hole (148). Lipases have been classified according to the residues forming their oxyanion hole as GGGX- or GX-. Upon determination of the structure of Cal-A, it was revealed that Cal-A did not fit into any known lipase class as it had Y93 forming the oxyanion hole. Therefore, the new Y-class was created to accommodate Cal-A(145). To date only two crystallographic structures of Cal-A have been

deposited in the Protein Data Bank (PDB ID: 2veo and 3guu). Crystallographic data has also pointed to residues D95 and G185 contributing to the oxyanion hole as stabilizers of the acyl-enzyme tetrahedral intermediate (149).

When its structure was first determined in 2007 (PDB ID: 2veo), Cal-A was assigned to a new superfamily of lipases due to its distinctive structural features. Apart from its unique oxyanion hole configuration, it was the first time the cap domain was described in an enzyme within the α/ß hydrolase class (145). The cap domain is formed by residues 217-308 which are partly involved in forming the putative tunnel of Cal-A that allows binding of long-chain acyl moieties (149). The putative tunnel region was identified from a polyethylene glycol (PEG) molecule bound in the crystal structure and computational docking of *p*-nitrophenyl palmitate in the PEG bound region (149). Its ~30 Å-long acyl-binding tunnel can accommodate up to 20-25 methylene units in length (149).

Access to the active site and tunnel region is modulated by conformational changes in the lid domain in the presence of a hydrophobic layer, termed interfacial activation (149). This phenomenon is common to many lipases as these generally operate in lipid-water interfaces (150, 151). The lid domain in Cal-A, often referred to as a flap due to its small size, is formed by residues 426-436 (149, 152). The flap region is formed mainly by hydrophobic residues and is flanked by G426 and G436 that have been proposed to act as “hinges” (152). It should be noted that the cap domain is sometimes referred to as the lid domain as when the crystal structure was first described, it was thought that this region regulated access to the active site cavity (149).

1.7.2.2 Current knowledge on Cal-A reaction mechanism

Prior to deciphering the 3D structure of Cal-A, titration with an irreversible inhibitor and site-directed mutagenesis studies determined the catalytic triad of this enzyme to be formed by S184, D334 and H366 (153, 154). These results were confirmed by analysis of the positioning of the residues in its structure (149). While no study on the reaction mechanism of Cal-A reaction mechanism has been reported, the catalytic triad is presumed to operate as in other α/ß hydrolases (144). Briefly, the nucleophilic residue S184 would be activated by its neighboring H366 that can accept a proton from S184 due to the negative charge of the side-chain of its

neighboring residue D334 (Figure 1.14). The activated S184 then performs the nucleophilic attack on the ester bond of a triglyceride, forming an acyl-enzyme tetrahedral intermediate that is stabilized by oxyanion hole-forming residues D95 and G185. Subsequently, the diglyceride product leaves and the acyl-intermediate is deacylated with the aid of a catalytic water. It should be noted that, despite having been isolated from a cold-adapted organism, Cal-A presents its optimum activity at 50°C (155).

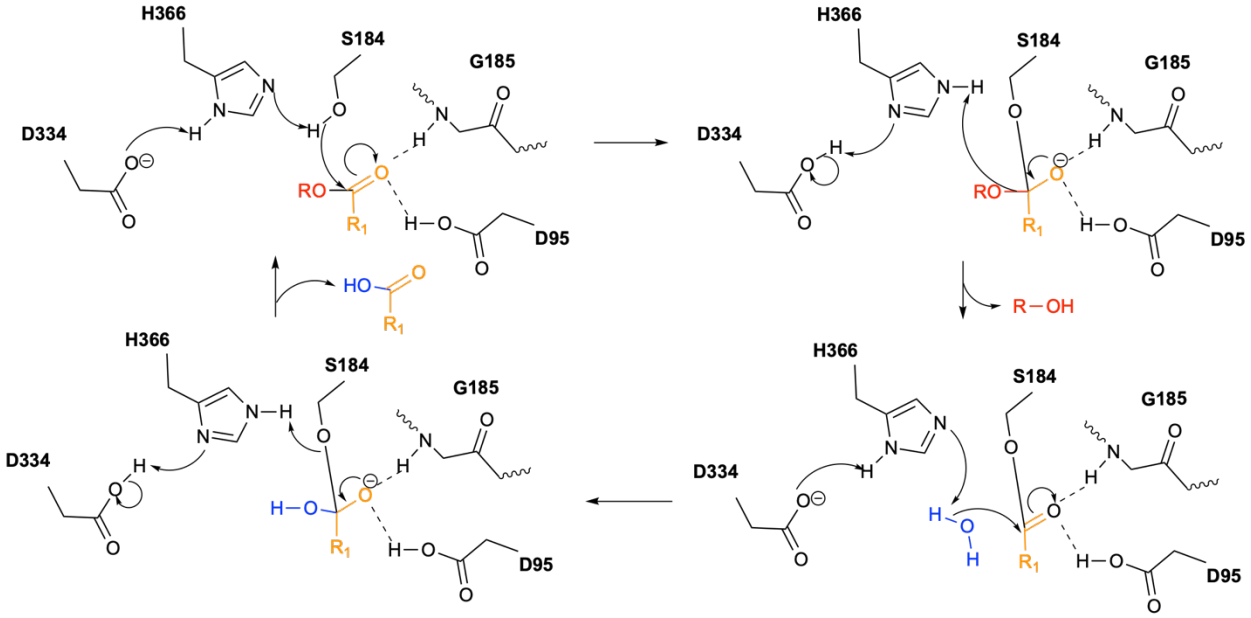


Figure 1.14. – Schematic representation of the reaction mechanism of the hydrolysis of an ester by Cal-A lipase.

Several steps typical of lipase catalyzed hydrolysis of ester are shown. First, the activation of the catalytic S184 and the subsequent nucleophilic attack to create the acyl-enzyme and liberate the carboxylic acid (red). Deacylation and regeneration of the enzyme is done with the aid of a catalytic water. Adapted from Monteiro, Virgen-Ortiz (144). Figure created in ChemDraw Professional 16.0.

1.7.2.3 Enzyme versatility

Cal-A has been shown to be a highly versatile enzyme capable of performing hydrolysis, ammonolysis, esterification, interesterification and transesterification reactions (144). It shows activities rarely observed in other lipases such as the hydrolysis of amide bonds in N-acylated α- and ß-amino acids (156) or the enantioselective N-acylation of ß-amino acids (144, 157).

It is considered a highly promiscuous lipase as it has been shown to hydrolyze 36 esters of different chemical nature (triglycerides, monoglycerides, aromatic) and chain-length (158). Regarding its natural substrates, triglycerides, Cal-A presents higher activity towards short-chain (C3-C4) than long-chain triglycerides (155, 158). This trend to preferably hydrolyze esters of fatty acids of C3-C4 lengths is also maintained in esters of other chemical nature such as vinyl esters or *p*-nitrophenyl esters (155, 158, 159). Interestingly, Cal-A has a long-acyl binding tunnel and was first proposed to bind preferably to long-chain fatty acids (149).

Other relevant substrate preferences of Cal-A are its preference for saturated fatty acids, trans fatty acids and fatty acids at the second position in triglycerides (*sn*-2 preference) (144, 146, 155, 160, 161). This preference towards *trans*-fatty acid over *cis*, shared by other enzymes belonging to the Cal-A family, is rarely observed in other lipases that contain an L-shape acyl-binding tunnel (161). The *sn*-2 preference is also quite singular, as most lipases show *sn*-1,3-regioselectivity. This feature is highly interesting from a biotechnological point of view to produce free fatty acids present at the *sn*-2 position of triglycerides from natural sources (155, 162).

In general, Cal-A shows high enantioselectivity and has been used to generate industrially relevant compounds (144, 161). Another relevant feature is its ability to accept highly sterically hindered alcohols such as tertiary alcohols and cyclic alcohols for hydrolysis and esterification reactions (161). Overall, Cal-A presents itself as a lipase with unusual substrate selectivity having wide applicability in industrial applications (144, 161). Its high thermal stability and solvent tolerance add further value to its use in biotechnology.

1.7.2.4 Directed evolution of Cal-A lipase

Due to its interest for potential use in biotechnological applications, Cal-A has been subjected to directed evolution to improve its selectivity. As such, Cal-A has been successfully engineered to improve its selectivity towards α-esters, trans-fatty acids or tertiary alcohols (144, 163). Contrary to TEM-1 ß-lactamase, few theoretical studies to understand its evolution or mechanism have been reported (152, 153).

The generation of Cal-A variants with improved enantiospecificity and enanstioselectivity is of great interest in the food and the pharmaceutical industries (143, 144, 161). In this thesis,

directed evolution of Cal-A has been performed to alter its chain-length selectivity. This property is of high interest in the food and pharmaceutical industries to enrich certain free fatty acids present in natural oils or to create diglycerides or monoglycerides that have been shown to have health benefits (143, 144, 147, 161, 162, 164). Certain mutations that increase selectivity towards short-chain fatty acids have even been patented by the related biotechnology companies Novo Nordisk and Novozymes (165, 166) and by the snack company Frito-Lay in collaboration with the research group of Dr. Uwe Bornscheuer in Germany (147, 167). Another study has revealed the existence of an synergistic interactions between mutations at the end of the acyl-binding tunnel that prevent the hydrolysis of *p*-nitrophenyl (*p*-NO_2) esters of long fatty acids (159). The position of previously identified short chain-selective variants in Cal-A is shown in Figure 1.15. Most of the mutations are found distal from the catalytic triad in the active site or in the cap domain that contains the putative acyl-binding tunnel. The fact that mutations distal to key regions have also been identified suggests the implication of long-range interactions. Concerning long-chain selectivity in Cal-A, there is limited information, despite its long-acyl chain binding tunnel. Prior research on other lipases has shown that long-chain fatty acid hydrolysis can be enhanced by physically widening the active site entrance (168, 169), favoring open conformations (170), improving lid flexibility (171, 172) or swapping the lid with another lipase (173, 174).

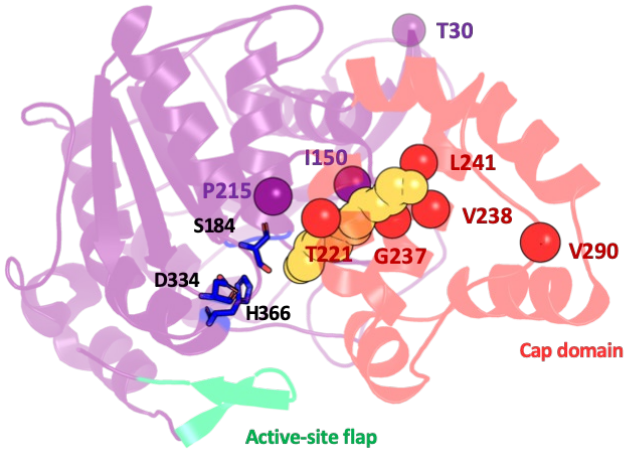


Figure 1.15. – Previously identified chain-length selective Cal-A mutations.

The 3D structure of Cal-A (PDB ID: 2veo) with key features highlighted: the flap in green, the cap in red, the catalytic triad as blue sticks and the co-crystallized PEG molecule within the putative tunnel in yellow. Previously identified positions of mutations that generate chain-

length selective phenotypes T30, I150, P215, T221, G237, V238, L241 and V290 are shown as spheres. Figure created with PyMOL Molecular Graphics System, Version 2.1 Schrödinger, LLC.

1.8 Description of the selected approaches and methodologies

The specific objectives of the thesis are based on the properties of each model system chosen to study the research questions pertaining to epistatic interactions in laboratory evolution. The approaches and methodologies chosen are justified based on the specific objectives, on structure-function knowledge and on screening methods available for each system.

1.8.1 Objective 1: Interplay of protein dynamics and epistasis on the evolution of new protein function

To address this objective, cTEM2m and cTEM17m variants derived from TEM-1 ß-lactamase constituted our model system. These variants exhibit increased protein dynamics at slow timescales in the all-α domain (cTEM2m) or in the all-α and α/ß domains (cTEM17m) compared to TEM-1. By using these models as start-points in evolution, we can dissect the impact of slow timescales (µs-ms) in the development of new protein function. Objective 1 was addressed by two distinct approaches described below. The study addressing Objective 1 was published in the journal Frontiers in Molecular Biosciences in 2020 and is presented as Chapter 3 of this thesis.

In the evolution and screening of evolved TEM-1 ß-lactamase variants, we used a ß-lactam antibiotic that has been used in directed molecular evolution and in clinical settings: cefotaxime. The extensive use of this antibiotic guarantees that we have prior information on mutations that give rise to this resistance in both natural and laboratory settings. Cefotaxime is commercially available and can be incorporated into agar plates containing growth media to easily screen for resistance variants. In addition, its extinction coefficient is known and can be used to perform kinetic characterization.

1.8.1.1 Objective 1.1: Epistasis and evolution in slow dynamic variants

Prevalent epistatic mutations in the natural and directed evolution of TEM-1 towards the antibiotic cefotaxime were introduced into cTEM2m and cTEM17m. The introduction of mutations E104K/G238S was made through site-directed mutagenesis using whole-plasmid site-

directed mutagenesis or site-directed overlap extension mutagenesis. Both methods are described in Chapter 2 of this thesis, within a critical overview of methods for mutagenic library building.

In order to understand epistasis upon the introduction of E104K/G238S in the slow dynamic variants both *in vivo* and *in vitro* characterization are of use. *In vivo* characterization allows to reveal epistatic effects while showing the mutational paths that are realistically possible. This was done by doing Minimal Inhibitory Concentration (MIC) assays that give a measure of the resistance conferred by the different ß-lactamase variants. *In vitro* characterization allows to decipher the basis for the epistatic effects observed. The analyses chosen for the *in vitro* characterization were the determination of thermal stability and kinetic parameters for ligand-binding and catalytic turnover. Thermal stability allows to evaluate If the loss of activity in MIC is due to loss of available protein to carry on the function. This was assayed by measuring the melting temperature. Kinetic characterization allows to evaluate the catalytic efficiency of the variants independently from other factors that might be affecting *in vivo* activity such as stability or expression. To rationalize the effects on catalytic efficiency in the different variants we employed flexible ligand docking. This approach was possible with the ligand used, as the rate-limiting step in cephalosporin hydrolysis is acylation, which occurs immediately after ligand-binding. Flexible ligand docking allows small angle rotations and translations in the ligands while exploring different conformations of the target protein.

1.8.1.2 Objective 1.2 Impact of slow timescale dynamics on the accessibility to evolutionary trajectories.

Directed molecular evolution was used to evaluate whether mutations accessed by the less dynamic TEM-1 were accessed by the slow dynamic cTEM2m and cTEM17m variants. Random mutagenesis was used to have an unbiased evaluation of the selected mutations. The *in vivo* activity using MIC assays was measured once we observed dominance of particular variants throughout the generations.

1.8.2 Objective 2: Characterization of protein epistasis in a biotechnologically relevant enzyme

To address this objective the industrially relevant enzyme Cal-A lipase was used. More specifically we focused on fatty acid chain-length selectivity. As less structural information and fewer directed evolution experiments are available to direct mutagenesis efforts in this system, Objective 2 was achieved according to the approaches defined in this section.

1.8.2.1 Objective 2.1: Identification of chain-length selectivity hot-spots

The results addressing Objective 2.1 were published in the journal Plos One in 2019 and are included as Chapter 4 of this thesis.

The sequence of Cal-A was divided into three segments and subjected to focused random mutagenesis. Random mutagenesis allows a blind exploration of the system. This mutagenesis approach coupled with colony screening on agar emulsions containing short-chain or long-chain triglycerides allowed identification of a hot-spot region where mutations modulated chain-length selectivity.

Direct selection methods on agar plates based on survival in the presence of specific triglycerides are not feasible using Cal-A. This justifies the use of focused random mutagenesis to reduce library size to reduce screening effort. The triglyceride plate screening consisted of spotting individual colonies onto agar plates containing triglyceride emulsions of different chain length. Active variants were isolated based on the formation of halos around colonies expressing an active Cal-A variant. If the emulsion is opaque, a clear halo is formed. When the emulsion is not opaque, a fluorescent pH marker is added to create a fluorescent halo. As the results obtained through this method are qualitative, we performed further tests using hydrolysis of *p*-NO_2-phenyl esters of fatty acids of different chain-lengths. This assay is the industry standard to quantify lipase activity *in vitro*: *p*-NO_2-phenol derivatives of fatty acids are highly activated, allowing rapid enzymatic hydrolysis and easy colorimetric detection (175).

Identification of a triglyceride binding hot-spot was done by correlating the observed phenotype (short or long-chain selectivity) to the genotype. As focused random mutagenesis yields a

considerable number of variants, a computational approach to correlate phenotype to genotype was used. Mapping of these correlations to the 3D structure allowed to uncover the key regions implicated in selectivity.

1.8.2.2 Objective 2.2 Characterization of chain-length selective variants to assess prevalence of epistasis

The results addressing Objective 2.2 are included in Chapter 5 of this thesis. They are subject of an invited article to the journal Frontiers in Molecular Biosciences.

Following identification of a hot-spot for chain-length selectivity, we sought to characterize well-performing chain-length-selective variants obtained. This allows to decipher whether the phenotype observed is due to additivity or epistasis. To do so we chose to characterize variants that hadn't been previously described in the literature. The variants were deconvoluted to analyze whether the single mutants were responsible for the observed variations in chain-length selectivity. Characterization of the variants was done as in Objective 2.1. In addition, we also used a lesser-known but promising high-throughput assay that employs the pH indicator phenol red to quantify hydrolysis of triglycerides *in vitro* (176). This allows to have a quantitative measure with the substrates used in the initial screening. Flexible docking and analysis of the potential tunnels available was used to rationalize the effect of the mutations on chain-length selectivity.

Chapter 2 – Methods for enzyme library creation: which one will you choose? A guide for novices and experts

Preface to Chapter 2

The success of laboratory evolution of enzymes depends on the mutagenesis approach chosen, as well as the screening methods available to detect activity. Focused mutagenesis is often sought to reduce screening efforts, as well as the cloning effort of creating a large library. The main limitation of this approach is that it needs prior knowledge of the biophysical and structural properties of the system to decide which region to target. In addition, to perform a focused mutagenesis approach, one needs to identify the methodology that best suits their goal, as well as the degree of technical expertise required. In this Chapter, we review established methodologies in the enzyme engineering field, as well as recently reported improvements. This review is aimed at novice and experienced users as an instrument to help them decipher which methodology is the best match based on their research questions and molecular biology expertise.

This is an invited review for the Methods, Models & Techniques rubric of the journal BioEssays. My main contribution to this review lies in the description of established methods in enzyme engineering and their recent development, as well as the identification of popular commercial kits to easily implement these methodologies. The introductory section in this review has been described by Dr. Daniela Quaglia with some contributions from my part. I have also prepared the figures. The original draft was written by Dr. Quaglia and myself. Prof. Joelle N. Pelletier contributed to the review and editing of the final draft.

Manuscript in preparation - Article 1. Methods for enzyme library creation: which one will you choose? A guide for novices and experts to introduce genetic diversity

Lorea Alejaldre[1,2] Joelle N. Pelletier[1,2,3], Daniela Quaglia[3,4]*

[1] Biochemistry Department, Université de Montréal, Montréal, Québec, Canada

[2] PROTEO, The Québec Network for Research on Protein Function, Engineering and Applications, Québec, QC, Canada G1V 0A6

[3]Chemistry Department and Center in Green Chemistry and Catalysis (CGCC), Université de Montréal, Montréal, QC, Canada H3T 1J4

[4] School of Chemistry, University of Nottingham, NG7 2RD, United Kingdom

***Correspondence:**

Daniela Quaglia

daniela.quaglia1@nottingham.ac.uk

Invited review BioEssays

2.1. Summary

Enzyme engineering allows exploring sequence diversity in the search for new properties. The scientific literature is populated with methods to create enzyme libraries for engineering purposes, however, choosing a suitable method for the creation of mutant libraries can be daunting, in particular for the novices. The majority of the methods are reported in the context of a specific problem, often in terms suited to experts. Here, we address both novices and experts: how can one enter the arena of enzyme library design and what guidelines can advanced users apply to select strategies best suited to their purpose?

The novice will find an easy-to-consult segment to quickly understand which strategy will get the job done painlessly. While the expert will find recent improvements to already established methods. In addition, we point to commercialized and proprietary methods that will allow novice and experienced users to speed up sequence space exploration. We focus primarily on *in vitro* methods, presenting the advantages of each method. Our ultimate aim is to offer a selection of methods/strategies that we believe to be most useful to the enzyme engineer, whether a first-timer or a seasoned user.

2.2. Introduction

Theoretically, the number of combinations of mutations that can be made in a protein of 100 amino acids works out to more than 10^{130} possible combinations. If we consider that the largest screened libraries contain approximately 10^{15} members, (1) it is clear that only an infinitesimal part of the vast theoretical sequence space is accessible through experiments and will remain so even as automation increases screening capacity. For this reason, researchers have focused on improving library quality because the 'smarter' the library, the greater the odds of identifying the improved variants you are looking for (2-6). From the perspective of screening capacity, we consider 'small' a library of less than 10^3 members, 'medium-sided' one containing between 10^3 to 10^5, and 'large' when the library contains more than 10^5 members. The ability to generate a library that is diverse, while maintain theoretical library size as small as possible so that it can be more thoroughly screened with currently available methods increases the chance of identifying functional variants (7).

Mutagenesis is used to introduce diversity and navigate the sequence space by modifying the wild-type DNA through substitutions, insertions and deletions. Different exploration strategies are available: (i) random mutagenesis; (ii) recombination/gene shuffling and (iii) targeted approaches. (5, 8-15).

Methods based on random mutagenesis generally make use of an error-prone polymerase in a PCR reaction (epPCR) to introduce mutations at random positions within the coding sequence. This is most useful when the structure or the functional regions of the protein we seek to alter are unknown (8, 12). However, random mutagenesis requires a powerful high-throughput screening method to isolate improved variants because beneficial mutations occur 2 to 3 orders of magnitude less frequently than do mutations that are deleterious to function (9, 16). Furthermore, randomized libraries severely limit the introduction of consecutive mutations in the context of a single round; even with the most modern of approaches, a single round of random mutagenesis seldom suffices to provide significant improvement (11, 17, 18).

Recombination methodologies involve fragmenting mutated copies of the gene of interest and reassembling them to give rise to new combinations of mutations (8). Arguably the most popular application of recombination is SCHEMA, which involves the guided recombination of fragments of homologous genes from the same family, giving rise to chimeric proteins (19). SCHEMA is particularly of interest because it is characterized by somewhat smaller theoretical library sizes (~10^3-10^4 variants) than other recombination methods (20-23) while generating diverse libraries with a higher proportion of functional variants than randomization methods (8, 24).

While these options are powerful and widely adopted, our review emphasizes more focused approaches. They involve targeting specific residues or regions for mutation to one or many possibilities, often in a combinatorial manner, or they might imply the use of insertions and deletions in the gene sequence (3, 18, 25-28).

Targeted mutagenesis is limited by the requirement for structural or biophysical knowledge of the enzyme. To address this limitation, it is common to perform exploratory rounds of randomization that inform subsequent rounds of targeted mutagenesis: Frances Arnold once stated that bringing together rational design and directed evolution is the way forward to the

generation of quality libraries containing a high number of sequences with the desired properties (4, 27, 29, 30). A targeted strategy might include randomization of certain parts of a gene, but not the whole sequence (10, 12, 17, 18, 31).

Targeted mutagenesis can produce smaller theoretical libraries (~10-10^4) that may hold a high proportion of functional variants, which in turn translates in a lower screening effort (11, 17). This has shown to be especially useful in improving specific properties of an enzyme: activity, selectivity and thermal resistance (8).

Although of limited scope for sequence diversification, the simplicity of targeted approaches has its benefits, including limiting the number of variants to be screened to find an improved enzyme.

The most basic approach for the generation of targeted mutations is site-directed mutagenesis (SDM), introduced by Nobel laureate Michael Smith to create a point mutation (32). Its derivatives include site-saturation mutagenesis (SSM), discussed in **section I-A**, for introduction of degenerate codons at a specific position (9, 33-35), and combinatorial saturation mutagenesis (CSM, **section I-B**) for mutating more than one site at a time.

Single site saturation libraries: Saturation libraries offer high quality and small size while giving access to greater sequence diversity than do point mutations. However, building saturation libraries one position at a time, severely limits exploration of sequence diversity; in particular, it curtails the discovery of positive epistatic effects, where specific combinations of mutations (whether proximal or distal) can give a greater improvement than could be expected from combining effects of the point mutants (36, 37). These complex interactions may result from direct interactions between residues or complex interactions involving conformational dynamics or stability (37). Epistasis is challenging to foresee, limiting predictability in the exploration of the sequence space (38, 39).

Combinatorial saturation libraries: to address the inaccessibility of epistasis in single site saturation libraries, combinatorial approaches can be useful. In this case, more sites are mutated at once and the combination of possible mutations are assessed. Examples of strategies of this kind are ISM (Iterative Saturation Mutagenesis) and derivatives.

In ISM, sites of interests are chosen (as many as necessary) that generally contain 1 to 3 amino acids (or more). These are then randomized individually using saturation mutagenesis. The best hits selected from the first randomization are subjected to further iterations of randomization with saturation mutagenesis at the other positions (5, 9, 28, 39, 40). ISM allows for a *fast convergence of beneficial mutations*, mainly because the choice of positions to mutate is informed by prior structural or biochemical knowledge (41). The choice of position to mutate are chosen depending on the final goal (i.e. B-FIT for a gain in thermostability (41), CASTing for improvement of catalytic properties (28)(35)). A novel iterative methodology (FRISM), inspired by CAST/ISM, includes a rational prediction aspect that restricts the number of mutants to be screened. FRISM has been proven useful for engineering stereoselectivity in enzyme variants (5, 42, 43).

Below, we review a broad selection of methods and strategies to produce the targeted library that best suits your needs. **Section I** presents an overview of more established approaches for standard library creation and popular commercial solutions to speed up the creation of targeted libraries. **Section II** presents popular commercial solutions to speed up the creation of targeted libraries

2.3. Section I: Established methods for site-directed and saturation mutagenesis for targeted library creation

Site-directed mutagenesis (SDM) allows an amino acid in the target protein to be substituted using mutagenic primers. This is easily achieved using commercially available kits such as QuikChange by Agilent (discussed in **section I-B 7**). A direct derivative from SDM is site-saturation mutagenesis, commonly referred to as SSM, for mutation at a single site, or CSM (Combinatorial Saturation Mutagenesis) for more than one randomization site (8, 9, 33, 35, 36). In SSM and CSM, degenerate mutagenic primers introduce the nineteen other amino acids (or a defined subset of them) at each site.

The main advantage of saturation mutagenesis is that it allows for non-conservative codon substitutions that are unlikely to arise by random mutagenesis, giving access to "*non-natural evolution pathways"* (44).

Direct improvement of saturation mutagenesis entails the use of primers that reduce codon degeneracy to tune it to the needs of the user (36). For instance, one can restrict codon redundancy or the presence of stop codons (8, 33, 45). The twenty possible naturally occurring amino acids are encoded by NNN degeneracy that codes for 64 possible codons (including 3 stop codons), where N is either A, T, G or C. The NNK degeneracy (K = T or G, refer to the IUPAC nomenclature for standard letters of degeneracy (46)) limits the number of codons to 32, while encoding all possible amino acids and only 1 stop codon to reduce screening efforts. Although the library size is indeed reduced compared to NNK if a single position in the gene is mutated, it is important to remember that when saturation mutagenesis is used in combinatorial strategy, this might not be the case as the redundancy and library size increase considerably, as it is a nonlinear relationship (i.e. for 1 position mutated with NNK, ~34% of hits are redundant, yet for 4 position, ~81% of the hits are redundant, with a library size of $10^{\wedge 6}$).

If the user is instead willing to sacrifice some of the possible amino acids, NNT and NNG code for 16 possibilities, while NDT (D = A, G, T) provides 12 codons, for an even smaller library (36). More recently, Tang et al. (45) devised a method called SILM (Small Intelligent Library Method) (8) as an alternative to NNK that includes all possible amino acids but reduces degeneracy. In SILM, four primers (with NDT, VMA, ATG and TGG, where V = A, C or G and M = A or C at the site to be mutated) (45, 47) are mixed at a specified ratio. Another method, called 22c-trick reduces degeneracy by combining three primers (NDT, VHG and TGG, where H= A, C, T) that contain 22 codons that code for 20 aminoacids (48). For an ever-tighter control over codon degeneracy, several tools are available to the users, such as: the DC-ANALYZER (8, 45), MDC-Analyzer (49), SwiftLib (50) or DeCoDe (51).

Many methods have been developed to introduce mutations in the targeted gene at specific positions. Here we describe a selection of the most popular methods employed in the field of enzyme engineering, in order of increasing complexity. We also mention methods that are less commonly used but that could provide advantages for users having more expertise in molecular biology or having specific requirements. Most of the methods described in this section can be adapted to perform site-directed mutagenesis (SDM), site-saturation mutagenesis (SSM) or combinatorial site-saturation mutagenesis (CSM). It is important to remember that the efficiency

of mutagenesis is often dependent on the availability of easy-to-use cloning and assembly methods, the most popular of which we describe in **section I-B**.

A. Established methods for SDM and SSM

The whole plasmid approach to site-directed mutagenesis is the most popular method for substituting one or a few consecutive amino acids in a target enzyme (Table 2.1). First commercialized by Stratagene in 1996 as the QuikChange kit (see **section I-B 7** and **section II**), the standard protocol relies on the design of overlapping primers containing the desired mutagenic codon(s) (Figure 2.1) (52). These are used to amplify the whole plasmid with a high-fidelity DNA polymerase. The resulting PCR product is a mutated and nicked double-stranded DNA. Treatment with *Dpn*I enzyme eliminates non-mutated methylated wild-type DNA template; transformation into a bacterial host allows repair of the nicked, mutated strand with the recombination machinery of the host. Although straightforward, this results in poor transformation efficiency due to nicked DNA; furthermore wild-type contamination resulting from incomplete *Dpn*I digestion has a higher transformation efficiency, which can increase the cost of identifying mutated variants.

These limitations make the standard method unsuited to the generation of larger libraries (>10^4 transformants, as also reported in the QuikChange Site Directed Mutagenesis Manuals). Other drawbacks include annealing of the complementary primers to each other instead of to the target plasmid (primer-dimers) and poor PCR amplification. Nonetheless, the whole-plasmid approach remains one of the fastest and simplest ways to introduce mutations, making it attractive to novice as well as experienced users. To enhance its benefits, numerous improvements have been made concerning primer design (53-55), incorporating DNA insertions or deletions (56), enhancing transformation efficiency and including mutations at multiple sites (multi-site directed mutagenesis) (56, 57). A recent, noteworthy improvement for difficult targets (large plasmid or difficult to deal with due to the presence of stable secondary structures) has also been proposed by Li & al. (58).

2.Whole-plasmid site-directed mutagenesis with template/product modification for improved mutational efficiency

Whole-plasmid site-directed mutagenesis using template/product modification are not popular amongst novices, as the protocols are lengthier than with other methods. However, they have proven useful to reduce wild-type template contamination, offering a mutational efficiency of 50-100% (59, 60).

The Kunkel method and its derivatives (61) require the generation of whole-plasmid single-strand template DNA containing dUTP instead of dTTP. This often implies using a phage to infect *E. coli* cells defective in dUTPase (dut^-) and uracil N-glycosylase (ung^-). The lack of these enzymatic activities allows the cells to propagate the phasmid DNA containing the target gene with uracil residues. Synthesis of a complementary strand using a mutated primer restores dTTP. Upon transformation into bacteria, the original dUTP-containing strands are cleaved, eliminating the wild-type template.

Although the basic method does not involve PCR, a remarkably improved protocol relies on the use of a high-fidelity polymerase which prevents uracil stalling, *Taq* ligase, *in vitro* uracil DNA glycosylase and exonuclease III to cleave the uracil-containing template and non-desired DNA products and transform double-stranded DNA with higher transformation efficiency (59). This protocol gives further accessibility to the Kunkel method by providing a description to generate uracil containing templates without the need to use phages (59). Other studies to improve the mutation efficiency of Kunkel-based methods have focused on improving conditions for DNA amplification such as primer design, extension time and annealing temperature (62) or the use of phi29 polymerase (63).

Oligonucleotide-directed mutagenesis by elimination of unique restriction site (USE) (60, 64) also relies on whole-plasmid amplification of DNA. It employs a mutagenic primer and a primer that eliminates a unique restriction site in the mutated plasmid. Subsequent digestion with the corresponding restriction enzyme eliminates wild-type template. This product is transformed into a mutS *E. coli* strain deficient in repair of mismatched bases. The obtained colonies are pooled, the plasmid DNA is obtained and digested with the same restriction enzyme to ensure

complete elimination of wild-type template, and transformed in a standard *E. coli* strain. In conjunction with the Kunkel-based methods, USE can serve to increase mutational efficiency (65).

3. Site-directed mutagenesis by overlap extension for large plasmids or difficult constructs

First described in 1989 as a PCR-based method to introduce mutations, insertions and deletions (66), the main advantages of Site-directed mutagenesis by Overlap Extension (SOE) are high product yield and traceability of product formation. It is mostly used for large plasmids, difficult-to-amplify genes or when whole-plasmid amplification consistently fails due to other reasons (*i.e*. formation of primer dimers). In SOE, overlapping primers containing the mutation(s) are designed, similarly to the commercial QuikChange method (see **section I-B 7)**. These are used separately in two PCR reactions to introduce the mutation(s), paired with a primer complementary to either the 5' or 3' terminus of the gene (Figure 2.1). A final PCR reaction combining the two mutated fragments and extending to full length with the terminal primers then recreates the whole, mutated gene with efficiencies of up to 98% (66). An improvement of this method allows one to carry it out in a single tube: in this instance, a megaprimer carrying the mutation of interest is generated using one mutagenic primer and a non-mutagenic one that starts at either of the gene's termini. The megaprimer is then used to amplify the whole gene in a second PCR step (67).

A variation of SOE for more efficient removal of impurities such as template DNA or excess primers is PAGE-mediated Overlap Extension PCR (POEP) (68). This modification increases mutational efficiency, especially for multiple-site insertions where efficiency of SOE is near 75% (and reaching 100% for POEP) (68). Improvements for creation of insertions and deletions >30 bp have also been described (69), as well as multi-site directed mutagenesis by creating mutagenic fragments with homologous regions that act as megaprimers (70-72).

The main limitation of SOE is the general requirement for subsequent subcloning of the mutated DNA fragment in the desired vector. Homologous recombination *in vitro* and *in vivo* can alleviate this limitation and will be discussed in **section I-B 6 and 7**. Commercialized methods for this purpose are listed in **section II**.

4. Inverse PCR for the introduction of deletions and insertions

Developed in 1989 (73), this approach is particularly advantageous for introducing deletions and insertions, and for saturation mutagenesis (74). Inverse PCR relies on the amplification of a targeted sequence using back-to-back primers for outward amplification of regions flanking the targeted sequence (Figure 1). Amplification generates a linear product that is subsequently ligated to regenerate the whole plasmid containing the target gene. Ligation requires a phosphate group at the 5' terminus; this can be achieved with T4 polynucleotide kinase or by use of phosphorylated primers. As with whole-plasmid amplification, digestion of the methylated template DNA with *Dpn*I prior to transformation increases transformation efficiency. An improvement of this method consists in generating complementary overhangs for a ligation-independent approach (SLIM) (75).

Table 2.1. – Quick overview of selected established mutagenesis methods

Category	Method	Mutagenesis efficiency[a]	Estimated time (h) [b]	Enzymes required	Sub-cloning required	Recommended protocol
Whole-plasmid site-directed mutagenesis		Less than 100%	~20	▪ High-fidelity polymerase ▪ *Dpn*I	No	Laible and Boonrod (76)
Whole plasmid site-directed mutagenesis with template/product modification	Pfunkel[c]	70-100%	~20	▪ T4 Polynucleotide kinase ▪ High-fidelity polymerase that reads uracil-containing DNA ▪ *Taq* Ligase ▪ Uracil DNA glycosylase ▪ Exonuclease III	No	Firnberg and Ostermeier (59)
	USE[d]	50-90%	~40	▪ T4/T7 DNA polymerase ▪ T4 DNA ligase ▪ Restriction enzyme	No	Forloni, Liu and Wajapeyee (60)
Site-directed mutagenesis by overlap extension	SOE	>98%	~24	▪ High-fidelity polymerase ▪ *Dpn*I (optional)	Yes	Sambrook and Russell (77)
Inverse PCR	SLIM	93%	~20	▪ *Taq* DNA polymerase ▪ High-fidelity polymerase ▪ *Dpn*I	No	Chiu, March, Lee and Tillett (75)

[a]Mutagenesis efficiency is as calculated in each of the recommended protocols

[b] Estimated time includes transformation and overnight culture.

[c] dut^-/ung^-: *E. coli* strain required to generate uracil-containing template

[d] mutS: *E. coli* strain required to avoid repair of the mutated template

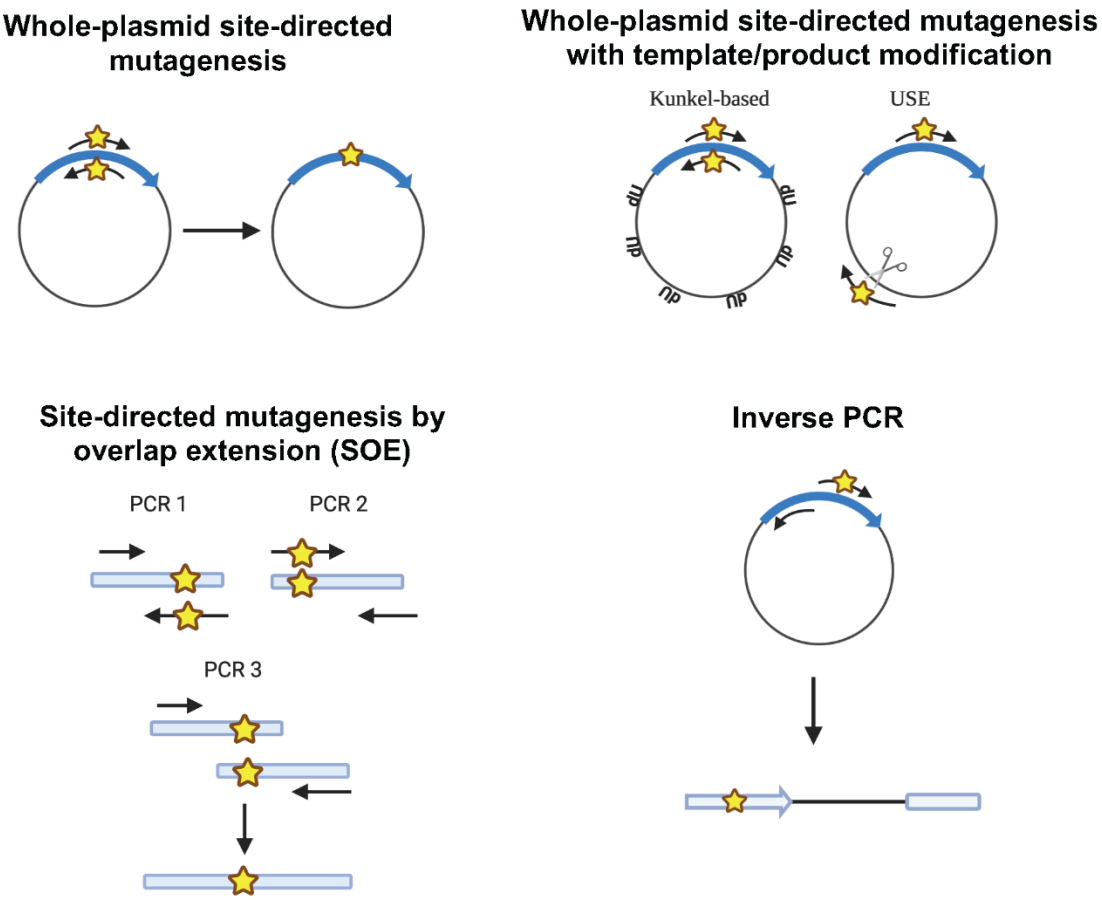


Figure 2.1. – Representation of PCR-based methods described in Section I-A

*Schematic representation of methods described in **section I-A**. For simplicity schematic representations correspond to double-stranded DNA except for Kunkel-based and USE methodologies where single-stranded DNA is represented. Mutagenic primers and mutated sites are represented with a yellow star, whereas restriction enzymes are represented with scissors. Uracil is indicated by dU on the plasmid. Created in BioRender.com.*

B. Flexible and easy-to-use assembly/cloning strategies instrumental to mutagenesis

The development of straightforward cloning and assembly methodologies in conjunction with site-saturation mutagenesis methods has accelerated the creation of 'smarter' methods for library generation. Sub-cloning of fragments into vectors constitutes a major bottleneck in library generation. This is reflected in the popularity of the QuikChange kit (**section I-B 7** and **section II)** that requires no restriction/ligation step. Ligation-independent methods have provided noteworthy solutions to this problem, accelerating library creation to explore sequence space more efficiently. The major advantage of ligation-independent methods resides in providing

increased flexibility in library generation. Mutations introduced in different parts of the genes can subsequently be reassembled to explore interactions between selected mutations or within a defined region in the protein (*i.e*. a tunnel, the *C*-terminus, etc.) (26). A few *ex vivo* and *in vivo* methods will also be mentioned briefly.

1. Restriction-free cloning for versatility

Restriction-Free (RF) cloning can be used to accelerate single or multiple site-directed mutagenesis, insertion and/or deletions (78). It consists in amplifying the vector using high-fidelity PCR and an insert (megaprimer) containing regions homologous to the vector in its 3' and 5' termini (Figure 2.2) (78, 79). This method is versatile as any vector/insert pair can be used as long as the primer design is carefully performed (79). Digestion with *Dpn*I prior to transformation is used to eliminate wild-type template DNA. However, gene replacement of a lethal gene in the original vector can eliminate the need for digestion with *Dpn*I (80). The main disadvantages of RF cloning include low product yield and low efficiency for large insertions. To circumvent these limitations, Exponential Megapriming PCR (EMP) adds forward or reverse primers to the megaprimer reaction, which contains a single overlapping region to the vector, to amplify the whole plasmid (81). A recent improvement of RF cloning, Modified Restriction-Free cloning (MRF) (82) allows insertion of fragments of up to 20 kb into the cloning vector.

2. Multiple overlap extension PCR for fragment assembly

Multiple Overlap Extension PCR (MOE-PCR) can be used to assemble up to eight DNA fragments in a single PCR reaction without the need for additional enzymes (T4 DNA ligase, exonuclease, etc…) (70). In MOE, fragments containing the desired mutations and 50 bp homologous to the contiguous fragment are produced in independent PCR reactions. These are subsequently mixed in a single PCR reaction and used as a template for the assembly of the whole gene or plasmid in a single step (Figure 2). Careful design of the annealing temperature of termini is needed for specificity and efficiency. A previous version of MOE-PCR, called multi-fragment site-directed mutagenic overlap extension PCR, applies the same principle of homologous

recombination of fragments termini and amplification: by assembling fragments two by two it can be used to efficiently assemble 13 fragments (72).

3. Combinatorial Codon Mutagenesis based on SOE and megaprimer PCR for tunable mutational frequency

The concept of Combinatorial Codon Mutagenesis (CCM) (83) is based on SOE-PCR and other recent work (84-86). This method allows for the generation of mutant libraries containing between one and seven codon mutations per gene. Two parallel PCR reactions are performed, which contain either a forward primer binding at 5′ terminus of the target gene and several mutagenic reverse primers, or a reverse primer binding at 3′ terminus and several mutagenic forward primers. These parallel amplifications generate numerous fragments containing combinations of the mutations introduced in the same PCR reaction (Figure 2.2). Analogous to SOE, a subsequent PCR containing forward and reverse primers and the generated fragments acting as megaprimers regenerates the whole gene length. The whole gene generated contains different combinations of the mutagenic codons introduced. This method has the advantage of easy "tunability" of mutational frequency as it is a function of fragmentation PCR cycle number (more cycles produce more mutations) and '*the number of rounds of fragmentation and joining PCRs conducted.'*

4. Ligation-independent cloning (LIC) and derivatives (SLIC and SLiCE) to ligate inserts into vectors without the need of a ligase

Ligation-Independent Cloning (LIC) allows assembly of insert(s) into vector without ligase as the overhangs created are completely complementary. LIC takes advantage of the exonuclease activity of T4 DNA polymerase to create overhangs in homologous regions between the desired mutated insert(s) and vector (87). This exonuclease activity is favored in the absence of dNTPs and stops when adding specific dNTPs.

Sequence and Ligation-Independent Cloning (SLIC) (88) (Figure 2.2), a modification of LIC, allows the recombination of up to 10 fragments. In SLIC, imperfect overhangs are created with T4 DNA polymerase and RecA protein is used to catalyze homologous recombination *in vitro* prior to transformation where the cell endogenous repair machine finishes recombination.

Recombination *ex vivo* using cell lysates of bacteria instead of the RecA protein is possible with a further modification: SLiCE (Seamless Ligation Cloning Extract) (89). Its application to enzyme engineering (90) will be described in the *in vivo* and *ex vivo* homology-based recombination methods **section I B-6**).

The main disadvantage of creating overhangs with T4 DNA polymerase is that ssDNA overhangs created may have strong secondary structures. A recent improvement of this method to recombine short fragments aims at reducing this disadvantage by optimizing incubation temperature and time with T4 DNA polymerase (91). A recent variation of SLIC consists in the use of T5 exonuclease instead of T4 DNA polymerase and its transformation in *E. coli* for assembly (92).

5. OmniChange for simultaneous mutations in distal parts of the gene

OmniChange can be used to mutagenize up to 5 codons simultaneously in up to the same number of fragments. As a first step, fragments are created via PCR amplification using forward primers and reverse mutagenic primers, or vice versa. Digestion with *Dpn*I prevents contamination with wild-type template DNA. The mutated fragments are cleaved with an iodine treatment to generate cohesive ends for subsequent reassembly (Figure 2.2). Nicks are repaired *in vivo* after transformation in *E. coli* BL21-Gold (DE3) $lacI^{Q1}$ (common laboratory strains can be used).

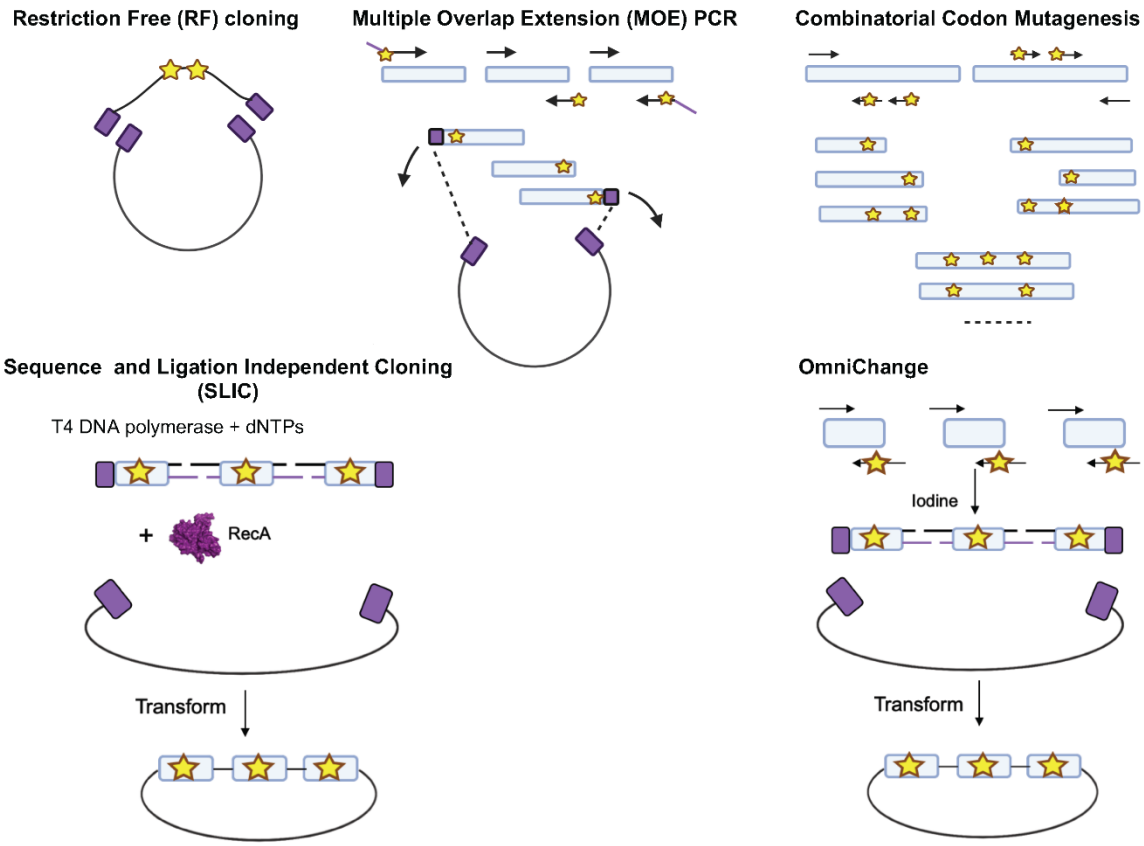


Figure 2.2. – Representation of assembly/cloning strategies described in Section I-B

Mutations are represented by yellow stars and overlapping segments as purple squares. Figure inspired by graphical descriptions from references (70, 79, 83, 88, 93). For simplicity, schematic diagrams represent double-stranded DNA. Created in BioRender.com.

6. In vivo and ex vivo homology-based recombination

Several methods described above generate nicked products (*i.e.* whole-plasmid amplification and OmniChange) that are repaired *in vivo* by common laboratory strains of *E. coli*. We deem it useful to briefly enumerate selected examples of easy-to-implement *in vivo and ex vivo* (cell lysate) homology-based recombination methods (Table 2). The methods use either *E. coli* or *S. cerevisiae*, the most common organisms for the heterologous expression and directed molecular evolution of prokaryotic and eukaryotic enzymes. We selected two that take advantage of the recA-independent recombination pathway of *E. coli* present in common cloning strains such as DH5α, XL10-Gold or Stbl3: SLiP (*in vivo*) (89, 90) and IVA cloning (*ex vivo*) (94). REPLACR mutagenesis (95) instead uses bacteria expressing Red/ET with recA-recombineering activity whereas IVOE (96)

uses *S. cerevisiae*. These approaches can be also applied by more experienced users to avoid the cost of commercial kits.

Table 2.2. – Selected *in vivo* and *ex vivo* homology-based recombination methods

Method	Short Description	Organism
SLiP - SLiCE-mediated PCR-based site-directed mutagenesis (89, 90)	Complementary to SOE, the target gene is amplified in two separate PCR reactions, generating two gene fragments containing the desired mutation(s) and short (15-20 bp) overlapping regions to each other and the receiving vector. Assembly of the linearized vector and the mutated fragment is done *ex vivo* by using SLiCE extracts.	*E. coli* lysate using the recA-independent recombination pathway (SLiCE extract) (common cloning strains, such as DH5α, XL10-Gold or XL1-Blue).
IVA cloning – *In Vivo* Assembly (94)	Fragments containing short (15-20 bp) 5' and 3' overlaps are directly transformed for *in vivo* assembly. Used to do assembly, mutagenesis, insertions and deletions.	*E. coli* cells using the recA-independent recombination pathway (common cloning strains, such as DH5α, XL10-Gold or XL1-Blue).
IVOE - *In Vivo* Overlap Extension (96)	Complementary to SOE, the target gene is amplified in two separate PCR reactions, generating two gene fragments containing the desired mutation(s) and 40-66 bp 5' and 3' overlaps with each other and the linearized receiving vector. These are subsequently transformed for	Protease-deficient *S. cerevisiae* strain BJ 5465, ATCC 208289.

	in vivo assembly to regenerate the full gene-length and sub-clone to the vector.	
REPLACR mutagenesis - Recombineering of Ends of linearised PLAsmids after PCR (95)	The vector is linearized by inverse PCR with primers designed to generate short (~17bp) overlaps at 5' and 3'. The linear PCR product is directly transformed for assembly *in vivo*. The method can be used to do site-directed mutagenesis, insertions and deletions.	Bacteria expressing Red/ET with recA-recombineering activity.

2.2. Section II: Commercial and proprietary methods to facilitate mutagenesis

Several kits are available for mutagenesis and assembly. They can be useful to novices for an easy start or as a standard quick alternative for experienced users. We present a selection of the most widely used kits in Table 2.3.

Until recently, gene synthesis was not broadly applied to generate libraries of very high quality (namely minimum biases). Nowadays the cost of gene synthesis and its quality have improved making this strategy more accessible. For instance, Twist Bioscience has recently developed a proprietary method for the synthesis of combinatorial libraries that they refer to as '*nearing perfection'* making commercial gene libraries even more accessible. Their synthesis capability was successfully validated by Li and colleagues, demonstrating how a high-quality combinatorial saturation mutagenesis library can be obtained through high-fidelity solid-phase chemical gene synthesis on silicon chips coupled with gene assembly (97). Their method has shown higher mutational efficiency, fewer incomplete sequences and less wild-type contamination than library generation through traditional saturation mutagenesis methods (97).

The same laboratory recently presented a second example (98, 99) of how commercially available methods (in this instance from LabGenius) can be of aid in the generation of smart libraries. Their method relies on previously developed techniques, in particular DNA assembly and on USER (Uracil-Specific Excision Reagent) cloning and it allows more accurate gene synthesis of '*long double stranded polynucleotides'* (99, 100).

While the use of synthetic libraries is no substitute for optimal library design, the steadily decreasing cost of gene synthesis definitely makes it a viable complement to smart library design strategies. Other companies that offer library generation services include ATUM, Codexis, Creative Biogene, Creative Biostructure and ThermoFisher.

Table 2.3. – Selected commercial mutagenesis kits.

Goal	Product name	Company	Cost per reaction [a]	Estimate dTime [b] (h)	Number of transformants (t) or fragments (f)
Site directed Mutagenesis	QuikChange® Lighting Site-Directed Mutagenesis kit	Agilent	$	20	t: 10^2-10^4
	QuikChange® Lighting Multi Site-Directed Mutagenesis kit	Agilent	$$$	20	t: 10-10^3
	Q5® Site directed Mutagenesis kit	New England Biolabs	$	20	t: 10^4
	GeneArt™ Site-Directed Mutagenesis System	ThermoFisher	$$	22	n/a [a]
Random mutagenesis	Genemorph II Random Mutagenesis kit	Agilent	$	22	t: 10^3-10^6 [d]
	Diversify™ PCR Random Mutagenesis kit	Takara	$	22	t: 10^3-10^6 [d]
Ligation-independent	GenBuilder™ DNA Assembly cloning kits	GenScript	$	17	t: n/a[c] / f: 1 to 12

commercialized homology-based recombination methods	NEBuilder® HiFi DNA assembly cloning kit	New England Biolabs	$	19	t: 10^3- 10^4/ f: (1 to 11)
	NEB Gibson Assembly® Cloning kits	New England Biolabs	$	19	t: 10^3-10^4 / f:(1 to 6)
	NEB® Golden Gate Assembly kits	New England Biolabs	$	19	t: 10^3-10^6/ f: (1 to 24)
	In-Fusion® HD cloning kit	Takara	$	17	t: 10^2- 10^3/ f: (1 to 2)
	GeneArt Type IIs Assembly kits	ThermoFisher	$$	20	t:10^2-10^5/ f: (1 to 8)
	GeneArt™ Gibson Assembly EX cloning kit	ThermoFisher	$	20	t: 10^2-10^3/ f: (1 to 15)
	GeneArt™ Gibson Assembly HiFi cloning kit	ThermoFisher	$	19	t:10^3/ f: (1 to 6)

[a] Cost per reaction based on the average of the formats available and compared within each category. Prices vary from 5-60 USD per reaction.

[b] Estimated time includes transformation and overnight culture.

[c] n/a Information not available.

[d] Depending on cloning efficiency.

2.3. Conclusions

Identifying a suitable method for the creation of a library for enzyme engineering is not as straightforward as it may seem, in particular for novices. A plethora of established and more recent methods described in the literature make it difficult to determine '*which method suits my needs*?' The overview of the most commonly used methods for the generation of targeted mutant libraries in **section I** should facilitate this task and open the door for novice users to enter this area.

While the field of enzyme engineering expands, new approaches addressing *where* to mutate or *how* to mutate genetic sequences are increasingly sought for. The 'smarter' the method used for

generating the library of variants, the more efficient the exploration of theoretical protein sequence space, and the lower the throughput of screening needed to uncover improved variants. Even the most innovative of advances will never enable exploring all of sequence space. Nonetheless, the advent of methods that facilitate generation of libraries with improved quality will make us better at navigating sequence diversity.

Looking ahead to the future necessarily brings computer aided strategies into play, to realize the goal of '*generating small(est) and high(est) quality mutant libraries*' (101). The main barrier to using predictive or computational methods is likely the unwillingness of chemists and biologists of '*getting acquainted with computational techniques*' (101). Happily, many computational tools for enzyme engineering no longer require being a computational expert (102). In the not-too-distant future, machine learning will procure a '*third approach'* together with rational design and directed evolution, to be used in complementary and ever-more-powerful ways (103). These approaches may be useful to better target the mutational hot-spots in a protein, hence, together with the selection of a smart method for library construction, they have the potential to reduce even further the number of variants that need to be screened before the improved enzyme can be identified.

2.4. References

1. Baker M. Protein engineering: navigating between chance and reason. Nat Methods. 2011;8(8):623-6.

2. Gumulya Y, Sanchis J, Reetz MT. Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. Chembiochem. 2012;13(7):1060-6.

3. Lutz S. Beyond directed evolution--semi-rational protein engineering and design. Curr Opin Biotechnol. 2010;21(6):734-43.

4. Chica RA, Doucet N, Pelletier JN. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotech. 2005;16(4):378-84.

5. Qu G, Li A, Sun Z, Acevedo-Rocha CG, Reetz MT. The crucial role of methodology development in directed evolution of selective enzymes. Angew Chem Int Ed. 2019; 59(32):13204-31

6. Emond S, Petek M, Kay EJ, Heames B, Devenish SRA, Tokuriki N, et al. Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. Nat Commun. 2020;11(1):3469.

7. Bunzel HA, Garrabou X, Pott M, Hilvert D. Speeding up enzyme discovery and engineering with ultrahigh-throughput methods. Curr Opin Struct Biol. 2018;48:149-56.

8. Ruff AJ, Dennig A, Schwaneberg U. To get what we aim for--progress in diversity generation methods. FEBS J. 2013;280(13):2961-78.

9. Gillam EM, Copp JN, Ackerley D. Directed evolution library creation. Gillam EMJ, Copp, Janine N., Ackerley, David editor. Totowa, NJ: Humana Press; 2014.

10. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH. Evolving strategies for enzyme engineering. Curr Opin Struct Biol. 2005;15(4):447-52.

11. Goldsmith M, Tawfik DS. Directed enzyme evolution: beyond the low-hanging fruit. Curr Opin Struct Biol. 2012;22(4):406-12.

12. Ali M, Ishqi HM, Husain Q. Enzyme engineering: Reshaping the biocatalytic functions. Biotech Bioeng. 2020;117(6):1877-94.

13. Tee KL, Wong TS. Polishing the craft of genetic diversity creation in directed evolution. Biotechnol Adv. 2013;31(8):1707-21.

14. Packer MS, Liu DR. Methods for the directed evolution of proteins. Nat Rev Genet. 2015;16(7):379-94.

15. Denard CA, Ren H, Zhao H. Improving and repurposing biocatalysts via directed evolution. Curr Opin Chem Biol. 2015;25:55-64.

16. Goldsmith M, Tawfik DS. Enzyme engineering by targeted libraries. Methods Enzymol. 2013;523:257-83.

17. Martínez R, Schwaneberg U. A roadmap to directed enzyme evolution and screening systems for biotechnological applications. Biol Res. 2013;46:395-405.

18. Wong TS, Roccatano D, Schwaneberg U. Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. Environ Microbiol. 2007;9(11):2645-59.

19. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH. Protein building blocks preserved by recombination. Nat Struct Biol. 2002;9(7):553-8.

20. Otey CR, Landwehr M, Endelman JB, Hiraga K, Bloom JD, Arnold FH. Structure-guided recombination creates an artificial family of cytochromes P450. PLoS Biol. 2006;4(5):e112.

21. Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, Wang ZG, et al. Library analysis of SCHEMA-guided protein recombination. Protein Sci. 2003;12(8):1686-93.

22. Ostermeier M, Shim JH, Benkovic SJ. A combinatorial approach to hybrid enzymes independent of DNA homology. Nature Biotechnology. 1999;17(12):1205-9.

23. Sieber V, Martinez CA, Arnold FH. Libraries of hybrid proteins from distantly related sequences. Nature Biotechnology. 2001;19(5):456-60.

24. Bedbrook CN, Rice AJ, Yang KK, Ding X, Chen S, LeProust EM, et al. Structure-guided SCHEMA recombination generates diverse chimeric channelrhodopsins. Proc Natl Acad Sci U S A. 2017;114(13):E2624-E33.

25. Schmitzer AR, Lépine F, Pelletier JN. Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations. Protein Eng Des Sel. 2004;17(11):809-19.

26. Quaglia D, Alejaldre L, Ouadhi S, Rousseau O, Pelletier JN. Holistic engineering of Cal-A lipase chain-length selectivity identifies triglyceride binding hot-spot. PLoS One. 2019;14(1):e0210100.

27. Reetz MT, Wilensek S, Zha D, Jaeger KE. Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. Angew Chem Int Ed 2001;40(19):3589-91.

28. Reetz MT, Bocola M, Carballeira JD, Zha D, Vogel A. Expanding the range of substrate acceptance of enzymes: Combinatorial active site saturation test. Angew Chem Int Ed. 2005;44(27):4192-6.

29. Williams GJ, Goff RD, Zhang C, Thorson JS. Optimizing glycosyltransferase specificity via "hot spot" saturation mutagenesis presents a catalyst for novobiocin glycorandomization. Chem Biol. 2008;15(4):393-401.

30. Cheng F, Zhu L, Schwaneberg U. Directed evolution 2.0: Improving and deciphering enzyme properties. Chem Commun (Camb). 2015;51(48):9760-72.

31. Tracewell CA, Arnold FH. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr Opin Chem Biol. 2009;13(1):3-9.

32. Hutchison CA, 3rd, Phillips S, Edgell MH, Gillam S, Jahnke P, Smith M. Mutagenesis at a specific position in a DNA sequence. J Biol Chem. 1978;253(18):6551-60.

33. Acevedo-Rocha CG, Reetz MT, Nov Y. Economical analysis of saturation mutagenesis experiments. Sci Rep. 2015;5(1):10654.

34. Smith M. Nobel lecture. Synthetic DNA and biology. Biosci Rep. 1994;14(2):51-66.

35. Acevedo-Rocha CG, Sun Z, Reetz MT. Clarifying the difference between iterative saturation mutagenesis as a rational guide in directed evolution and OmniChange as a gene mutagenesis technique. ChemBioChem. 2018;19(24):2542-4.

36. Siloto RMP, Weselake RJ. Site saturation mutagenesis: Methods and applications in protein engineering. Biocatal Agric Biotechnol. 2012;1(3):181-9.

37. Starr TN, Thornton JW. Epistasis in protein evolution. Protein Sci. 2016;25(7):1204-18.

38. Miton CM, Tokuriki N. How mutational epistasis impairs predictability in protein evolution and design. Protein Sci. 2016;25(7):1260-72.

39. Reetz MT. The importance of additive and non-additive mutational effects in protein engineering. Angew Chem Int Ed 2013;52(10):2658-66.

40. Reetz MT, Prasad S, Carballeira JD, Gumulya Y, Bocola M. Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: Rigorous comparison with traditional methods. J Am Chem Soc. 2010;132:9144-52.

41. Reetz MT, Carballeira JD. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat Protoc. 2007;2(4):891-903.

42. Li D, Wu Q, Reetz MT. Focused rational iterative site-specific mutagenesis (FRISM). Methods Enzymol. 2020;643:225-42.

43. Xu J, Cen Y, Singh W, Fan J, Wu L, Lin X, et al. Stereodivergent protein engineering of a lipase to access all possible stereoisomers of chiral esters with two stereocenters. J Am Chem Soc. 2019;141(19):7934-45.

44. Miyazaki K, Arnold FH. Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. J Mol Evol. 1999;49(6):716-20.

45. Tang L, Gao H, Zhu X, Wang X, Zhou M, Jiang R. Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. Biotechniques. 2012;52(3):149-58.

46. Cornish-Bowden A. Nomenclature for incompletely specified bases in nucleic acid sequences: rcommendations 1984. Nucleic Acids Research. 1985;13(9):3021-30.

47. Gaytan P, Roldan-Salgado A. Elimination of redundant and stop codons during the chemical synthesis of degenerate oligonucleotides. Combinatorial testing on the chromophore region of the red fluorescent protein mKate. ACS Synth Biol. 2013;2(8):453-62.

48. Kille S, Acevedo-Rocha CG, Parra LP, Zhang Z-G, Opperman DJ, Reetz MT, et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. ACS Synthetic Biology. 2012;2(2):83-92.

49. Tang L, Wang X, Ru B, Sun H, Huang J, Gao H. MDC-Analyzer: a novel degenerate primer design tool for the construction of intelligent mutagenesis libraries with contiguous sites. Biotechniques. 2014;56(6):301-2, 4, 6-8, passim.

50. Jacobs TM, Yumerefendi H, Kuhlman B, Leaver-Fay A. SwiftLib: rapid degenerate-codon-library optimization through dynamic programming. Nucleic Acids Res. 2015;43(5):e34.

51. Shimko TC, Fordyce PM, Orenstein Y. DeCoDe: degenerate codon design for complete protein-coding DNA libraries. Bioinformatics. 2020;36(11):3357-64.

52. Papworth, C, B, JC, B, J, et al. QuikChange site-directed mutagenesis. Strategies. 1996;9(8):3-4.

53. Zheng L. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. 2004;32(14):e115.

54. Miyazaki K, Takenouchi M. Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. Biotechniques. 2002;33(5):1033-4, 6-8.

55. Tseng WC, Lin JW, Wei TY, Fang TY. A novel megaprimed and ligase-free, PCR-based, site-directed mutagenesis method. Anal Biochem. 2008;375(2):376-8.

56. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. 2008;8:91.

57. Hogrefe HH, Cline J, Youngblood GL, Allen RM. Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. Biotechniques. 2002;33(5):1158-60, 62, 64-5.

58. Li A, Acevedo-Rocha CG, Reetz MT. Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. Appl Microbiol Biotechnol. 2018;102(14):6095-103.

59. Firnberg E, Ostermeier M. PFunkel: Efficient, expansive, user-defined mutagenesis. PLoS ONE. 2012;7(12):e52031.

60. Forloni M, Liu AY, Wajapeyee N. Oligonucleotide-directed mutagenesis by elimination of a unique restriction site (USE Mutagenesis). Cold Spring Harb Protoc. 2019(1):pdb.prot097782.

61. Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci USA. 1985;82:488-92.

62. Liu B, Long S, Liu J. Improving the mutagenesis efficiency of the Kunkel method by codon optimization and annealing temperature adjustment. Nat Biotechnol. 2020;56:46-53.

63. Huovinen T, Brockmann E-C, Akter S, Perez-Gamarra S, Ylä-Pelto J, Liu Y, et al. Primer extension mutagenesis powered by selective rolling circle amplification. PLoS ONE. 2012;7(2):e31817.

64. Deng WP, Nickoloff JA. Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. Anal Biochem. 1992;200(1):81-8.

65. Markvardsen P, Lassen SF, Borchert V, Clausen IG. Uracil-USE, an improved method for site-directed mutagenesis on double-stranded plasmid DNA. Biotechniques. 1995;18(3):370-2.

66. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene. 1989;77(1):51-9.

67. Forloni M, Liu AY, Wajapeyee N. Megaprimer polymerase chain reaction (PCR)-Based Mutagenesis. Cold Spring Harb Protoc. 2019(6):pdb.prot097824.

68. Peng RH, Xiong AS, Yao QH. A direct and efficient PAGE-mediated overlap extension PCR method for gene multiple-site mutagenesis. Appl Microbiol Biotechnol. 2006;73(1):234-40.

69. Lee J, Shin MK, Ryu DK, Kim S, Ryu WS. Insertion and deletion mutagenesis by overlap extension PCR. Methods Mol Biol. 2010;634:137-46.

70. Kadkhodaei S, Memari HR, Abbasiliasi S, Rezaei MA, Movahedi A, Shun TJ, et al. Multiple overlap extension PCR (MOE-PCR): an effective technical shortcut to high throughput synthetic biology. RSC Advances. 2016;6(71):66682-94.

71. Wei H, Hu J, Wang L, Xu F, Wang S. Rapid gene splicing and multi-sited mutagenesis by one-step overlap extension polymerase chain reaction. Anal Biochem. 2012;429(1):76-8.

72. Waneskog M, Bjerling P. Multi-fragment site-directed mutagenic overlap extension polymerase chain reaction as a competitive alternative to the enzymatic assembly method. Anal Biochem. 2014;444:32-7.

73. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ. A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Res. 1989:7.

74. Jain PC, Varadarajan R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. Anal Biochem. 2014;449:90-8.

75. Chiu J, March PE, Lee R, Tillett D. Site-directed, ligase-independent mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic Acids Res. 2004;32(21):e174.

76. Laible M, Boonrod K. Homemade site directed mutagenesis of whole plasmids. J Vis Exp. 2009;27:e1135.

77. Sambrook J, Russell DW. Site-specific mutagenesis by overlap extension. Cold Spring Harb Protoc. 2006;1(1):pdb.prot3468.

78. Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y. Applications of the restriction free (RF) cloning procedure for molecular manipulations and protein expression. J Struct Biol. 2010;172(1):34-44.

79. van den Ent F, Löwe J. RF cloning: A restriction-free method for inserting target genes into plasmids. J Biochem Bioph Methods. 2006;67(1):67-74.

80. Lund B, Leiros H-K, Bjerga GE. A high-throughput, restriction-free cloning and screening strategy based on ccdB-gene replacement. Microbial Cell Factories. 2014;13(1):38.

81. Ulrich A, Andersen KR, Schwartz TU. Exponential megapriming PCR (EMP) cloning--seamless DNA insertion into any target plasmid without sequence constraints. PLoS One. 2012;7(12):e53360.

82. Zeng F, Hao Z, Li P, Meng Y, Dong J, Lin Y. A restriction-free method for gene reconstitution using two single-primer PCRs in parallel to generate compatible cohesive ends. BMC Biotechnol. 2017;17(1):32.

83. Belsare KD, Andorfer MC, Cardenas FS, Chael JR, Park HJ, Lewis JC. A simple combinatorial codon mutagenesis method for targeted protein engineering. ACS Synth Biol. 2017;6(3):416-20.

84. Bloom JD. An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol Biol Evol. 2014;31(8):1956-78.

85. Heckman KL, Pease LR. Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc. 2007;2(4):924-32.

86. Jin P, Kang Z, Zhang J, Zhang L, Du G, Chen J. Combinatorial evolution of enzymes and synthetic pathways using one-step pcr. ACS Synth Biol. 2016;5(3):259-68.

87. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 1990;18:6069-74.

88. Li MZ, Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods. 2007;4(3):251-6.

89. Motohashi K. A simple and efficient seamless DNA cloning method using SLiCE from Escherichia coli laboratory strains and its application to SLiP site-directed mutagenesis. BMC Biotechnol. 2015;15(1):47.

90. Motohashi K. Seamless Ligation Cloning Extract (SLiCE) Method using cell lysates from laboratory escherichia coli strains and its application to SLiP site-directed mutagenesis. In: Reeves A, editor. In Vitro Mutagenesis. 1498. New York, NY: Springer New York; 2017. p. 349-57.

91. Islam MN, Lee KW, Yim H-S, Lee SH, Jung HC, Lee J-H, et al. Optimizing T4 DNA polymerase conditions enhances the efficiency of one-step sequence- and ligation-independent cloning. BioTechniques. 2017;63(3):3.

92. Xia Y, Li K, Li J, Wang T, Gu L, Xun L. T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. Nucleic Acids Res. 2019;47(3):e15.

93. Dennig A, Shivange AV, Marienhagen J, Schwaneberg U. OmniChange: The sequence independent method for simultaneous site-saturation of five codons. PLoS ONE. 2011;6(10):e26222.

94. Garcia-Nafria J, Watson JF, Greger IH. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Sci Rep. 2016;6:27459.

95. Trehan A, Kielbus M, Czapinski J, Stepulak A, Huhtaniemi I, Rivero-Muller A. REPLACR-mutagenesis, a one-step method for site-directed mutagenesis by recombineering. Sci Rep. 2016;6:19121.

96. Alcalde M, Zumarraga M, Polaina J, Ballesteros A, Plou F. Combinatorial saturation mutagenesis by in vivo overlap extension for the engineering of fungal laccases. Comb Chem High Throughput Screen. 2006;9(10):719-27.

97. Li A, Acevedo-Rocha CG, Sun Z, Cox T, Xu JL, Reetz MT. Beating bias in the directed evolution of proteins: Combining high-fidelity on-chip solid-phase gene synthesis with efficient gene assembly for combinatorial library construction. Chembiochem. 2018;19(3):221-8.

98. Li A, Acevedo-Rocha CG, D'Amore L, Chen J, Peng Y, Garcia-Borras M, et al. Regio- and stereoselective steroid hydroxylation at the C7-position by cytochrome P450 monooxygenase mutants. Angew Chem Int Ed. 2020;59:12499.

99. Field JEJ, Rickerby HF, inventors; LabGenius LTD, assignee. Compositions and methods for polynucleotide assembly (WO 2017/046594) patent WO 2017/046594. 2017.

100. Nour-Eldin HH, Geu-Flores F, Halkier BA. USER cloning and USER fusion: The ideal cloning techniques for small and big laboratories. In: Fett-Neto A, editor. Plant Secondary Metabolism Engineering. Methods in Molecular Biology. Totowa, NJ: Humana Press; 2010. p. 185-200.

101. Li G, Dong Y, Reetz MT. Can machine learning revolutionize directed evolution of selective enzymes? Adv Synth Catal. 2019;361:2377 – 86.

102. Ebert MC, Pelletier JN. Computational tools for enzyme improvement: why everyone can - and should - use them. Curr Opin Chem Biol. 2017;37:89-96.

103. Mazurenko S, Prokop Z, Damborsky J. Machine learning in enzyme engineering. ACS Catal. 2019;10(2):1210-23.

Chapter 3 – Known evolutionary paths are accessible to engineered ß-lactamases having altered protein motions at the timescale of catalytic turnover

Preface to Chapter 3

To better guide enzyme engineering projects, it is crucial to identify the features that make an enzyme easily evolvable towards a new function. In this sense, knowledge of the three-dimensional structure of proteins, as well as substrate reactivity and stability have considerably accelerated enzyme engineering projects. Another feature, protein dynamics, has recently attracted more attention due to its structural importance and relevance for function. In fact, a number of studies have shown changes in protein dynamics along directed molecular evolution trajectories that contributed to improved activity. Protein dynamics is an inherent and complex biophysical feature that encompasses multiple timescales and regions of proteins. In this study we use as start-point in directed evolution, TEM-1 ß-lactamase variants that differ broadly in the dynamics at the millisecond to picosecond timescale. Motions relevant for enzymatic catalysis as well as loop rearrangements occur at this timescale. To examine whether slow dynamics at the outset of evolution affect access to different evolutionary trajectories, variants were subjected to a known evolutionary trajectory through site-directed mutagenesis and to directed molecular evolution using error-prone PCR to randomly explore alternative evolutionary trajectories.

The results of this chapter were published in 2020 under the topic 'Molecular Evolution: You learn from your Mistakes' in the Structural Biology section of the journal Frontiers in Molecular Biosciences. This study was designed by Prof. Joelle Pelletier and myself. Site-directed mutagenesis and directed molecular evolution were done by myself. The enzymatic characterization was done with the aid of my then-summer intern Claudèle Lemay-St-Denis. To rationalize our kinetic results, we included computational simulations through a collaboration with the group of Prof. Victor Guallar at the Barcelona Supercomputing Center. PhD candidate Ferran Sancho-Jodar from the Guallar group started the collaboration as a visiting student; the collaboration was continued by Prof. Guallar. Analysis of the simulations was done by myself with

the help of PhD candidate Carles Perez-Lopez from the Guallar group. I wrote the original draft and Profs Pelletier and Guallar contributed to writing the subsequent versions. All authors reviewed and revised the manuscript.

Article 2. Known evolutionary paths are accessible to engineered ß-lactamases having altered protein motions at the timescale of catalytic turnover

Lorea Alejaldre[1,2,3], Claudèle Lemay-St-Denis[1,2,3], Carles Perez[4], Ferran Sancho-Jodar[4], Victor Guallar[4], Joelle N. Pelletier [1,2,3,5*]

[1]Biochemistry Department, Université de Montréal, Montréal, Québec, Canada

[2]PROTEO, the Québec Network for Research on Protein, Function, Engineering and Applications, Qc, Canada;

[3]CGCC, Center in Green Chemistry and Catalysis, Montréal H3A 0B8, Qc, Canada;

[4]Barcelona Supercomputing Center, Barcelona, Spain

[5]Chemistry Department, Université de Montréal, Montréal, Québec, Canada

***Correspondence:**

Joelle Pelletier

joelle.pelletier@umontreal.ca

Keywords: Enzyme engineering, epistasis, protein dynamics, protein engineering start-points, TEM-1 beta lactamase, slow timescales, protein evolution.

2020 Frontiers in Molecular Biosciences

3.1. Abstract

The evolution of new protein functions is dependent upon inherent biophysical features of proteins. Whereas it has been shown that changes in protein dynamics can occur in the course of directed molecular evolution trajectories and contribute to new function, it is not known whether varying protein dynamics modify the course of evolution. We investigate this question using three related ß-lactamases displaying dynamics that differ broadly at the slow timescale that corresponds to catalytic turnover yet have similar fast dynamics, thermal stability, catalytic and substrate recognition profiles. Introduction of substitutions E104K and G238S, that are known to have a synergistic effect on function in the parent ß-lactamase, showed similar increases in catalytic efficiency towards cefotaxime in the related ß-lactamases. Molecular simulations using Protein Energy Landscape Exploration reveal that this results from stabilizing the catalytically-productive conformations, demonstrating the dominance of the synergistic effect of the E104K and G238S substitutions *in vitro* in contexts that vary in terms of sequence and dynamics. Furthermore, three rounds of directed molecular evolution demonstrated that known cefotaximase-enhancing mutations were accessible regardless of the differences in dynamics. Interestingly, specific sequence differences between the related ß-lactamases were shown to have a higher effect in evolutionary outcomes than did differences in dynamics. Overall, these ß-lactamase models show tolerance to protein dynamics at the timescale of catalytic turnover in the evolution of a new function.

3.2. Introduction

Intragenic epistasis describes the non-additive effect of mutations on protein fitness. It has been shown to hinder predictability in protein engineering experiments (1-6), highlighting the need for investigation into its causes and effects. In recent years, deep-mutational scanning has been applied to extensively investigate both the impact of single mutations on a given function (7-9) and, of particular interest with respect to intragenic epistasis, all possible combinations of double mutants (10). However, exhaustive investigation of combinatorial mutations is resource intensive and out of reach for the average-sized protein (a 50 amino acid protein requires a library of 20^{50} $=1.1x10^{65}$ variants). Therefore, focused combinatorial libraries and directed molecular evolution experiments are currently the most useful approach to uncover epistatic interactions (11-13).

Intragenic epistasis comes in different forms. Mutations that individually have a beneficial effect on fitness can act synergistically (positive epistasis) or can be deleterious when combined (negative sign epistasis); on the other hand, deleterious mutations can be beneficial when combined (positive sign epistasis) (3). The variety of epistatic interactions imposes constraints on mutational pathways leading to a new protein function: pathways that include a node with lower fitness are not evolutionarily accessible whereas specific pathways can be decided by early, positive epistatic effects in evolution (2, 4, 6, 14, 15). Factors known to influence epistasis include direct inter-residue interactions as well as long-range interactions mediated by dynamic networks and protein stability (2, 3). Despite increasing evidence that protein dynamics modulate function and fitness (6, 16-21), the impact of protein dynamics on intragenic epistasis is not well understood.

Dynamics are a complex feature of proteins. They encompass motions of varied magnitude and timescales; no single methodology can readily evaluate the full range of motion describing a protein (21-23). Protein dynamics are selected for during natural evolution of enzymes, consistent with dynamics being a determinant of function (24-26). Tracking protein dynamics during the course of experimental evolution has equally revealed correlations between protein dynamics, function and evolution (27, 28). Ancestral reconstruction of ß-lactamases revealed that protein motions at fast timescales decreased in conjunction with evolution of specificity (29). The directed evolution of a phosphotriesterase to an arylesterase, and back to a phosphotriesterase, revealed that the majority of mutations, although distant from the active site, were epistatic and caused conformational changes (16). Similarly, the additive combination of mutations distant from the active site increased protein motions of inactive proline isomerase CypA, restoring its activity (30). Finally, simulations of a beneficial sign epistatic interaction obtained by laboratory evolution of a metallo-ß-lactamase (31) were experimentally verified to result in increased protein motions on the millisecond to microsecond timescales (32).

Evolution is thus shown to modulate protein dynamics; do dynamics modulate the course of evolution? Subjecting to evolution proteins having differing dynamic patterns may reveal alternative evolutionary paths. Alternatively, differing dynamics at the start-point of evolution may not be determinant in evolutionary mechanisms. This might depend on the nature of the

protein motions in play: their location and frequency may determine their impact on evolution.

To address this question, we performed the laboratory evolution of two variants of TEM-1 ß-lactamase, a widely-used model for the study of enzyme evolution (14, 33-35) that readily evolves towards hydrolysis of new ß-lactam antibiotics (36, 37). Previous studies have revealed the most prevalent evolutionary paths and shown that the evolution of TEM-1 ß-lactamase is highly constrained due to epistasis (14, 38, 39). In addition, high active site dynamics have been proposed both to promote and to prevent protein evolvability in TEM-1 ß-lactamase (29, 34). This suggests that evolving variants having different motions may result in different endpoints, or in different paths to reach the same point. The two specific TEM-1 ß-lactamase variants were chosen because they share similar catalytic activity and substrate recognition, thermal stability, as well as conserved motions at fast (ps-ns) timescales, but they differ greatly in the location and extent of their motions at slow timescales (μs-ms) (22) (Figure 3.1).

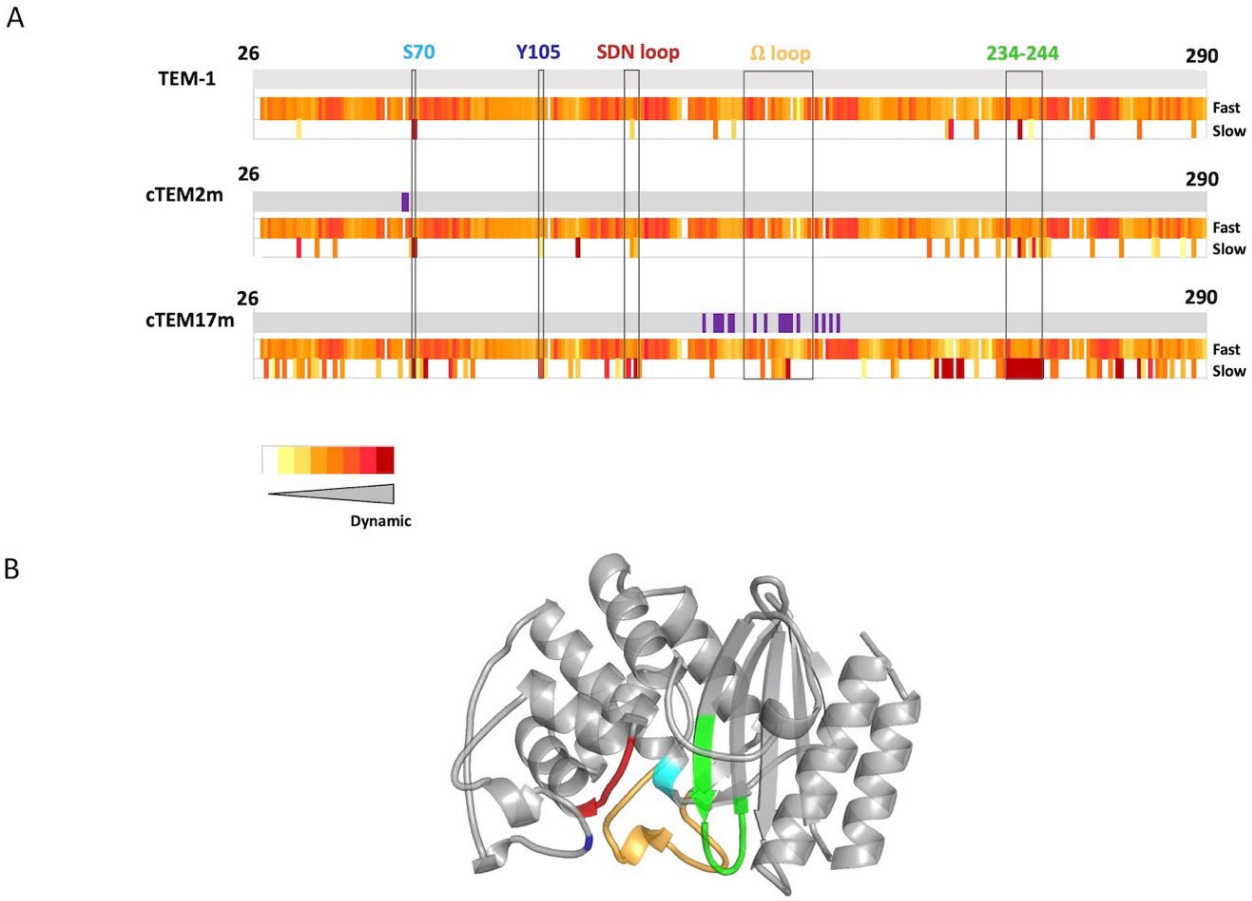


Figure 3.1. – Schematic representation of protein dynamics and 3D structure of TEM-1 ß-lactamase highlighting catalytically-relevant regions.

(A) Schematic representation of protein dynamics determined at fast time scales (ps-ns) and slow time scales (µs to ms or slower) below sequence diagrams for TEM-1, cTEM-2m and cTEM-17m (22). Amino acid substitutions with respect to TEM-1 are shown in purple. The slow timescale coincides with the timescale of catalytic turnover (TEM-1 k_{cat} for benzylpenicillin = 450 s^{-1}) (40). (B) Structure of TEM-1 highlighting catalytically-relevant regions: S70 (cyan), Y105 (dark blue), SDN loop (red), Ω-loop (yellow) and 234-244 wall (green)

Mutations E104K/G238S appear in the most efficient evolutionary pathways in natural and directed molecular evolution of TEM-1 towards cefotaxime, exhibiting positive epistasis (14, 37, 41, 42). We therefore introduced mutations E104K and G238S into the two variants of TEM-1 that differ in slow dynamics but exhibit similar catalytic reactivity, to examine the impact of protein dynamics on the evolution of new cefotaximase activity. As for TEM-1, this resulted in synergistic

improvement of cefotaximase activity, demonstrating that this evolutionary trajectory remained efficient in different dynamic backdrops. To examine the broader impact of protein dynamics on evolution, we performed three rounds of directed molecular evolution with random mutagenesis to allow exploration of alternative pathways to cefotaximase activity. These revealed that mutations previously reported in the evolution of TEM-1 also appear in the variants with differing slow dynamics. Our results show that TEM-1 ß-lactamase is a robust system where evolvability is compatible with diverse dynamic backgrounds, and where variants possessing slow timescale dynamics that vary in extent and in location can access known evolutionary pathways.

3.3. Materials and methods

3.3.1. Site-directed mutagenesis

Mutations E104K and G238S were introduced individually and jointly into ß-lactamase constructs TEM-1, cTEM-2m and cTEM-17m. The ß-lactamase genes were fused to an OmpA signal peptide in 3' for periplasmic export and cloned into the plasmid pET-24 as previously reported (43). Whole plasmid site-directed mutagenesis (44) was done using primers (Sigma Aldrich) designed according to the QuikChange Lightning kit: E104K-F (5'-TATTCTCAGAATGACTTGGTTAAGTACTCACCAGTCACAG-3'), E104K-R (5'-CTGTGACTGGTGAGTACTTAACCAAGTCATTCTGAGAATA-3'), G238S-F (5'-GATAAATCTGGAGCCAGTGAGCGTGGGTCTC-3') and G238S-R (5'-GAGACCCACGCTCACTGGCTCCAGATTTATC-3'). Pfu polymerase (Agilent) was used, with reaction conditions according to the manufacturer's recommendations and an extension time of 6 min (~1 min/kb). The reaction product was digested with DpnI (NEB) for 1h at 37ºC to eliminate the template and one-tenth of the reaction was transformed into $CaCl_2$-competent *E. coli* XL1-Blue prepared following the Inoue method (45). Where mutagenesis was unsuccessful by that method, mutagenesis was undertaken by overlap extension (46). Briefly, the forward primers (above) were employed with the T7- reverse primer (5'- ATGCTAGTTATTGCTCAGC-3') and the reverse primers (above) with the T7+ forward primer (5'-TAATACGACTCACTATAGGG-3') in separate PCR reactions with the Phusion polymerase (Thermo Scientific) according to the manufacturer's recommendations. The full-length mutated gene was reassembled from the resulting mutated

PCR products in a third PCR reaction using both T7+ and T7- primers. The purified PCR products were subcloned into pET-24 using NdeI and HindIII restriction enzymes (NEB) and DNA ligation kit (Takara). The reaction product was transformed into $CaCl_2$-competent *E. coli* XL1-Blue cells. Following selection on Luria-Bertani agar plates containing 50 µg/mL kanamycin, sequences were confirmed by DNA sequencing (IRIC Genomics Platform at Université de Montréal).

3.3.2. DNA library generation

Libraries were generated according to the protocol of Copp, Hanson-Manful (47). Briefly, random mutagenesis was performed on pET24-cTEM-2m and pET24-cTEM-17m using T7+ and T7- primers with the GeneMorph II kit (Agilent) to obtain a low mutation rate (1-3 mut/kb). Reaction products were sub-cloned into pET-24 as described above, but using T4 DNA ligase (NEB) overnight at 4ºC. Following DNA purification using Monarch PCR and DNA purification kit (NEB), half of the reaction was transformed into electrocompetent E. cloni cells (Lucigen) and spread onto Luria-Bertani agar plates containing 50 µg/mL kanamycin. After overnight incubation at 37ºC, the DNA of 10-20 colonies was amplified using 2x PCR Precision Master mix (ABM) and sequenced to verify the mutation rate. DNA sequencing was done by the Genomic Platform of IRIC or the Genome Quebec Innovation Center at McGill University, Canada. Library size was estimated by counting the resulting colony forming units after overnight incubation at 37ºC in Luria-Bertani agar plates containing 50 µg/mL kanamycin. Subsequently, the libraries were pooled, miniprepped using the Monarch Plasmid miniprep kit (NEB) and transformed into *E. coli* BL21(DE3) for expression and screening.

3.3.3. Directed molecular evolution

Libraries were screened on agar plates containing the antibiotic cefotaxime (Sigma-Aldrich) at concentrations 0.08-0.032 µg/mL. Several dilutions of the library were plated on non-selective medium to ensure that colony forming units were 10-20-fold higher than the estimated library size. To favor genetic diversity and avoid early evolutionary dead-ends, between 25 and 70 colonies were pooled and selected at each round. Plasmid DNA was extracted with Monarch plasmid miniprep kit (NEB), amplified and used as a template for random mutagenesis to produce the next generation. This process was repeated for a total of three generations. The selection

process was tracked by sequencing between 5 and 14 colonies after each selection (IRIC Genomics Platform at Université de Montréal).

3.3.4. Protein expression and purification

Expression and purification of the ß-lactamase variants were performed as in Gobeil, Clouthier (48). Briefly, 5 mL of an overnight culture were inoculated in 400 mL of ZYP-5052 autoinduction medium containing 100 µg/mL kanamycin. After initial growth at 37ºC until $OD_{600\ nm}$=0.6, expression was carried out overnight at 22ºC. After centrifugation at 3,000 rpm for 30 min, the cell pellet was resuspended in 10 mM Tris-HCl pH 7.0 buffer and lysed using a cell disrupter (Constant Systems). After centrifugation at 20,000 rpm at 4ºC for 30 min, the supernatant was filtered through a 0.2 µm filter and injected onto a DEAE-Sepharose Fast Flow column (1.6 cm x 30 cm). A linear gradient from 10 mM to 200 mM Tris-HCl pH 7.0 was applied. If purity was < 80%, size exclusion chromatography was performed over a Superdex 75 column (1.6 cm x 55 cm) equilibrated with 50 mM Tris-HCl pH 7.0. Protein concentration and buffer exchange were done in 10 MWCO Ultra Centrifugal Filter Units MilliporeSigma Amicon (Fischer Scientific).

3.3.5. Minimal inhibitory concentration (MIC) assays

Minimal inhibitory concentration (MIC) assays were performed according to Wiegand, Hilpert (49) using the agar plate method. Briefly, *E. coli* BL21(DE3) cells expressing the ß-lactamase variants were propagated overnight in Luria-Bertani medium containing 50 µg/mL kanamycin. An inoculum of 10^4 cfu was spotted onto Luria-Bertani agar plates, containing 0.25 mM IPTG and 0.004-1048 µg/mL cefotaxime in two-fold dilutions. After overnight incubation at 37ºC, the plate which had the lowest cefotaxime concentration at which there was no visible growth was considered as the minimal inhibitory concentration. Each construct was assayed in triplicate.

3.3.6. Enzyme kinetics

Kinetic parameters for cefotaxime hydrolysis were determined at 27ºC in a Cary 100 Bio UV-Visible (Agilent) spectrophotometer, where $\Delta\varepsilon_{264\ nm}$ = 7250 $M^{-1}cm^{-1}$ (40). Cefotaxime concentrations ranged between 12.5 µM – 200 µM and enzyme concentration 5-70 nM. Kinetic constants were calculated using GraphPad Prism 6 using Michaelis-Menten equation when K_M was lower than 150 µM. Otherwise, the Lineweaver-Burk representation was used to calculate

kinetic parameters due to inability to saturate the enzyme with CTX, a common issue with cephalosporins (50-53).

3.3.7. Thermal stability

Differential scanning fluorimetry was used to determine the melting temperature of the ß-lactamase variants. Protein concentrations of 1 µM – 8 µM in sodium phosphate buffer 50 mM pH 7.0 were combined with 2.5 ×, 3.3 × and 5 × SYPRO Orange in a final volume of 20 µL. A temperature gradient of 20-95ºC with a 0.04ºC/s increase was applied in a Light Cycler 480 qPCR instrument (Roche). Fluorescence was measured with λ_{exc} = 483 nm and λ_{em} = 568 nm. Melting temperature values were calculated from the negative derivative of the fluorescence value versus temperature where the minimum represents the melting temperature (54).

3.3.8. System preparation for computational analysis

The crystal structures 1XPB for TEM-1, 4MEZ for cTEM-2m and 4ID4 for cTEM-17m were prepared with Schrodinger's Preparation Wizard. Active site waters blocking substrate docking were manually removed, taking care to keep the water molecule bridging S70 and E166. Missing side chains were added and sampled through Prime. Hydrogens were added and the protonation states manually checked with the interactive optimizer. For TEM-1, this resulted in H26 and H96 in epsilon protonation, the remaining histidines being delta-protonated. In cTEM-2m, H96 was additionally delta-protonated; it is quite solvent exposed and far from the active site. For cTEM-17m, H26 and H29 were epsilon-protonated and H112 was doubly protonated (thus having a +1 charge). Although the Wizard suggested that K234, immediately next to the nucleophilic S70, should be neutral, we kept it positive as it might be crucial for substrate binding and for stabilizing deprotonation S70 (55).

We performed Glide docking of cefotaxime using S70 as the grid center and increasing the (grid) inner box diameter to 16 Å to guarantee substrate accommodation. We attempted docking before and after computing quantum charges for the ligands, at the B3LYP DFT and 6/31G** level of theory (56) (57) but obtained analogous results. We thus adopted the classical OPLSAA atom parameters for the ligand since in our experience they might represent better average ones (58). Docking used default SP Glide parameters for the three systems resulting in good catalytic

positions, with catalytic substrate-serine distances < 4 Å. After docking cefotaxime, double mutants were generated with Maestro and the mutated side chains were sampled with Prime prior to performing PELE simulations.

3.3.9. Protein Energy Landscape Exploration method (PELE)

Using the cefotaxime-docked structures as input, local PELE explorations were run with the recent AdaptivePELE version (59). PELE is a Monte Carlo protein-ligand sampling procedure capable of accurately mapping ligand migration and induced fit mechanisms (60). Each simulation involved 20 adaptive epochs of 75 PELE steps each, running on 48 computing cores for a wall clock of ~15 hours. The system was described using PELE's default energy function, the 2005 OPLS-AA level of theory with a surface generalized Born model implicit solvent (61). For each sampled pose, substrate binding energies were computed using the interaction force field energy: $E_{complex} - E_{receptor} - E_{substrate}$.

3.3.10. PELE analysis

In-house Python scripts were used to identify poses with the lowest interaction energies where the catalytic S70 was at a distance of < 3.5 Å from C3 of the cefotaxime ß-lactam ring. Scripts are available upon request. Visual inspection of the poses as well as identification of productive poses was done using VMD (62). The salt bridges plugin in VMD was used to identify interactions. This considers that a salt bridge is formed when any oxygen atom of a residue identified as acidic and any nitrogen atom of a residue identified as basic are at a cut-off distance of < 3.2 Å in at least one frame. Distance between residue 104 and P167 was calculated using the Tk console of VMD with a freely available script at https://github.com/ipudu/useful-vmd-scripts/blob/master/distance.tcl. Distances were calculated between the center of mass of two residues, as well as between the α-carbons. Active site volume was calculated using F-Pocket (63).

3.4. Results

3.4.1. Positive epistasis is maintained in a known evolutionary path despite altered protein dynamics.

Protein dynamics on slow timescales (µs to ms or slower) see large conformational

rearrangements associated with ligand-binding and substrate turnover. We previously reported two variants of TEM-1 ß-lactamase where functionality is maintained despite sequence differences and marked differences in the magnitude and location of slow protein motions (Figure 3.1) (22, 48). Variants cTEM-2m and cTEM-17m include, respectively, two and 17 substitutions in the active site region. The substitutions of cTEM-2m are M68L and M69T, the immediate neighbors of the nucleophilic S70; those of cTEM-17m are on an opposite face of the active site, within and flanking the catalytic Ω-loop (Figure 3.2). In contrast to TEM-1, where only 13 residues display slow motions near the frequency of turnover (µs to ms or slower, where the k_{cat} of TEM-1 for benzylpenicillin is 450 s^{-1} (40)), variant cTEM-2m exhibits slow motions in 29 residues; these are predominantly located in the active site cavity at the catalytically relevant Y105 and SDN loop but exclude the Ω-loop. Variant cTEM-17m differs yet more strikingly, with slow motions mapped to 82 residues that broadly span the enzyme and include the Ω-loop (Figure 3.1) (22). Nonetheless, their catalytic efficiencies (k_{cat}/K_M) are within one order of magnitude of the native TEM-1 ((22, 40) and Table 3.1). Conservation of their kinetic properties correlates with conservation of their fast motions (ps-ns), despite wide-ranging alterations of their motions on the timescale of catalytic turnover (24).

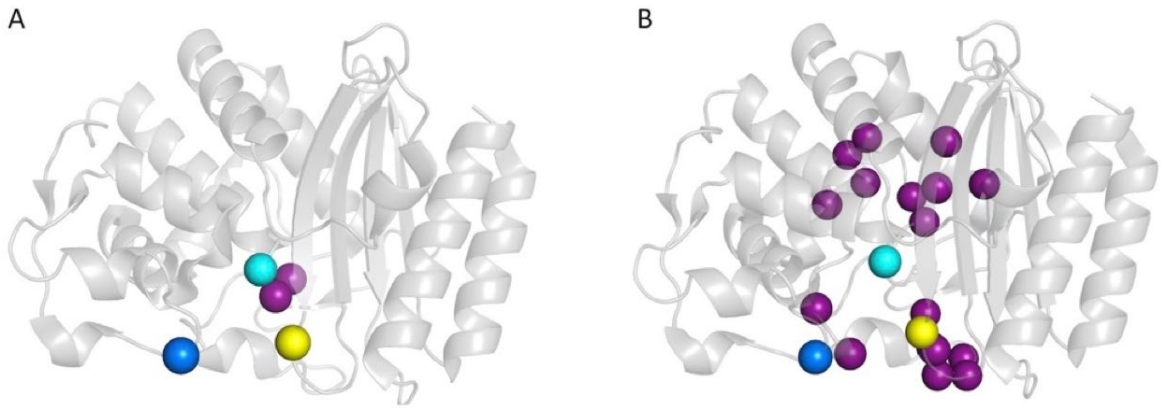


Figure 3.2. – Location of mutations E104K and G238S in host ß-lactamases.

(A) cTEM-2m (PDB 4mez) and (B) cTEM-17m (PDB 4id4). Mutated residues with respect to TEM-1 are highlighted as purple spheres and the catalytic serine 70 is in cyan, E104K in dark blue and G238S in yellow.

The synergistic mutations E104K/G238S are prevalent in cefotaxime-resistant clinical isolates and consistently appear in directed molecular evolution experiments (36, 37, 64). These mutations were introduced into the two TEM-1 variants to determine whether the E104K/G238S mutational pathway to cefotaxime resistance is accessible in the context of altered active site residues and slow motions. The impact of the two and 17 mutations of variants cTEM-2m and cTEM-17m on the newly inserted E104K/G238S mutations was evaluated by determining the *in vitro* kinetics of cefotaxime hydrolysis (Table 3.1). In all cases, inclusion of E104K/G238S procured remarkable increases in catalytic efficiency (~600-fold in cTEM-2m; ~360-fold in cTEM-17m), comparable to the ~320-fold increase observed for TEM-1 E104K/G238S. These results immediately demonstrate that, despite exhibiting unique and widespread patterns of increased slow dynamics throughout the active site (Figure 3.1), cTEM-2m and cTEM-17m can evolve cefotaximase activity via the same mutations as TEM-1 ß-lactamase (Table 3.1).

Variant cTEM-17m is the most genotypically distant from TEM-1 (17 substitutions) and differs most in the range and extent of its slow dynamics, yet inclusion of mutations E104K/G238S had essentially the same effect on catalytic efficiency (190 mM^{-1} s^{-1}, relative to 220 mM^{-1} s^{-1} in TEM-1, Table 3.1); variant cTEM-2m was only 4-fold less efficient. Catalytic turnover (k_{cat}) was strongly improved by inclusion of mutations E104K/G238S in all three settings: a ~180-fold increase in cTEM-2m and cTEM-17m, significantly greater than the ~30-fold increase in TEM-1. This was accompanied by a modest increase in productive binding to cefotaxime: K_M was reduced by ~2-3-fold in cTEM-2m and cTEM-17m, which both had higher affinity at the outset, compared to ~10-fold in TEM-1. Although mutations E104K/G238S have been found to be destabilizing in TEM-1 and are often accompanied with the stabilizing mutation M182T (42), it has been reported that thermostability is not a major driving force in the evolution of CTX resistance in TEM-1 ß-lactamase (50). It is interesting to note that inclusion of mutations E104K/G238S decreased thermostability by ~7°C in both TEM-1 and cTEM-2m; despite cTEM-17m displaying the most widespread slow dynamics, its melting temperature was hardly affected by inclusion of mutations E104K/G238S (~2°C decrease).

Insertion of the individual E104K and G238S substitutions revealed a clear increase on cefotaximase specific activity in the three host ß-lactamases (Table 3.1). Inclusion of G238S

procured 10- to 100-fold greater improvement to specific activity against cefotaxime than E104K in each host, consistent with the prevalence of the G238S substitution in clinical isolates/evolution towards cephalosporins (36, 37, 64). Most of the catalytic improvement resulted from increased k_{cat} due to G238S (30 to160-fold). In contrast, the effects of the individual mutations on K_M were modest, whether in improving K_M in TEM-1 (0.2 to 0.5-fold) or weakening it in cTEM-2m and cTEM-17m (1.2 to 4-fold) (Table 3.1). Indeed, sign epistasis was observed for K_M in cTEM-2m and cTEM-17m; in both cases, one or both substitutions weakened K_M and their combination had a favorable effect on K_M. This demonstrates that substitutions E104K and G238S in all three host ß-lactamases increase catalytic activity, yet the pattern of improvement is modulated by the differences between the hosts.

Table 3.1. – Cefotaximase activity *in vitro* and in *E. coli*, and thermostability of host ß-lactamases TEM-1, cTEM-2m, cTEM-17m and their corresponding E104K/G238S variants.

	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}\ s^{-1}$)	MIC (μg/mL)	T_m (˚C)
TEM-1	1500 ± 900	1.0 ± 0.5	0.68 ± 0.53	0.03	49.9 ± 0.3
TEM-1 E104K	750 ± 560 (0.5 ×) [a]	1.0 ± 0.6 (1 ×)	1.3 ± 1.2 (2 ×)	0.1	49.4 ± 0.1
TEM-1 G238S	240 ± 80 (0.2 ×)	30 ± 8 (30 ×)	130 ± 100 (190 ×)	0.6	43.8 ± 0.1
TEM-1 E104K/G238S	150 ± 50 (0.1 ×)	33 ± 9 (33 ×)	220 ± 90 (320 ×)	130	43.5 ± 0.1
cTEM-2m [b]	240 ± 40	0.02 ± 0.01	0.083 ± 0.044	0.02	49.6 ± 0.2
cTEM-2m E104K	680 ± 170 (3 ×)	0.3 ± 0.1 (15 ×)	0.44 ± 0.25 (5 ×)	0.02	47.5 ± 0.4
cTEM-2m G238S	290 ± 40 (1.2 ×)	1.4 ± 0.3 (70 ×)	4.8 ± 1.4 (60 ×)	0.04	42.3 ± 0.2
cTEM-2m E104K/G238S	77 ± 10 (0.3 ×)	3.8 ± 0.2 (190 ×)	49 ± 7 (600 ×)	0.09	42.0 ± 0.4
cTEM-17m [c]	260 ± 100	0.14 ± 0.04	0.54 ± 0.26	0.03	49.0 ± 0.8
cTEM-17m E104K	920 ± 90 (4 ×)	3.1 ± 0.5 (20 ×)	3.4 ± 0.7(6 ×)	0.02	48.3 ± 0.3
cTEM-17m G238S	730 ± 290 (3 ×)	23 ± 8 (160 ×)	32 ± 17 (60 ×)	43	48.0 ± 0.5
cTEM-17m E104K/G238S	130 ± 30 (0.5 ×)	25 ± 3 (180 ×)	190 ± 50 (350 ×)	109	46.9 ± 0.5

[a] Fold increase relative to the respective host ß-lactamase is given in parentheses. Changes equal to or greater than one order of magnitude are highlighted in red.

[b] Values taken from (22).

[c] Values taken from (40).

At a glance, the MIC values against CTX of the individually and doubly substituted variants did not show a clear correlation with the kinetic parameters (Table 3.1). A closer look shows similar trends in improvement of specific activity and increased MIC for the TEM-1 series and the cTEM-17m series. In both cases, G238S had the greatest individual effect and combination of E104K and G238S showed clear positive epistasis, similar to effects on specific activity. This relation did not hold in cTEM-2m. cTEM-2m despite a 600-fold improvement in catalytic efficiency showed an increased MIC of only ~4-fold (Table 3.1). Expression of all variants in *E. coli* was similar, but factors inherent to the *in vivo* MIC assay such as altered mRNA or protein stability, folding or periplasmic export may play an additional role. For instance, the correlation between MIC and activity assays increased when kinetic stability, that is aggregation and degradation in native conditions, was considered for BcII metallo-ß-lactamase (4). In the current study, thermostability is unlikely to be in play since TEM-1 and cTEM-2m showed a similar loss of thermostability upon inclusion of E104K/G238S (Table 3.1).

Overall, *in vitro* assays demonstrate that the E104K/G238S combination considerably increases catalytic activity in all three host ß-lactamases. This is observable *in vivo* through MIC assays only for host ß-lactamases TEM-1 and cTEM-17m, indicating that the improvement in catalytic function or expression in host cTEM-2m was insufficient to overcome microbial inhibition with CTX. These considerations will influence the outcome of directed molecular evolution, where cefotaximase function is selected via an *in vivo* process.

3.4.2. Rationalization of catalytic improvement through flexible molecular docking.

The inclusion of E104K and G238S resulted in a similar extent of catalytic improvement in the three host ß-lactamases despite the variation in active site sequence and dynamics at the timescale of cefotaxime turnover. The mechanism by which these mutations provide a synergistic effect in TEM-1 is not yet fully understood (2, 37). Here, we examined whether the mechanism of catalytic improvement was conserved by performing Protein Energy Landscape Exploration (PELE) (65) simulations. PELE is a Monte Carlo approach that combines random localized perturbations of the substrate and protein backbone with side-chain sampling and minimization cycles. Such a combination has been shown to provide robust descriptions of substrate migration and active site induced-fit (66, 67). This is key, as previous research has found that docking to a single structure

does not accurately predict increases in catalytic efficiency (68). Here, the rate-limiting step for cephalosporin hydrolysis is the acylation process (69), which is the first step in the reaction mechanism. As a result, correlation of the enzyme-substrate catalytic distances with the Michaelis-Menten kinetic parameters is pertinent.

Our simulations identify such a correlation with catalytic efficiency. Each of the E104K/G238S variants presents significantly more favorable (lower) interaction energies at the catalytic distance monitored (distance from catalytic S70 to C3 carbon in cefotaxime; Figure 3.3) than their respective non-mutated hosts, reflecting the improved catalytic efficiency of each E104K/G238S variant (Figure 3.4). In TEM-1, cTEM-17m and their E104K/G238S variants, the energy minima observed at catalytically relevant distances (~3 Å) are sharper, showing a more productive exploration of the active site by the substrate than in cTEM-2m and its double mutant (Figure 3.4). The non-mutated TEM-1 has a second energy minimum at ~3.7 Å (almost degenerate with that at ~3 Å). Inclusion of E104K/G238S in TEM-1 produced a greater stabilization of the catalytically productive minimum than the non-productive minimum, contributing to the increase in catalytic efficiency. This is consistent with previous results, where mutations E104K/G238S were shown to reduce the incidence of non-productive conformations (68).

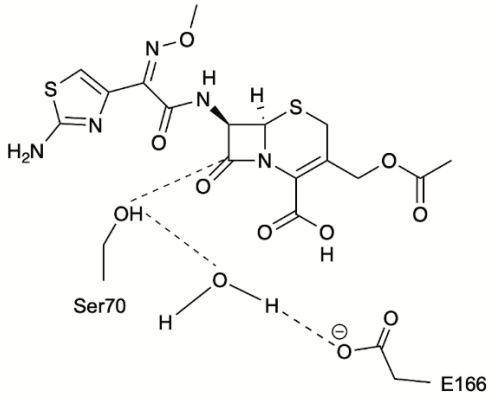


Figure 3.3. – Schematic representation of the catalytic distances monitored during PELE simulations for identification of catalytically relevant frames

S70 hydroxyl oxygen to CTX(C3), S70 hydroxyl hydrogen to catalytic water, and E166 carboxylate oxygen to catalytic water.

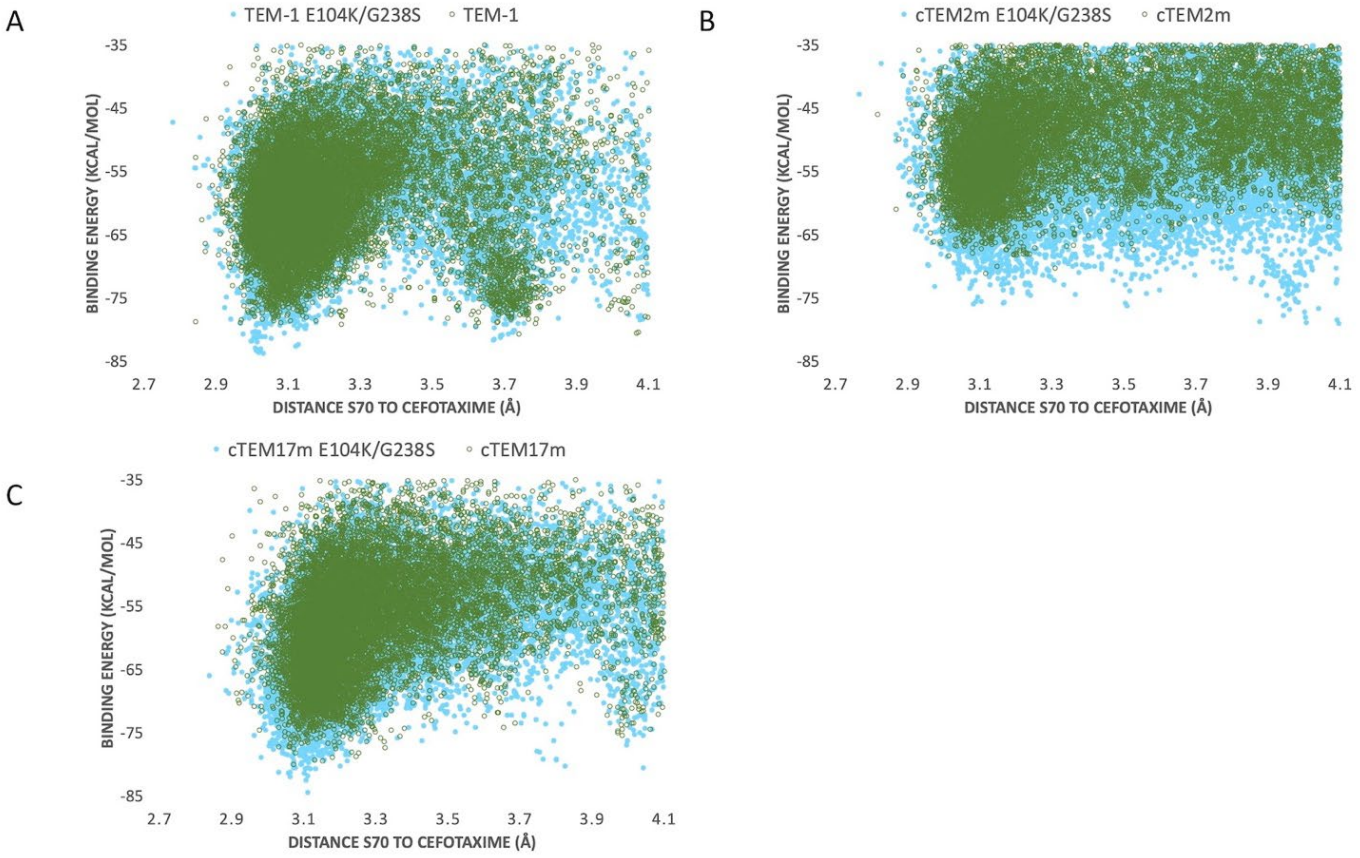


Figure 3.4. – PELE energy profiles for the host ß-lactamases

(A) TEM-1, (B) cTEM-2m and (C) cTEM-17m (empty green circles) and their respective E104K/G238S variant (full blue circles). Binding energy is plotted against distance of S70 hydroxyl oxygen to CTX(C3).

The conformers corresponding to the lowest interaction energies and shortest distances between the catalytic S70 and the CTX ß-lactam ring for each variant were defined as catalytically productive (Figure 3.3). This excluded binding in catalytically disfavored conformations, such as conformers displaying favorable binding energy but Ser-CTX distances incompatible with catalysis (> 3.7 Å) (Figure 3.4). They were extracted using the lowest interaction energies where the catalytic S70 was within 3.5 Å of C3 in the CTX ß-lactam ring. The average catalytic distances are similar (Table 3.2). Then, the 50 best frames with < 3.5 Å between the catalytic water and the

hydroxyl oxygen of S70 as well as a carboxylate oxygen in E166 were selected, again excluding any catalytically disfavored conformation (Figure 3.3). The predicted binding energies of the catalytically productive structures improved upon inclusion of E104K/G238S into each of the three host ß-lactamases (Table 3.2). Despite a 2-fold greater improvement in host cTEM-2m (average 6.1 kcal/mol) than in hosts TEM-1 and cTEM-17m (average ~3 kcal-mol), the predicted binding energy of cTEM-2m E104K/G238S was weaker than the other double mutants. This correlates with its lower catalytic efficiency and poor performance in MIC assays (Table 3.1). Further analysis revealed a new salt bridge in each of the double mutants (Table A1.1 in annex 1). The E104K-E240 salt bridge appeared in 7 out 50 poses in TEM-1 E104K/G238S, 40 out of 50 poses in cTEM-2m E104K/G238S and 12 out of 50 poses in cTEM-17m E104K/G238S. Although it did not appear in all poses, it signals a potential contribution to stabilizing catalytically relevant conformations.

In addition, the inclusion of E104K/G238S decreased the distance between residue 104 and P167 in all variants, by an average of 1.4 Å (1.1 to 1.8 Å; Table 3.2). P167 belongs to the Ω-loop and is the neighbor of the catalytic E166; its van der Waals interaction with K104 has been proposed to stabilize catalytically relevant conformations (68). We note that this stabilization did not significantly modify the backbone distances between residue 104 and P167.

It is interesting to note that cTEM-17m has a larger active site pocket (5000 $Å^3$) than TEM-1 (2200 $Å^3$) and cTEM-2m (2600 $Å^3$). The active site cavity size was greatly reduced in cTEM-17m E104K/G238S (2320 $Å^3$), bringing it in the range of TEM-1 E104K/G238S (2350 $Å^3$) and cTEM-2m E104K/G238S (1980 $Å^3$) which showed smaller variation in cavity size with their respective hosts. The observations made with cTEM-17m demonstrate that variation in the active site cavity volume is tolerated in TEM-1 variants.

Table 3.2. – Analysis of the 50 catalytically productive variants having the lowest interaction energy

		Binding Energy (kcal/mol)	Distances (Å)				E104K-E240 salt bridge (%)
			S70-CTX(C3)	*S70-catalytic water*	*E166-catalytic water*	*104-P167*	
TEM-1	Average	-76.3	3.1	2.7	3.1	6.7	N/A
	Max	-74.9	3.5	2.8	3.3	7.3	
	Min	-78.8	2.8	2.6	2.9	5.7	
TEM-1 E104K/G238S	Average	-79.1	3.1	2.7	3.2	5.6	14
	Max	-77.0	3.5	2.8	3.5	6.8	
	Min	-83.7	2.9	2.7	3.0	5.0	
cTEM-2m	Average	-65.7	3.2	2.7	3.2	7.6	N/A
	Max	-64.1	3.5	2.7	3.5	8.1	
	Min	-70.3	3.0	2.6	3.0	7.1	
cTEM-2m E104K/G238S	Average	-71.8	3.2	2.7	3.1	5.8	80
	Max	-69.9	3.4	3.4	3.4	6.8	
	Min	-76.0	3.0	2.6	2.4	5.1	
cTEM-17m	Average	-76.4	3.2	2.7	2.6	6.8	N/A
	Max	-74.9	3.4	2.7	3.2	7.1	
	Min	-80.1	3.0	2.6	2.4	6.4	
cTEM-17m E104K/G238S	Average	-79.4	3.1	2.7	2.6	5.5	24
	Max	-77.9	3.2	2.9	3.2	6.1	
	Min	-84.5	3.0	2.7	2.4	5.2	

N/A: not applicable.

3.4.3. Directed molecular evolution towards cefotaxime hydrolysis in cTEM-2m and cTEM-17m.

Overall, the combined E104K/G238S substitutions consistently led to important improvements in catalytic efficiency in the three hosts, as reflected in the more favorable energies predicted for their cefotaxime complexes. Nonetheless, the effects of the E104K/G238S substitutions on MIC in *E. coli* were variable. This led us to question whether the evolutionary paths to cefotaxime resistance available to these three hosts could differ, should mutation interactions differ significantly due to initial slow motions.

Having demonstrated that the prevalent E104K/G238S evolutionary solution to cefotaxime resistance in TEM-1 ß-lactamase is valid in variants cTEM-2m and cTEM-17m, we investigated whether other known evolutionary paths towards cefotaxime resistance in TEM-1 (36, 37, 41, 70) are also accessible to cTEM-2m and cTEM-17m host ß-lactamases by performing three rounds of random mutagenesis and selection. TEM-1 has served in the past as a robust model to describe new mutagenesis methodologies (71, 72), deep-mutational scanning (8, 73) and to test protein evolution hypotheses (35), where screening vast libraries for improved cefotaximase activity has proven practical and informative. As a result, a large body of information describing evolutionary paths to higher cefotaximase activity in TEM-1 is available. Prevalent mutations known to confer cefotaxime resistance occur in the Ω-loop region (R164S, R164H, I173V), and its neighbouring regions: 234-244 region (A237T, G238S, E240K) and E104K. Some of these are known to act synergistically as E104K and G238S do, for example, E240K is often encountered in combination with R164S or G238S or A237T in combination with R164S.

We thus investigated evolutionary paths to cefotaximase activity using cTEM-2m and cTEM-17m ß-lactamases as starting points, to determine whether they include the same substitutions as TEM-1 ß-lactamase or if distinct ones are selected. Libraries consisting of $>10^5$ variants with a mutation rate of 1-5 mutations/1000 bp/generation were screened over three generations (Table 3.3). Each generation saw the libraries plated on several concentrations of cefotaxime. Colonies were collected and pooled at the lowest cefotaxime concentration where significant selection was observed. This strategy allowed for inclusion of moderate sequence diversity in each subsequent round of mutagenesis, thereby minimizing the occurrence of evolutionary dead ends.

The mutational process was tracked by sequencing randomly picked colonies throughout the process of laboratory evolution.

Table 3.3. – Mutation rate and library size for three rounds of directed molecular evolution, prior to selection.

	Mutation rate/1000 bp		*Library size (CFU)[a]*	
	cTEM-2m	**cTEM-17m**	**cTEM-2m**	**cTEM-17m**
Generation 1	2.6 (10) [b]	3.3 (16)	1.5×10^5	6.0×10^5
Generation 2	4.9 (13)	4.0 (10)	3.0×10^5	1.7×10^5
Generation 3	9.4 (12)	7.5 (10)	3.7×10^4	2.6×10^5

[a] CFU: colony-forming units.

[b] The number of variants sequenced is in parentheses.

Dominance of one variant was identified following two rounds of evolution for cTEM-2m: variant cTEM-2m(**L40F/G116S**/T160T/**P167L/R178C**; non-synonymous substitutions are in bold) was observed in 7 out of 9 colonies sequenced and was the sole variant identified after selection of the third generation (Table 3.4). PCR biases during the mutagenic process were ruled out as that variant was not predominant prior to selection (1 out of 13 sequences, Table A1.2 in annex 1).

Table 3.4. – Mutations identified upon screening cTEM-2m libraries against 0.016 μg/mL CTX. Non-synonymous mutations are highlighted in bold.

cTEM-2m	**40**	**57**	**91**	**104**	**116**	**118**	**160**	**167**	**169**	**170**	**178**
				E>V		**T>A**					
Generation 1										**N>I**	
		I>P									
Generation 2			**L>V**		**G>S**		T>T	**P>L**	L>L		**R>C**
	L>F				**G>S**		T>T	**P>L**			**R>C**
Generation 3	**L>F**				**G>S**		T>T	**P>L**			**R>C**

Non-synonymous mutations are highlighted in bold

Three generations of evolution were required to observe predominance of a variant of cTEM-17m (Table 3.5). Variant cTEM-17m(P62P/A135A/**K158N**/R164R/**K192N**/K234K) was identified in 10 out of 14 colonies sequenced. Of special interest is that mutations E104K and G238S were individually observed following the selection of the second generation, accompanied by one or

more non-synonymous mutations, but neither became dominant following selection of the third generation (Table 3.5).

Table 3.5. – Mutations identified upon screening cTEM-17m libraries against 0.016 μg/mL CTX

Generation	42	53	62	70	78	91	92	100	104	125	134	145	158	164	170	182	185	188	190	192	197	228	234	238	239	246	249	253	254	256	275
																			L>L								A>C				
															N>N															K>M	
1						L>F										T>T															
					G>G						A>A					T>T		S>G													
			P>P							A>A																			K>R		
				S>S								P>P																			
				S>S													L>I										L>L				R>G
2			P>P							A>A														G>S	E>G					
			P>P							A>A														G>S		I>V					
			P>P						E>K																E>G			G>S			
			P>P							A>A			K>N	R>R						K>N			K>K								
							G>D	N>N																							
3			P>P																	K>N											
	A>A												K>N								E>K										
		S>N								A>A				R>R								G>G							K>R		

Non-synonymous mutations are highlighted in bold

3.4.4. Dominant variants in directed evolution belong to known evolutionary paths

Sampling evolutionary paths towards cefotaxime resistance showed that cTEM-17m could access known prevalent mutations in directed evolution and clinical isolates. The experiment was continued until a variant was predominant in cTEM-2m and cTEM-17m evolution. The dominant variants were sub-cloned and transformed to confirm cefotaxime resistance. Predominance of a variant indicates that these clones contain mutations that provide an advantage with respect to the other clones in the pool when the mutational load is increased. The location and prior reports on these mutations are discussed in this section.

The predominant cTEM-2m(**L40F**/**G116S**/T160T/**P167L**/**R178C**) variant includes two non-synonymous mutations in the Ω-loop that had previously been reported in clinical isolates or directed molecular evolution experiments in TEM-1 (39, 64, 71). In particular, P167 is the neighbor to catalytic residue E166 and is a key residue for folding in TEM-1 (74). The P167L substitution observed here was previously found upon selection for CTX and ceftazidime hydrolysis, accompanied by further mutations known to confer ceftazidime resistance (39). Mutation of R178 to A has been observed in clinical isolate TEM-178, and to C in a multiply mutated variant evolved towards ceftazidime resistance (71). It is interesting to note that both previously reported P167L and R178C variants containing these mutations were accompanied by mutations R164H and I173V, known to confer ceftazidime and cefotaxime resistance individually. Therefore, these mutations might act synergistically with other cephalosporinase-conferring mutations. Residue L40 is mutated to V in clinical isolate TEM-164, or to W with another mutation known to confer ceftazidime resistance (71), yet to our knowledge the L40F mutation has not been observed in the context of resistance. Furthermore, no resistance-conferring mutations of residue G116 have been reported. Deep-mutational scanning data (73) agree that its mutation does not confer a significantly adaptive phenotype suggesting that it is not a significant contributor to CTX resistance here. Despite fixation of variant cTEM-2m(**L40F**/**G116S**/T160T/**P167L**/**R178C**) after two rounds of evolution, it conferred no observable increase in resistance *in vivo* (MIC = 0.016 μg/mL, equal to cTEM-2m). This is consistent with the slight increases in MIC that result from introduction of the synergistic mutations E104K/G238S into cTEM-2m. Overall, cTEM-2m did not evolve cefotaxime resistance in *E. coli* as readily as TEM-1 and cTEM-17m.

The predominant cTEM-17m(P62P/A135A/**K158N**/R164R/**K192N**/K234K) variant conferred a 4-fold increase in MIC compared to cTEM-17m (MIC = 0.11 μg/mL). It includes two non-synonymous mutations, K158N and K192N, that have both been previously identified. A mutation in residue 158 to asparagine defines the clinical isolate TEM-127 (H158N) (64), whereas K192N has been identified along with other mutations conferring cefotaximase activity (37).

Overall, the most prevalent mutations in TEM-1 laboratory experiments and clinical isolates such as E104K, R164S, A237T, G238S or E240K were not fixated. However, evolutionary path sampling did reveal previously reported mutations in cefotaximase variants. And even though it was not fixated, variants containing E104K and G238S were obtained in cTEM-17m evolution, the variant with higher slow dynamics.

3.5. Discussion

Evolution of a new protein function often requires the initial protein to possess at least some promiscuous activity (75). It has been proposed that proteins exist as an ensemble of conformers in an equilibrium that can shift through evolution to favor conformers having promiscuous function (21, 76). It is not known whether higher protein dynamics may facilitate modulating the conformational ensemble, potentially making a protein more amenable to evolve towards new function (77). Prior studies have reported changes in protein dynamics along evolutionary trajectories that were correlated with higher activity, in conjunction with the observation of epistasis in the mutations (16, 27, 30, 32). Protein dynamics may thus promote non-additive interaction of mutations (2). Increasing our understanding of the biophysical features that facilitate protein evolution finds practical applications in better directing protein engineering (6, 21, 27, 28, 77).

We considered that TEM-1 ß-lactamase was a good model to examine the impact of initial dynamics on evolutionary trajectory because ancestral enzyme reconstruction of TEM-1 indicated that higher active site flexibility was correlated with higher promiscuity and evolvability (29). In contrast, a study of TEM-1 variant R164S/G238S, displaying negative epistasis toward function, attributed high Ω-loop flexibility to evolutionary dead-ends (34). Both reports point to a link

between protein dynamics and evolutionary outcomes, and suggest that the relationship is complex and depend on their location and the type of protein motions.

In this study, introduction of the epistatic mutations E104K/G238S in the host ß-lactamase variants cTEM-2m and cTEM-17m maintained their effect on acquisition of *in vitro* cefotaximase activity, despite the variants' differences in protein motions at the timescale of catalytic turnover. The mechanism by which these mutations act synergistically is poorly understood (37, 78). In fact, even the mechanism by which G238S provides such an increase in catalytic activity remains elusive (37). It has been proposed that this mutation increases catalytic activity through a new hydrogen bond between G238S and the oxime oxygen of cephalosporins (79) or by enlargement of the active site either through repositioning of the B3 ß-strand (80) or by repositioning of the Ω-loop (69, 81). More recently, it has been hypothesized that E104K/G238S act synergistically through long-range interactions (2). Moreover, these mutations have been shown to reduce Ω-loop motions, stabilizing a conformer already present in TEM-1 that correlates with higher cefotaximase activity (68). Interestingly, in this first use of PELE to examine cefotaxime binding, we similarly observe the stabilization of catalytic enzyme-substrate poses in the E104K/G238S mutants. Although we did not conclusively determine the mechanism by which these mutations increase catalytic efficiency, a salt bridge observed between the new E240 and K104 in all three E104K/G238S mutants has the potential to stabilize productive enzyme-substrate conformations by forming a fluctuating gate that helps to enclose the substrate within the active site cavity (Figure 3.5). Crystal structure analysis of ligand-free TEM-52 containing E104K/G238S and the stabilizing mutation M182T was also consistent with formation of that salt bridge: E240 was repositioned with its side-chain pointing towards E104K and the distance between those residues was decreased by 2 Å, to 4.8 Å (78). This suggests that the gate is not triggered by substrate binding but may nonetheless contribute to keeping the substrate inside the active site. We note that the occurrence of the salt bridges trends with the improvement in catalytic efficiency of the E104K/G238S variants; cTEM-2m E104K/G238S (600-fold improvement in k_{cat}/K_M) displays the highest occurrence of the salt bridge (80%, Table 3.2). Additionally, we observed reduced distances between residue 104 and P167 in the Ω-loop of the double mutants. This has been proposed to increase the distribution of conformers able to hydrolyze cefotaxime by restricting

Ω-loop motions through van der Waals contacts, thereby stabilizing the Ω-loop (68) (Figure 3.5). Interestingly, the Ω-loop stabilization was observed in all variants despite important slow dynamics and sequence variation occurring in the Ω-loop of cTEM-17m only (Figure 3.1A). The combined effects appear to contribute to the approximately 2-order of magnitude improvement in k_{cat} conferred by the synergistic mutations.

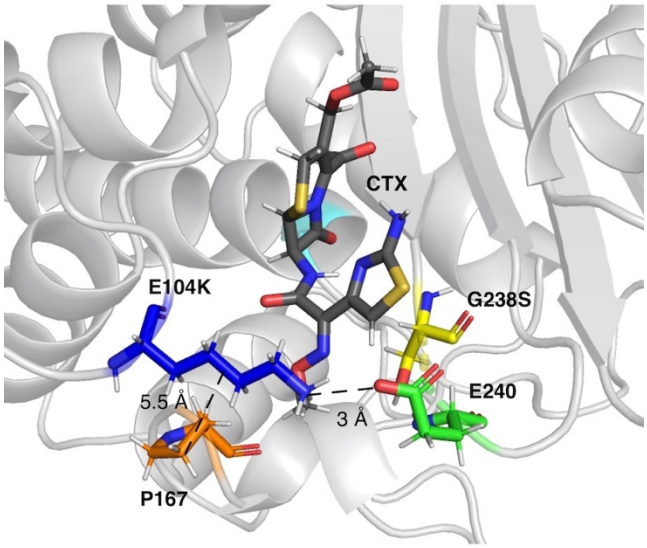


Figure 3.5. – The E104K-E240 salt bridge and E104K-P167 van der Waals interaction in a representative pose of TEM-1 E104K/G238S.

E104K, G238S, E240 and P167 are shown as sticks in dark blue, yellow, green and orange, respectively. S70 is in cyan and CTX in grey.

Considering that the double mutants of cTEM-2m and cTEM-17m show a similar increase in catalytic efficiency as the double mutant of WT TEM-1; that the crystal structures and NMR analysis of those two start-points for evolution are virtually indistinguishable from the WT TEM-1 (22, 48); and that PELE simulations point to populations of statistically stabilized, cefotaximase-competent conformers in the double mutants, consistent with the analysis of Hart (68), we can reasonably infer that the mechanism of increased catalytic efficiency is unchanged albeit incompletely understood in the ß-lactamase variants cTEM-2m and cTEM-17m.

The impact of mutations E104K/G238S on catalytic efficiency, on CTX resistance in *E. coli* (MIC) and on binding energy predicted upon docking CTX was most similar upon introduction into TEM-1 and cTEM-17m. This was unexpected since the latter includes 17 mutations in the active site region relative to TEM-1 and displays vastly increased slow dynamics (Figure 3.1A). Despite cTEM-2m E104K/G238S displaying the strongest increase in catalytic efficiency *in vitro*, with a 180-fold increase in k_{cat} and 600-fold improvement of k_{cat}/K_M, the resulting cefotaximase activity did not suffice to promote high CTX resistance in *E. coli* and yielded a lower binding energy profile. The host ß-lactamase cTEM-2m has only two mutations relative to TEM-1 and displays an intermediate extent of slow dynamics; however, those two mutations – M68L and M69T – are the immediate neighbors of the catalytic S70 and they result in a 50-fold reduction in k_{cat} in cTEM-2m relative to TEM-1. Overall, the inclusion of E104K/G238S in cTEM-2m, despite resulting in the greatest increase in catalytic efficiency (k_{cat}/K_M), did not suffice to procure CTX resistance in *E. coli*.

Similar parallels in evolutionary outcome were observed when sampling evolutionary paths in directed evolution. Mutations E104K, G238S and further mutations reported in clinical isolates of TEM-1 were selected in cTEM-17m libraries (36, 37, 64, 82). Whereas variants selected upon directed evolution of cTEM-2m did not include prevalent cefotaximase mutations, other mutations previously identified in cefotaximase variants were reported (39, 71). However, fixation of a cTEM-2m variant in the third round of directed evolution did not significantly increase CTX resistance as evaluated with MIC, consistent with the failure of the E104K/G238S mutations in the same context. We therefore observe similar evolutionary outcome in *E. coli* for TEM-1 and for host cTEM-17m that displays extensive slow dynamics, despite its possessing greater sequence divergence than cTEM-2m. Furthermore, the strongest increase in catalytic activity observed *in vitro* for cTEM-2m E104K/G238S demonstrates that its differences in slow dynamics do not prevent evolvability; instead, the poor evolutionary outcome of cTEM-2m in *E. coli* appears to result from the deleterious effect of its initial mutations M68L and M69T on kinetics of cefotaxime hydrolysis.

Overall, three vastly differing patterns of protein dynamics at the timescale of catalytic turnover did not impede the high evolvability of TEM-1 ß-lactamase. Its tolerance to extensive protein

dynamics at slow timescales is consistent with the robustness of TEM-1 ß-lactamases and has been hypothesized to facilitate evolution toward the recognition of new substrates (22, 29, 48). Whereas loop movements and active site cavity fluctuations occur at slow timescales and are relevant for catalysis, fast timescales where the formation and breakdown of transition states occur are also important (24). Dynamics at fast timescales are largely conserved in the three host-ß-lactamases used in this study (Figure 3.1A) (22). This backdrop of conserved fast motions and diverse slow motions provides scaffolds that have the potential to evolve towards new protein function. To our knowledge this is the first study using ß-lactamases whose dynamics landscapes have been fully characterized to explore the impact of protein motions at a specific timescale and specific regions on the evolution of new protein function and epistasis.

3.6. Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.7. Author Contributions

LA and JP designed experiments and analyzed data. LA and CL performed protein expression, purification and kinetic experiments. LA performed directed molecular evolution. FSJ designed preliminary computational simulations. VG performed computational simulations. CP provided Python scripts. LA, JP and VG wrote the manuscript. All authors revised and approved the final version of the manuscript.

3.8. Funding

This work was supported by operating grant RGPIN-2018-04686 to JNP from the Natural Science and Engineering Research Council of Canada (NSERC), infrastructure grant 11510 to JNP from the Canada Foundation for Innovation and operating grant PID2019-106370RB-I00 to VG from the Spanish Ministry of Science and Innovation. LA and CLSD are grateful to FQRNT, Université de Montréal and NSERC for scholarships.

3.9. Acknowledgments

We thank Pouria Dasmeh, Adrian Serohijos and Nicolas Doucet for fruitful discussions, Adem Hadjabdelhafid-Parisien for technical support and Daniela Quaglia for proofreading.

3.10. References

1. Carneiro M, Hartl DL. Adaptive landscapes and protein evolution. Proc Natl Acad Sci USA. 2010;107(suppl 1):1747-51.

2. Miton CM, Tokuriki N. How mutational epistasis impairs predictability in protein evolution and design: How Epistasis Impairs Predictability in Enzyme Evolution. Protein Sci. 2016;25(7):1260-72.

3. Starr TN, Thornton JW. Epistasis in protein evolution. Protein Sci. 2016;25(7):1204-18.

4. Meini MR, Tomatis PE, Weinreich DM, Vila AJ. Quantitative Description of a Protein Fitness Landscape Based on Molecular Features. Mol Biol Evol. 2015;32(7):1774-87.

5. Reetz MT. The importance of additive and non-additive mutational effects in protein engineering. Angew Chem Int Ed Engl. 2013;52(10):2658-66.

6. Yang G, Miton CM, Tokuriki N. A mechanistic view of enzyme evolution. Protein Sci. 2020;29(8):1724-47.

7. Klesmith JR, Bacik J-P, Wrenbeck EE, Michalczyk R, Whitehead TA. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. Proc Natl Acad Sci USA. 2017;114(9):2265-70.

8. Firnberg E, Labonte JW, Gray JJ, Ostermeier M. A Comprehensive, High-resolution map of a gene's fitness landscape. Mol Biol Evol. 2014;31(6):1581-92.

9. Shin H, Cho BK. Rational Protein Engineering guided by deep mutational scanning. Int J Mol Sci. 2015;16(9):23094-110.

10. Olson CA, Wu Nicholas C, Sun R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Curr Biol. 2014;24(22):2643-51.

11. Sato TK, Tremaine M, Parreiras LS, Hebert AS, Myers KS, Higbee AJ, et al. Directed evolution reveals unexpected epistatic interactions that alter metabolic regulation and enable anaerobic xylose use by Saccharomyces cerevisiae. PLoS Genet. 2016;12(10):e1006372.

12. Poelwijk FJ, Socolich M, Ranganathan R. Learning the pattern of epistasis linking genotype and phenotype in a protein. Nat Commun. 2019;10(1):4213.

13. Acevedo-Rocha CG, Hoebenreich S, Reetz MT. Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. Methods Mol Biol. 2014;1179:103-28.

14. Salverda MLM, Dellus E, Gorter FA, Debets AJM, Oost Jvd, Hoekstra RF, et al. Initial mutations direct alternative pathways of protein evolution. PLoS Genet. 2011;7(3):e1001321.

15. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature. 2007;445(7126):383-6.

16. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, et al. The role of protein dynamics in the evolution of new enzyme function. Nat Chem Biol. 2016;12(11):944-50.

17. Doucet N, Savard PY, Pelletier JN, Gagne SM. NMR Investigation of Tyr105 mutants in TEM-1 beta-lactamase: Dynamics are corrleated with function. J Biol Chem. 2007;282(29):21448-59.

18. Duff MR, Jr., Borreguero JM, Cuneo MJ, Ramanathan A, He J, Kamath G, et al. Modulating enzyme activity by altering protein dynamics with solvent. Biochemistry. 2018;57(29):4263-75.

19. Osuna S, Jimenez-Oses G, Noey EL, Houk KN. Molecular dynamics explorations of active site structure in designed and evolved enzymes. Acc Chem Res. 2015;48(4):1080-9.

20. Eisenmesser EZ, Bosco DA, Akke M, Kern D. Enzyme dynamics during catalysis. Science. 2002;295(5559):1520-3.

21. Maria-Solano MA, Serrano-Hervás E, Romero-Rivera A, Iglesias-Fernández J, Osuna S. Role of conformational dynamics in the evolution of novel enzyme function. Chem Commun. 2018;54(50):6622-34.

22. Gobeil SMC, Ebert MCCJC, Park J, Gagné D, Doucet N, Berghuis AM, et al. The structural dynamics of engineered β-lactamases vary broadly on three timescales yet sustain native function. Sci Rep. 2019;9(1):6656.

23. Pandya MJ, Schiffers S, Hounslow AM, Baxter NJ, Williamson MP. Why the energy landscape of barnase is hierarchical. Front Mol Biosci. 2018;5:115.

24. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature. 2007;450(7172):964-72.

25. Liu Y, Bahar I. Sequence evolution correlates with structural dynamics. Mol Biol Evol. 2012;29(9):2253-63.

26. Petrovic D, Risso VA, Kamerlin SCL, Sanchez-Ruiz JM. Conformational dynamics and enzyme evolution. J R Soc Interface. 2018;15(144).

27. Johansson KE, Lindorff-Larsen K. Structural heterogeneity and dynamics in protein evolution and design. Curr Opin Struct Biol. 2018;48:157-63.

28. Gardner JM, Biler M, Risso VA, Sanchez-Ruiz JM, Kamerlin SCL. Manipulating conformational dynamics to repurpose ancient proteins for modern catalytic functions. ACS Catal. 2020;10(9):4863-70.

29. Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. Mol Biol Evol. 2015;32(1):132-43.

30. Otten R, Liu L, Kenner LR, Clarkson MW, Mavor D, Tawfik DS, et al. Rescue of conformational dynamics in enzyme catalysis by directed evolution. Nat Commun. 2018;9(1):1314.

31. Tomatis PE, Fabiane SM, Simona F, Carloni P, Sutton BJ, Vila AJ. Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. Proc Natl Acad Sci U S A. 2008;105(52):20605-10.

32. Gonzalez MM, Abriata LA, Tomatis PE, Vila AJ. Optimization of conformational dynamics in an epistatic evolutionary trajectory. Mol Biol Evol. 2016;33(7):1768-76.

33. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. Nature. 2006;444(7121):nature05385.

34. Dellus-Gur E, Elias M, Caselli E, Prati F, Salverda MLM, de Visser JAGM, et al. Negative epistasis and evolvability in TEM-1 β-lactamase—The thin line between an enzyme's conformational freedom and disorder. J Mol Biol. 2015;427(14):2396-409.

35. Bershtein S, Tawfik DS. Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. Mol Biol Evol. 2008;25(11):2311-8.

36. Barlow M, Hall BG. Predicting evolutionary potential: In vitro evolution accurately reproduces natural evolution of the TEM β-lactamase. Genetics. 2002;160(3):823-32.

37. Salverda MLM, De Visser JAGM, Barlow M. Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. FEMS Microbiol Rev. 2010;34(6):1015-36.

38. Guthrie VB, Allen J, Camps M, Karchin R. Network models of TEM beta-lactamase mutations coevolving under antibiotic selection show modular structure and anticipate evolutionary trajectories. PLoS Comput Biol. 2011;7(9):e1002184.

39. Schenk MF, Szendro IG, Salverda MLM, Krug J, de Visser JAGM. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. Mol Biol Evol. 2013;30(8):1779-87.

40. Clouthier CM, Morin S, Gobeil SMC, Doucet N, Blanchet J, Nguyen E, et al. Chimeric β-lactamases: Global conservation of parental function and fast time-scale dynamics with increased slow motions. PLoS One. 2012;7(12):e52283.

41. Palzkill T. Structural and mechanistic basis for extended-spectrum drug-resistance mutations in altering the specificity of TEM, CTX-M, and KPC β-lactamases. Front Mol Biosci. 2018;5.

42. Wang X, Minasov G, Shoichet BK. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol. 2002;320(1):85-95.

43. Morin S, Clouthier CM, Gobeil S, Pelletier JN, Gagné SM. Backbone resonance assignments of an artificially engineered TEM-1/PSE-4 Class A β-lactamase chimera. Biomol NMR Assign. 2010;4(2):127-30.

44. Laible M, Boonrod K. Homemade site directed mutagenesis of whole plasmids. J Vis Exp. 2009(27)

45. Sambrook J, Russell DW. The inoue method for preparation and transformation of competent e. Coli: "ultra-competent" cells. CSH Protoc. 2006;2006(1).

46. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene. 1989;77(1):51-9.

47. Copp JN, Hanson-Manful P, Ackerley DF, Patrick WM. Error-prone PCR and effective generation of gene variant libraries for directed evolution In: Gillam E. CJ, Ackerley D., editor. Directed Evolution Library Creation Methods in Molecular Biology. 1179. 2014 ed2014. p. 3-22.

48. Gobeil SMC, Clouthier CM, Park J, Gagne D, Berghuis AM, Doucet N, et al. Maintenance of native-like protein dynamics may not be required for engineering functional proteins. Chem Biol. 2014;21(10):1330-40.

49. Wiegand I, Hilpert K, Hancock RE. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nat Protoc. 2008;3(2):163-75.

50. Knies J, Cai F, Weinreich DM. Enzyme efficiency but not thermostability drives cefotaxime resistance evolution in TEM-1 β-lactamase. Mol Biol Evol. 2017:msx053.

51. Maryam L, Khan AU. Synergistic effect of doripenem and cefotaxime to inhibit CTX-M-15 type beta-Lactamases: Biophysical and microbiological views. Front Pharmacol. 2017;8:449.

52. Taibi-Tronche P, Massova I, Vakulenko SB, Lerner SA, Mobashery S. Evidence for structural elasticity of class A β-lactamases in the course of catalytic turnover of the novel cephalosporin cefepime. JACS. 1996;118(32):7441-8.

53. Vakulenko SB, Taibi-Tronche P, Tóth M, Massova I, Lerner SA, Mobashery S. Effects on substrate profile by mutational substitutions at positions 164 and 179 of the class A TEM pUC19 β-lactamase from Escherichia coli. J Biol Chem. 1999;274(33):23052-60.

54. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007;2(9):2212-21.

55. Atanasov BP, Mustafi D, Makinen MW. Protonation of the beta-lactam nitrogen is the trigger event in the catalytic action of class A beta-lactamases. Proc Natl Acad Sci USA. 2000;97(7):3160-5.

56. Ditchfield R, Hehre WJ, Pople JA. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. J Chem Phys. 1971;54(2):724-8.

57. Kim K, Jordan K. Comparison of density functional and MP2 calculations on the water monomer and dimer. J Phys Chem. 1994;98(40):10089-94.

58. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theory Comput. 2010;6(5):1509-19.

59. Lecina D, Gilabert JF, Guallar V. Adaptive simulations, towards interactive protein-ligand modeling. Scientific Reports. 2017;7(1):8466.

60. Gilabert JF, Lecina D, Estrada J, Guallar V. Monte Carlo techniques for drug design: The success case of PELE. Biomolecular simulations in structure-based drug discovery. Methods and Principles in Medicinal Chemistry. 2018 ed2018. p. 87-103.

61. Yu Z, Jacobson MP, Josovitz J, Rapp CS, Friesner RA. First-shell solvation of ion pairs: Correction of systematic errors in implicit solvent models. J Phys Chem B. 2004;108(21):6643-54.

62. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33-8, 27-8.

63. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinformatics. 2009;10:168.

64. NCBI. Bacterial Antimicrobial Resistance Reference Gene Database: [Available from: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047.

65. Borrelli KW, Vitalis A, Alcantara R, Guallar V. PELE: Protein energy landscape exploration. a novel Monte Carlo based technique. J Chem Theory Comput. 2005;1(6):1304-11.

66. Acebes S, Fernandez-Fueyo E, Monza E, Lucas MF, Almendral D, Ruiz-Dueñas FJ, et al. Rational enzyme engineering through biophysical and biochemical modeling. ACS Catal. 2016;6(3):1624-9.

67. Carro J, González-Benjumea A, Fernández-Fueyo E, Aranda C, Guallar V, Gutiérrez A, et al. Modulating fatty acid epoxidation vs hydroxylation in a fungal peroxygenase. ACS Catal. 2019;9(7):6234-42.

68. Hart KM, Ho CMW, Dutta S, Gross ML, Bowman GR. Modelling proteins' hidden conformations to predict antibiotic resistance. Nat Commun. 2016;7:12965.

69. Saves I, Burlet-Schiltz O, Maveyraud L, Samama J-P, Prome J-C, Masson J-M. Mass spectral kinetic study of acylation and deacylation during the hydrolysis of penicillins and cefotaxime by beta-lactamase TEM-1 and the G238S mutant. Biochemistry. 1995;34(37):11660-7.

70. Weinreich DM. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 2006;312(5770):111-4.

71. Fujii R, Kitaoka M, Hayashi K. One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Res. 2004;32(19):e145.

72. Firnberg E, Ostermeier M. PFunkel: efficient, expansive, user-defined mutagenesis. PLoS One. 2012;7(12):e52031.

73. Stiffler Michael A, Hekstra Doeke R, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell. 2015;160(5):882-92.

74. Vanhove M, Raquet X, Palzkill T, Pain RH, Frère J-M. The rate-limiting step in the folding of the cis-Pro167Thr mutant of TEM-1 β-lactamase is the trans to cis isomerization of a non-proline peptide bond. Proteins. 1996;25(1):104-11.

75. Peisajovich SG, Tawfik DS. Protein engineers turned evolutionists. Nat Methods. 2007;4(12):991-4.

76. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science. 2009;324(5924):203-7.

77. Trudeau DL, Tawfik DS. Protein engineers turned evolutionists-the quest for the optimal starting point. Curr Opin Biotechnol. 2019;60:46-52.

78. Orencia MC, Yoon JS, Ness JE, Stemmer WPC, Stevens RC. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nat Struct Biol. 2001;8(3):238-42.

79. Raquet X, Vanhove M, Lamotte-Brasseur J, Goussard S, Courvalin P, Frère JM. Stability of TEM β-lactamase mutants hydrolyzing third generation cephalosporins. Proteins: Structure, Function, and Bioinformatics. 1995;23(1):63-72.

80. Knox JR. Extended-spectrum and inhibitor-resistant TEM-type beta-lactamases: mutations, specificity, and three-dimensional structure. Antimicrob Agents Chemother. 1995;39(12):2593-601.

81. Cantu C, Palzkill T. The Role of Residue 238 of TEM-1 β-Lactamase in the hydrolysis of extended-spectrum antibiotics. J Biol Chem. 1998;273(41):26603-9.

82. Palzkill T. Structural and mechanistic basis for extended-spectrum drug-resistance mutations in altering the specificity of TEM, CTX-M, and KPC β-lactamases. Frontiers in Molecular Biosciences. 2018;5.

Chapter 4 – Holistic engineering of Cal-A lipase chain-length selectivity identifies triglyceride binding hot-spot

Preface to Chapter 4

Enzyme engineering is a tool that can be used to evolve proteins towards new functions, as well as to expand our knowledge of the systems under study. It can be achieved through the introduction of mutations in the genes coding for a specific protein. One way to do so is through directed evolution, in which mutations are introduced at random throughout the coding gene or in specific regions. This approach often reveals areas that do not tolerate mutations, interactions or mutational hot-spots that could not have been anticipated by knowing the structure alone.

In this chapter we use the industrially relevant enzyme Cal-A lipase, whose biophysical properties are poorly known. Currently, there are two crystal structures of this enzyme (PDB IDs 2veo and 3guu). Their analysis has shown that Cal-A has rare structural features such as a tyrosine in the oxyanion hole, a cap domain and a small active site flap. The existence of an acyl-binding tunnel has been suggested for this enzyme, although the mechanism for substrate recognition remains to be studied.

In order to explore the relevant regions implicated substrate recognition, the gene encoding Cal-A lipase was subjected to region-focused mutagenesis in three segments. These were subsequently recombined seamlessly to regenerate the full-length gene and the resulting mutants were screened against triglycerides of different chain-length. This focused approach allowed us to identify a mutational hot-spot in the putative acyl-binding tunnel, showing that this is the main tunnel implicated in binding the fatty acid moiety of the triglycerides.

The results of this chapter were published in 2019 in the journal PLoS ONE. This study was conceptualized and supervised by postdoctoral Fellow Dr. Daniela Quaglia and Prof. Joelle Pelletier. The experimental portion of the study was carried out by Dr. Quaglia and undergraduate intern Sara Ouadhi. Specifically, Golden-gate mutagenesis, screening, sequencing and lysate characterization was performed by them. My contribution to this manuscript was in

the analysis and discussion of genotype-phenotype relations of Cal-A variants. MSc candidate Olivier Rousseau and myself contributed to the data curation and analysis of sequencing results from random mutagenesis libraries. O. Rousseau designed a script to extract mutations from sequencing results, whereas I designed a script to correlate and map mutations with the phenotype observed in the experimental results. The original draft was written by Dr. Quaglia. I contributed in writing results, discussion, conclusions, to create figures 4.1, 4.3, 4.4 and supplemental figures, as well as to review and edit the overall manuscript. Prof. Pelletier reviewed and edited the final manuscript together with Dr. Quaglia. All authors reviewed and revised the manuscript.

Article 3. Holistic engineering of Cal-A lipase chain-length selectivity identifies triglyceride binding hot-spot

Daniela Quaglia[1,2], Lorea Alejaldre[2,3], Sara Ouadhi[1,2], Olivier Rousseau[1,2] and Joelle N. Pelletier[1,2,3]*

[1] Chemistry Department and Center in Green Chemistry and Catalysis (CGCC), Université de Montréal, Montréal, QC, Canada H3T 1J4

[2] PROTEO, The Québec Network for Research on Protein Function, Engineering and Applications, Québec, QC, Canada G1V 0A6

[3] Biochemistry Department, Université de Montréal, Montréal, Québec, Canada

***Correspondence:**

Joelle Pelletier

joelle.pelletier@umontreal.ca

Keywords: biocatalysis, cal-a lipase, fatty acid hydrolysis, golden gate, library screening, hot-spot identification

2019 PLoS ONE

4.1. Abstract

Through the application of a region-focused saturation mutagenesis and randomization approach, protein engineering of the Cal-A enzyme was undertaken with the goal of conferring new triglyceride selectivity. Little is known about the mode of triglyceride binding to Cal-A. Engineering Cal-A thus requires a systemic approach. Targeted and randomized Cal-A libraries were created, recombined using the Golden Gate approach and screened to detect variants able to discriminate between long-chain (olive oil) and short-chain (tributyrin) triglyceride substrates using a high-throughput *in vivo* method to visualize hydrolytic activity. Discriminative variants were analyzed using an in-house script to identify predominant substitutions. This approach allowed identification of variants that exhibit strong discrimination for the hydrolysis of short-chain triglycerides and others that discriminate towards hydrolysis of long-chain triglycerides. A clear pattern emerged from the discriminative variants, identifying the 217-245 helix-loop-helix motif as being a hot-spot for triglyceride recognition. This was the consequence of introducing the entire mutational load in selected regions, without putting a strain on distal parts of the protein. Our results improve our understanding of the Cal-A lipase mode of action and selectivity. This holistic perspective to protein engineering, where parts of the gene are individually mutated and the impact evaluated in the context of the whole protein, can be applied to any protein scaffold.

4.2. Introduction

Enzyme engineering is widely used to transform enzymes into better catalysts for improved specificity, selectivity, stability, and new chemistry (1-4). Methods available to produce the genetic diversity necessary to discover variants with improved characteristics range from focused point mutations to a completely randomized approach. The diversified library of variants is then screened for the desired properties (5-9). Many of these advances attempt to address two critical limitations to efficient enzyme engineering: the need to reduce library sizes by pinpointing mutational hot-spots where engineering will have the most effect, and the unavailability of effective high-throughput screening methods that directly elicit the desired property, rather than a proxy to that property. It is of paramount importance for advancement in the area to develop

better approaches for the creation of smart libraries of mutants (10, 11) and tools for direct screening.

Smart library creation is currently best addressed either by performing series of sequential and converging mutagenesis campaigns on the whole gene of interest, or by inclusion of computational predictions based on structure and sequence alignments (5, 6, 8, 9, 12-15). We tackle the problem from a dynamic and more holistic perspective, applying region-focused saturation mutagenesis and randomization with seamless recombination (11). In this approach, individual regions of the gene of interest are treated for mutagenesis and are readily recombined with other mutated or native regions. This enables customized generation of mutations in each region, followed by recombination into the full-length coding sequence through assembly using the Golden Gate method (11, 16-18). The flexibility of the method resides in the ability of introducing different patterns of mutations in distinct regions of the protein: each part can be mutated randomly or using a more targeted approach such as saturation mutagenesis.

Demonstrated at a small scale in previous work (11), here we validate this flexible, combinatorial approach by expanding the engineered libraries. Then, we demonstrate that this approach readily defines mutational hot-spots in the chosen target protein (*Candida antarctica Lipase A*, Cal-A). We rapidly discovered improved enzyme variants that were not restricted to point mutants but considered possible synergistic effects between substitutions (19). We further demonstrated that the variants allowed individuation of point mutants of interest, providing an important body of information on the target enzyme, Cal-A.

Growing industrial interest in the development of biocatalysts for selective hydrolysis of fatty acids from naturally sourced triglycerides has stimulated research and development concerning lipases. Cal-A has been investigated for a range of biocatalytic applications, such as using the native Cal-A to concentrate ω-3 polyunsaturated fatty acids (PUFA) from marine resources (20, 21) or engineering variants for enhanced *trans*-selectivity (22, 23). Lipases with the ability to hydrolyze fatty acids of specific chain-lengths are also highly desirable, particularly for the food industry (24, 25). For instance, medium-chain fatty acids constitute high-value food additives as

they provide quick access to energy. Furthermore, health benefits have been associated with milk-fat products rich in diglycerides composed of short-chain saturated fatty acids (24-26).

Cal-A is thermostable (retains catalytic activity at > 90 ˚C), stable at acidic pH and accepts a wide variety of substrates including sterically hindered ones. Furthermore, it differs from the majority of lipases in that it does not undergo extensive movement of a lid domain when undergoing interfacial activation (22, 27, 28). Interfacial activation is a fundamental property of most lipases, where association with lipids is required for a large conformational change that provides access to the otherwise masked active site. These qualities reduce the complexity of the system and make Cal-A a good target for engineering new properties. Prior reports interpreted a co-crystallized PEG-4 molecule as being a fortuitous substrate surrogate, which allowed to identify a docking region for *p*-NO_2-phenyl derivatives of fatty acids. This was supported by the demonstration of altered chain-length selectivity upon mutating that region (29). Specifically, mutations Gly237Lys/Val/Tyr and Val290Trp near the PEG-4 procured enhanced hydrolysis of medium-chain (C6-C10) and long-chain (C22) fatty acid substrates (22, 23, 30), demonstrating the potential for chain-length alteration in Cal-A.

In this work, we engineered Cal-A using the region-focused, random and targeted mutagenesis approach to enhance its discriminative capability to hydrolyze short- (tributyrin, C-4) *vs* long-chain (olive oil, approximately 80% C-18 and 20% C-16) fatty acid esters (31), obtaining variants displaying either selectivity. Identification of variants with the desired selectivities was made possible by applying a high-throughput, whole-cell visual screen on solid medium containing tributyrin or olive oil (11). A particular emphasis was placed on mutations in the PEG-4-binding region (29). Finally, we developed a straightforward script to visualize the sites of mutation associated with either discriminative phenotype. By this computational approach, a large body of mutational data was readily analyzed and clearly revealed the main hot-spot for modulation of substrate binding. Quantitative analysis of activity was achieved for some of the variants with a short- and long-chained *p*-NO_2-phenyl esters substrates, the industry standard for monitoring lipase activity. This confirms that variants with a markedly higher activity for both triglycerides and *p*-NO_2-phenyl derivatives of short- or long- chain fatty acids were identified.

4.3. Results and Discussion

4.3.1. Library generation and characterization

Cal-A was treated for random mutagenesis in three segments (referred to as 'parts') (Figure 4.1). We previously reported that the separation of Cal-A into three parts had facilitated random mutagenesis of Part 2 (residues 211-350) in isolation from the *N*-terminal Part 1 and *C*-terminal Part 3 (11). Part 2 includes a putative tunnel, hypothesized to bind the long chain of a fatty acid substrate (22). Following random mutation, it had been seamlessly recombined with the wild-type Parts 1 and 3 to give 'Random 2' library (Table 4. 1) (11). Recombination was achieved with the Golden Gate-based method that is ideally suited for inserting randomized segments into specific parts of a gene (11, 16, 17). In separate experiments, residues in Parts 1 and 3 that could potentially alter the discriminative hydrolysis had been individuated according to structural analysis: Tyr93, positioned near the active site, was selected for its potential ability to modulate substrate access; the sterically hindered Tyr183 appeared to open a putative tunnel during molecular dynamics simulations (11), whereas Phe431 is a hypothetical gate keeper, modulating substrate entry (27). These three positions were mutated using NDT saturation mutagenesis (N = A, C, G, T, and D = A, G or T; covering 12 codons, 12 amino acids) and recombined with the appropriate wild-type or mutated parts. Screening yielded between 10-40% of clones in the various libraries with improved discrimination towards either short- or long-chain fatty acid esters relative to wild-type Cal-A. This validated the method used to create the desired functional diversity, although the identity of variants of interest had not been determined.

Encouraged by the promising functional diversity that can be accessed and seeking further improvements in discrimination, we explored the sequence space of Cal-A more widely by applying broader randomization and seamless recombination. We complemented the existing libraries (11) by randomizing Part 1 and Part 3 (Figure 4.1). Each randomized part was integrated into the native gene, yielding libraries 'Random 1' and 'Random 3', respectively (Table 4.1) (11). The randomly mutated parts 1, 2 and 3 were also recombined together, yielding a reassembled, fully randomized Cal-A library (library 'Random Rec').

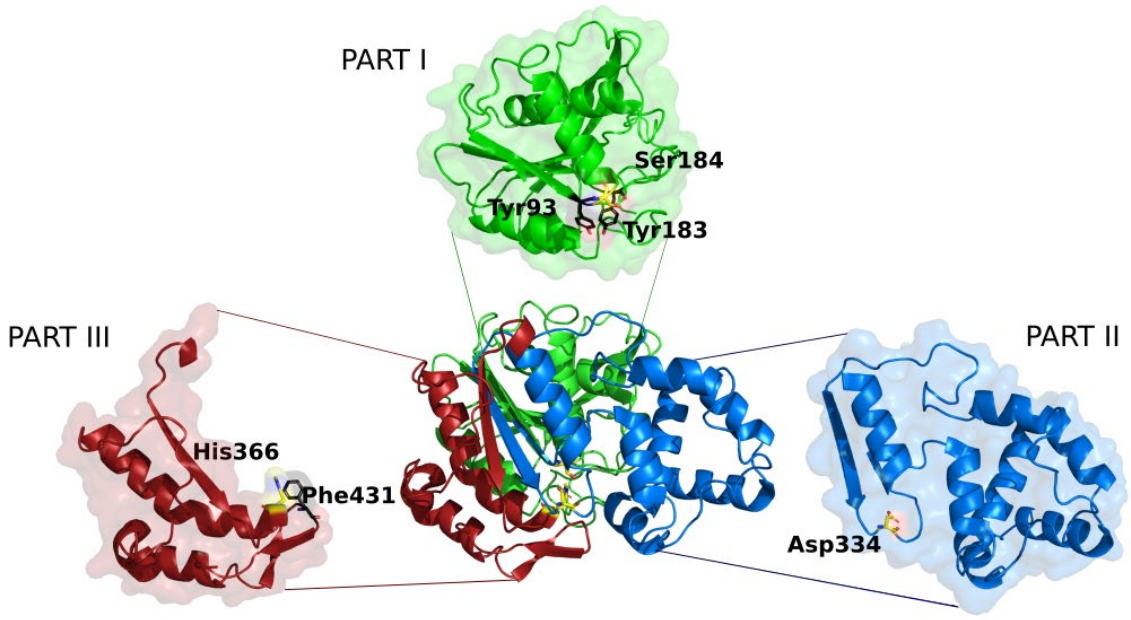


Figure 4.1. – Cal-A was treated in three segments for mutagenesis.

Parts 1 (encoding the N-terminal 210 residues), 2 (encoding residues 211 to 350) and 3 (C-terminal residues 351 to 446 plus 6His-tag) were individually randomized and recombined with wild-type or mutated parts. Parts 1 and 3 include residues treated by NDT saturation mutagenesis (Tyr93, Tyr183 and Phe 431), shown in black. The catalytic triad (Ser184, Asp334, His366) is shown in yellow. Image generated using PDB coordinates 2VEO (29).

Briefly, the parts of Cal-A were synthesized by ATUM (formerly DNA2.0) and delivered in plasmid pM269; these part-carrying constructs are referred to as "mother vectors". Reassembly of the parts using the Golden Gate strategy (through the use of *Bsa*I and *Sap*I type IIS restriction enzymes) was carried out in an Electra expression vector, referred to as a 'daughter vector' (pD441pelB, DNA2.0), where the clones gene is under the control of T5 promoter and carries a pelB leader sequence for periplasmatic expression, as previously described (11). For randomization, the parts were subjected to error-prone mutagenesis in the mother vectors and reassembled in the daughter vector with the complementary native parts, to obtain libraries Random 1 and Random 3. The randomized Parts 1 and 3 were also assembled with Random 2 (previously obtained) in the daughter vector to give the fully randomized Random Rec. In parallel, we also randomized the entire gene as a single entity (not in parts; library 'Random Tot', Table 4. 1) to compare the efficiency of randomizing and reassembling separate parts with the complete

randomization of the gene. DNA sequencing confirmed the quality of the newly obtained libraries (Table 4. 1).

Table 4.1. – Library generation and assessment of library quality.

Library	Randomization method	Mutations per gene segment /kb	Mutations per reassembled gene [a]	Total bp sequenced in Cal-A coding region
Random 1 (605 bp)	Randomized Part 1 + WT Parts 2,3	2.1	1.3	11000
Random 2[b] (420 bp)	Randomized Part 2 + WT Parts 1,3	4.1	1.7	8220
Random 3 (286 bp)	Randomized Part 3 + WT Parts 1,2	3.3	1	4480
Random Rec	Assembly of randomized Parts 1,2, 3	3.0	3.9	18304
Random Tot	Randomization of full coding sequence	4.2	5.5	19155

[a] Length of the reassembled gene: 1311 bp.

[b] Random 2 library was previously reported (11). Additional DNA sequencing was performed in this study.

The number of mutations obtained per individually mutated part (hence per gene, when reassembled with native parts) during randomization of Parts 1 and 3 (libraries Random 1 and 3, respectively) is comparable: an average of 1.3 and 1 mutation per gene (Table 4. 1). This mutation rate is similar to that previously obtained for randomization of Part 2 (library Random 2, 1.7

mutations per gene) (11). Libraries Random Rec and Random Tot have a greater mutational load because the entire length of the gene is randomized (either by full gene randomization or by recombination of randomized parts): 5.5 and 3.9 mutations per gene, respectively. It is satisfying to note that the mutational load of the Random Rec library, obtained by recombination of libraries Random 1, 2 and 3, is essentially equal to the sum of the average mutations of each part (1.3 +1.7 +1 = 4.0). This confirms the validity of the combinatorial assembly method as a reliable tool to obtain quality libraries.

4.3.2. Screening for discriminative hydrolysis of short- or long-chain esters

The randomized libraries were screened towards two classes of substrates: triglycerides and activated *p*-NO_2-phenyl-esters of different lengths. Triglycerides are the natural substrates of Cal-A and their hydrolysis is the ultimate industrial goal. We previously reported a rapid and robust automated method for screening against triglycerides on solid medium plates to procure a high-throughput platform for primary screening of libraries (11). Here, solid emulsions containing either tributyrin or olive oil show haloes surrounding colonies that hydrolyze that substrate. For tributyrin emulsions, clear haloes are observed; olive oil emulsions require the addition of rhodamine where the change of pH upon hydrolysis gives rise to increased fluorescence. Observation of halo formation on growth medium supplemented with emulsified triglycerides is amenable to high-throughput screening yet provides only a qualitative assessment. Selectivity was determined by comparing activity against individual substrates; we note that this method is incompatible with multi-substrate competition (32, 33). *p*-NO_2-Phenyl-esters were used as a secondary substrate in this study: assays using these substrates despite providing an indirect assessment of activity toward triglycerides, they are rapid, quantitative and practical and as thus widely used as standards to assess lipase activity ((34) and references therein). They are single-chain fatty acid esters rather than glycerol triesters, and the *p*-NO_2-substituent activates them toward hydrolysis. Despite differing significantly from native triglycerides in structure and in reactivity, they allow for direct quantitative activity measurements by spectrophotometry and provide a benchmark for comparison with current literature of the reactivity of key variants towards specific fatty acid esters.

We performed automated high-throughput screening on solid medium to evaluate the selectivity of Cal-A variant libraries towards the hydrolysis of short-chain triglycerides (tributyrin) and long-chain fatty acid esters (olive oil, which contains ≈ 70% oleic acid in a triglyceride form) (11). Briefly, rectangular dishes were cast with agar growth medium containing either tributyrin or olive oil/rhodamine to form an opaque emulsion. Using a liquid handler, cultures of *E. coli* BL21(DE3) carrying individual variants from the Cal-A libraries and propagated in 96 deep-well plates were spotted onto the solid media. The inoculated plates were incubated at 30°C overnight (approximately 16 hrs) to obtain colonies of 0.2 - 0.5 cm in diameter. The solid emulsions containing either tributyrin or olive oil show haloes surrounding colonies that hydrolyze that substrate. For tributyrin emulsions, clear haloes are observed; olive oil emulsions require the addition of rhodamine where the change of pH upon hydrolysis gives rise to increased fluorescence (11, 35, 36). Variants of interest were selected based on strong hydrolysis of one substrate and weak or no hydrolysis of the other (Table 4. 2).

Table 4.2. – Discriminative variants obtained upon screening the Cal-A libraries against short- and long-chain triglycerides.

Library	Clones screened	Active[a]	Discriminative towards	
			short-chain	long-chain
Random 1	384	89%	0	0
Random 2	384	66%	18	3
Random 3	384	89%	0	0
Random Rec	768	41%	12	6
Random Tot	768	23%	10	4

[a] Variants were screened for halo formation around colonies spotted via automation on solid medium. Variants were considered active only when halo formation was unequivocal.

Despite specific variants of Part 1 and Part 3 having been identified as being discriminative in point mutant libraries (refer to previous work (11), and below), none were identified upon

randomization of Parts 1 and 3, suggesting that there are few discriminative possibilities in those segments of Cal-A. In contrast, all libraries that include randomization of Part 2 (libraries Random 2, Random Rec and Random Tot) yielded several variants of both discriminative phenotypes. The region defined by Part 2 contains the putative fatty acid-binding tunnel (22). Interestingly, library Random 2 yielded a higher rate of active variants than Random Rec or Random Tot. This appears to result from having a lower but more focused mutational load. As a result, fewer screening efforts (384 *vs* 768) yielded a greater number of discriminative variants in library Random 2 (21 total) than in libraries Random Rec or Random Tot (18 and 14, respectively). These results clearly illustrate the additional level of control gained by a segment-wise approach to mutagenesis.

These and previous efforts (11) provided us with a diverse pool of variants that were either point-mutated at positions Tyr93 (Part 1), Tyr183 (Part 1) , Phe431 (Part 3) or randomized entirely or in part, and exhibited discrimination in hydrolysis of short- or long-chain triglycerides. In addition to the results presented here, we also applied the flexible reassembly strategy to recombine libraries point mutated in Parts 1 and 3 (namely, Try93/Phe431, and Tyr183/Phe431), to uncover potential synergistic effects. No improved discriminative variants were obtained from these combinations and they are not further discussed. All selected variants were characterized by DNA sequencing to gain insight into the acquired discriminative abilities.

4.3.3. Characterization of point substitutions procuring discrimination for the hydrolysis of short- or long-chain triglycerides

When Tyr93 was substituted with Gly, Cal-A lost its ability to process tributyrin while maintaining the ability to process olive oil, making Tyr93Gly discriminative towards hydrolysis of long-chain fatty acids (Figure 4.2). Substitution of Tyr with small, polar residues (Ser, Cys, Asn and Asp) resulted in loss of activity for both substrates while larger, hydrophobic substitutions (His, Phe, Val, Leu, Ile) were comparable in activity to the wild-type.

Most of the Tyr183 substitutions we identified abolished activity (Gly, Asn, Arg, Ser, His, Ile, Val and Leu). Only Tyr183Phe was active, and it was active only toward tributyrin, suggesting that Tyr183Phe is discriminative for hydrolysis of short-chain triglycerides. In contrast, Phe431 was

highly permissive to substitution. Activity was maintained toward long and short-chain triglycerides in all variants obtained: Gly, Asn, Cys, Ser, Arg, His, Val, Leu (Fig 2).

These results inform us of the utility of varying Cal-A positions 93 and 183 (belonging to Part 1) to alter substrate discrimination, while position 431 (belonging to Part 3) was not productively altered toward this goal. In addition to acquiring information on the tolerance of these positions to substitution, it is gratifying that we obtained both discriminative phenotypes with point mutations in a single region (Part 1).

	Tyr93		Tyr183		Phe431	
	Trib	Olive	Trib	Olive	Trib	Olive
Discriminative	Gly ✗	✓	Phe ✓	✗		
Non-discriminative	His ✓	✓	Arg ✗	✗	Arg ✓	✓
	Ile ✓	✓	Asn ✗	✗	Asn ✓	✓
	Leu ✓	✓	Gly ✗	✗	Cys ✓	✓
	Phe ✓	✓	His ✗	✗	Gly ✓	✓
	Val ✓	✓	Ile ✗	✗	His ✓	✓
	Asn ✗	✗	Leu ✗	✗	Leu ✓	✓
	Asp ✗	✗	Ser ✗	✗	Ser ✓	✓
	Cys ✗	✗	Val ✗	✗	Val ✓	✓
	Ser ✗	✗				

Figure 4.2. – Short- *vs* long-chain hydrolytic activity in point substituted libraries of Tyr93, Tyr183 and Phe431.

Variants were screened for halo formation around colonies spotted on solid medium. Hydrolysis of the short-chain tributyrin (Trib) or long-chain olive oil (Olive) were rated as active (green checkmark) or inactive (red 'X').

4.3.4. Identification of the substitutions in randomized variants procuring discriminative hydrolysis

Analysis of discriminative variants selected from randomly mutated libraries required a different approach because they generally included multiple amino acid substitutions. As a result, their acquired discriminative ability can be due to one or more among these. We turned to a 3D mapping approach to provide a holistic determination of the key regions that confer either discriminative capacity to Cal-A.

A script was created to align each variant sequence with the Cal-A reference sequence and identify amino acid substitutions. Each variant was assigned a phenotype describing either activity towards, or capacity to discriminate between, short- and long-chain triglycerides. The variants, which contain between 1 and 6 mutations, were ranked from inactive to highly active according to halo formation upon triplicate in-plate screening, the wild-type Cal-A serving as a reference activity level (A2.1-A2.3 Tables). Each defined phenotype, linked to the corresponding mutations in each variant, was assigned a value that served for 3D visualization on the structure of Cal-A. This approach allowed us to rapidly analyze the randomized variants from the perspectives of activity towards a given substrate and discrimination between substrates, and gain insight into key regions and residues that altered function.

4.3.5. Activity levels of discriminative variants

Visualizing the positions where substitutions were consistent with discriminative hydrolysis of short-chain triglycerides (tributyrin) or long-chain triglycerides (olive oil) immediately reveal patterns of modified levels of activity (Figure 4.3). Library Random 2 includes residues linked to phenotypes that confer higher than wild-type activity towards the short-chain substrate but none showed higher than wild-type activity towards the long-chain substrate (Figure 4.3; Table A2.1 in annex 2). In contrast, libraries Random Rec and Random Tot link a greater diversity of residues to high-activity phenotypes, including variants with activity greater than the wild-type, for both short- and long-chain activity (Figure 4.3; Tables A3.2 and A2.3 in annex 2). The residues that are substituted in high-activity variants are predominantly located in Part 2 (41 residues), while Part 1 contains 12 and Part 3 contains 20 residues associated with high activity, excluding repetitions (S1 Fig).

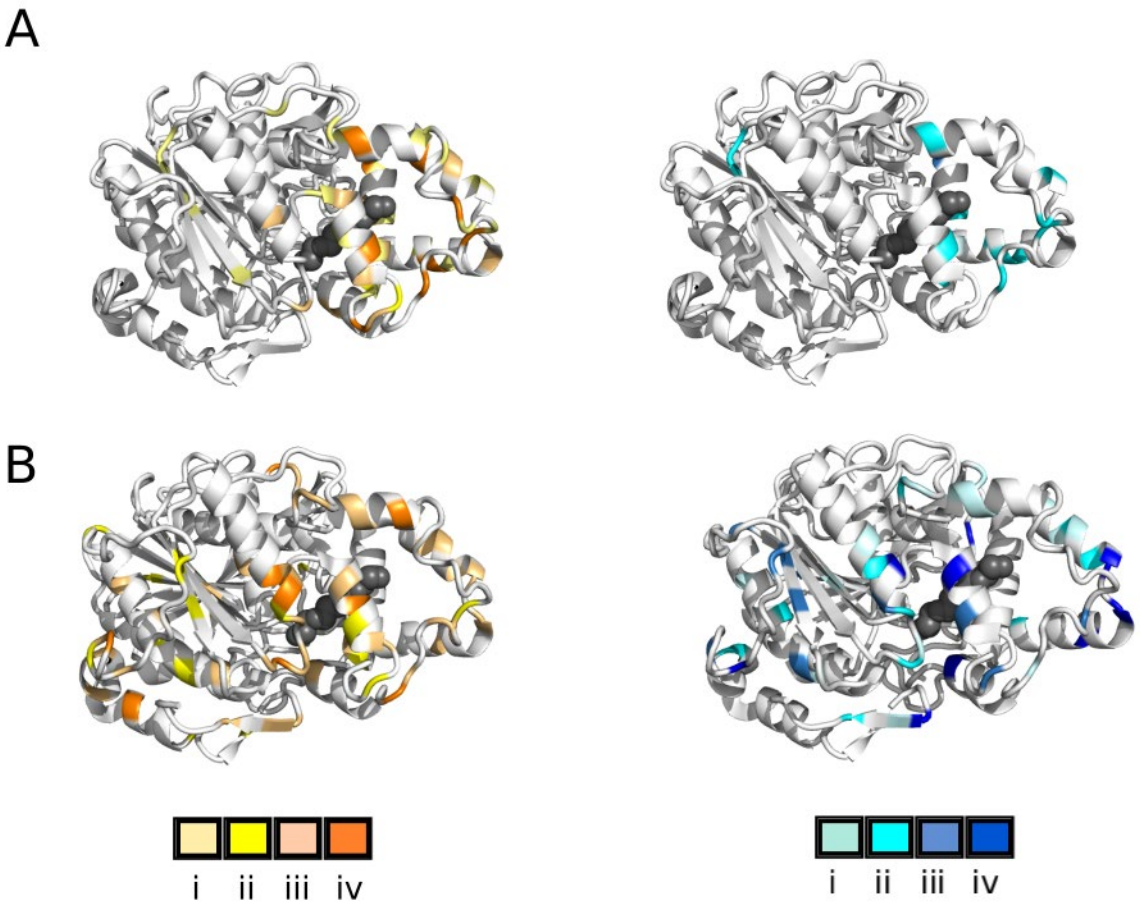


Figure 4.3. – Activity level of the discriminative variants towards short and long chain triglycerides

A: Variants selected from library Random 2. B: Variants selected from libraries Random Tot and Random Rec, combined. Left panels, short-chain activity (hydrolysis of tributyrin): gradient from light yellow (low activity) to orange (high activity). Right panels, long-chain activity (hydrolysis of olive oil): gradient from light (low activity) to dark blue (high activity). Inactive variants coloured as background (gray). Wild-type activity corresponds to shade iii. Where more than one discriminative variant was mutated at the same position, the position is colored according to the variant having the highest activity. A PEG-4 molecule is shown in black spheres, crystallized inside the putative tunnel (22, 29). Libraries Random Tot and Random Rec are individually represented in Figure A2.1 in annex 2. Libraries Random 1 and Random 3 did not yield discriminative variants and are not represented.

We noted that the majority of discriminative variants displaying activity greater than wild-type for short- or long-chain hydrolysis and including substitutions in Part 1 or 3 also included at least one substitution in Part 2, with the exception of Gln65Leu, Pro84Gln, Ser359Ile, Ser377Arg, and Ala444Thr. Although the role of individual substitutions in the randomized variants has not been

deconvoluted, this suggests that substitutions within Part 2 may be main contributors to the observed discriminative phenotypes. In one specific case, we were able to compare the point-substituted Arg262His variant alone and in conjunction with two additional substitutions. Interestingly, Arg262His (Part 2) alone conferred low activity towards long-chain triglycerides (Table A2.1 in annex 2), yet when selected in combination with substitutions Leu274Trp and Gly432Asp (Part 3), it conferred high activity towards the long-chain substrate (Table A2.3 in annex 2). The increased activity is apparently due to an additive or synergistic effect of the combined substitutions.

4.3.6. Analysis of discriminative quality of the variants

From the perspective of long-chain *vs* short-chain substrate discrimination, clear patterns were also revealed (Figure 4.4; S1-S4 Tables).

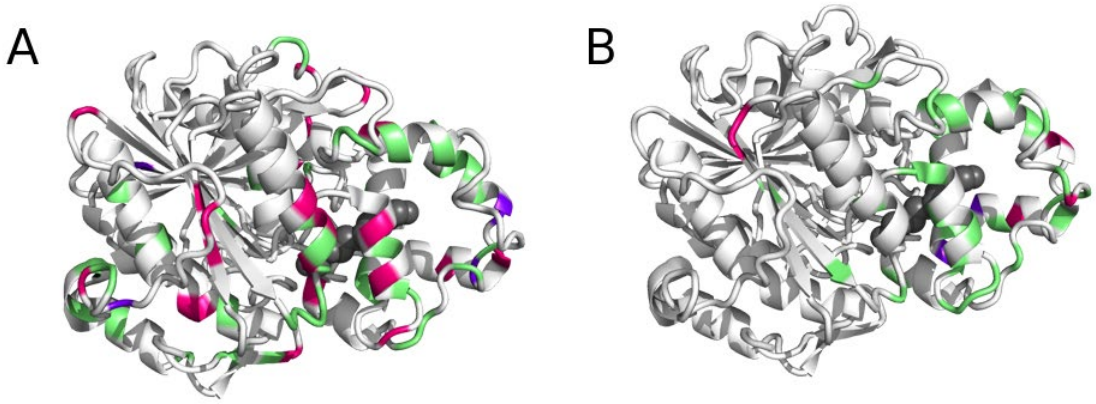


Figure 4.4. – Residues substituted in variants conferring discriminative activity.

A: Libraries Random Tot and Random Rec, combined. B: Library Random 2. Green: clear discrimination for hydrolysis of short-chain triglycerides (tributyrin). Magenta: clear discrimination for hydrolysis of the long-chain triglycerides (olive oil). Purple: residues substituted in distinct variants showing opposite discriminative phenotypes. A PEG-4 molecule is shown in black spheres, crystallized inside the putative tunnel (22, 29).

SHORT CHAIN DISCRIMINATION: Short-chain discrimination was strongly biased to substitutions in Part 2: it is compelling to note that more than half (34) of the total mutations found (55) in the fully randomized libraries (Random Tot and Random Rec) belong to Part 2, and eight among these 34 residues were also identified upon screening library Random 2 (Figure 4.3, Tables A2.1 and

A2.2 in annex 2). Furthermore, all but one of the additional 21 residues identified outside of Part 2 belong to variants including at least one substitution in Part 2. The sole exception, Ser377Arg, was a point-substituted variant offering weak discrimination for short-chain triglycerides (Figure A2.2 and Table A2.2 in annex 2). These results clearly identify this tunnel region as being key for fatty-acid substrate recognition, supporting the hypothesis based on co-crystallized PEG-4 and previous engineering results (22, 29).

A number of the selected, randomized variants held a single substitution, allowing for straightforward analysis of their impact on substrate recognition. Point-mutations that confer strong short-chain discriminative activity are Phe222Ile, Leu225Pro, Gly237Asp, Gly240Cys, and Ala244Val, while point-mutants Leu235Pro and Pro338Glu/Thr confer weak short-chain discrimination (Tables A2.1-A2.3 in annex 2). The helix-loop-helix motif in Part 2, formally delimited by residues Ser217-His245 according to the crystal structure and belonging to the Cal-A lid domain,(29) is thus a hot-spot for creation of short-chain discrimination.

Additional information was obtained from positions that where mutations were observed at a higher frequency, accompanied by different further mutations. This suggests that they are key to the observed property. Residues selected more than once within distinct variants are Phe222Ile (3×), Gly237 substituted either as Asp/Cys/Ser, Ala244 as Thr/Val, Pro338 substituted either as Thr/Gln and Ala342 substituted either as Val (2×)/Asp (Tables A2.2 and A2.3 in annex 2). In addition, the combination of Pro229Leu/Gly237Ser (selected 2×) conferred strong short-chain discrimination. Three further residues appeared twice each among the weak short-chain discriminative pool: Thr221Arg/Lys, Ala236Thr/Val and Asn291Lys/Tyr (Table A2.1 in annex 2). All the above residues belong to Part 2 (residues 211 – 350), confirming the importance of this region in conferring chain-length selectivity.

LONG CHAIN DISCRIMINATION: Variants giving rise to long-chain discrimination held mutations that were more scattered throughout the protein: 6 mutations were found in Part 1, 12 in Part 2 and 10 in Part 3 (Figure 4.4b). Pro84Gln (Part 1), Gly232Cys (Part 2) and Ser359Ile (Part 3) were point-mutated variants, and unequivocally confer long-chain discrimination. Together, substitutions Gln65Leu (Part 1) and Ala444Thr (Part 3) gave long-chain discrimination. All other long-chain

discriminative variants containing any substitution in Parts 1 and/or 3 also contain at least one substitution in Part 2. The positions substituted more than once for this phenotype were Arg262His (2×), and Ala402Val/Glu (2×). Overall, fewer substitutions were identified that give rise to long-chain discrimination relative to short-chain, as may be expected according to the ease of excluding large substrates relative to that of including only large substrates.

SUBSTITUTIONS FOUND BOTH IN SHORT- AND LONG-CHAIN DISCRIMINATIVE VARIANTS: Also of interest are positions where substitutions occurred both in variants that showed selectivity towards short and long chain triglycerides: Lys265Thr or Gln (short-chain or long-chain discrimination, respectively), Tyr136Asn or Phe (short/long-chain, respectively), Leu289Met or Val (short/long-chain, respectively), Ala402Val or Ala/Glu (short/long-chain, respectively) and Arg255His or Ser (short/long-chain, respectively) (Figure 4.4). Each of the above variants contained further mutations which may be determinants of selectivity. The impact of accompanying mutations is clearly illustrated by Gly232Cys: alone, it sufficed to confer long-chain discrimination, while its combination with Gly227Ser, and Lys265Thr inverted selectivity. Considering that Lys265 was substituted as Gln in a variant with long-chain discrimination (along with substitutions Val238Leu, Arg262His, Thr293Ser andPro325Thr), it is plausible that the inverted selectivity is attributable to Gly227Ser. Finally, Gly228Ser/Val were included in variants conferring short-chain discrimination (each holding a second substitution in Part 2), yet the same Gly228Ser was included in a variant conferring long-chain discrimination, accompanied by 2 substitutions in Part 3.

Globally, screening the diversified libraries against triglyceride substrates led to the unequivocal identification of previously unreported residues that contribute to substrate recognition and discrimination. Point-substituted variants that confer short-chain discrimination were identified at positions 183, 235, 240, 244, 338 and 377, whereas long-chain discriminators were identified at positions 84, 93, 232 and 359. In addition, our results confirm and lend further support to prior reports of key residues. In particular, Thr221 and Phe222 were found to be crucial to modulate selectivity in the hydrolysis of *trans*- and *cis*-fatty acids (22), while Gly237 had previously been identified as important for short-chain discrimination (23). Residues Thr221/Leu225/Phe233/Gly237 were involved in increased activity towards sterically hindered substrates (37).

4.3.7. Quantification of hydrolysis using *p*-NO_2 phenyl esters of different chain lengths

p-NO_2-Phenyl-fatty acids are the reference substrates used in industry assessment of lipase reactivity (34). *p*-NO_2-Phenyl-ester derivatives of single fatty acids differ importantly from the triglycerides used in the in-plate screening: *p*-NO_2-phenolate is an activated leaving group, facilitating hydrolytic reactivity; binding to the enzyme necessarily differs since the volume of a triglyceride is nearly 3× that of a *p*-NO_2-phenyl-fatty acid. We had envisaged applying the rapid triglyceride screening method described above to qualitatively identify promising variants for further quantitative *p*-NO_2-phenyl-fatty acid characterization. However, consistent with the reasons mentioned above, we obtained incomplete correspondence between the two assays, as recently observed in a study of *Candida rugosa lipase 1* (38). Nonetheless, screening with *p*-NO_2-phenyl-substrates allowed us to gather quantitative insights into chain-length discrimination for all variants that showed correspondence. We report results of assays on clarified *E. coli* lysates using *p*-NO_2-phenyl-butyrate, -decanoate, -dodecanoate and –palmitate; assays with *p*-NO_2-phenyl-octanoate produced lower quality results due to background hydrolysis. Results show that wild-type Cal-A displays a slight selectivity towards short-chain fatty acids (activity 60% higher with *p*-NO_2-phenyl-butyrate than *p*-NO_2-phenyl-palmitate, S3 Fig).

For variants belonging to the Tyr93 library, activity in clarified lysates is lost when Tyr93 is substituted with Val, Asn, Leu, Cys, Ile and Asp (Figure 4.5). Surprisingly, selectivity is reversed when Tyr is substituted with the conservative substitutions His and Phe, as well as Gly. Lysate holding the Tyr93Gly variant is 12 times more active toward *p*-NO_2-phenyl palmitate than *p*-NO_2-phenyl butyrate (compared to 2.5 times for Tyr93His), yet its maximum activity with *p*-NO_2-phenyl palmitate is importantly reduced (40% of the wild-type; Figure 4.5). This illustrates a trade-off between improved discrimination and residual activity, and confirms that activity of variant Tyr93Gly is markedly higher with long-chain substrate – whether a single-chain *p*-NO_2-phenyl derivative or a triglyceride – than with short-chain substrates.

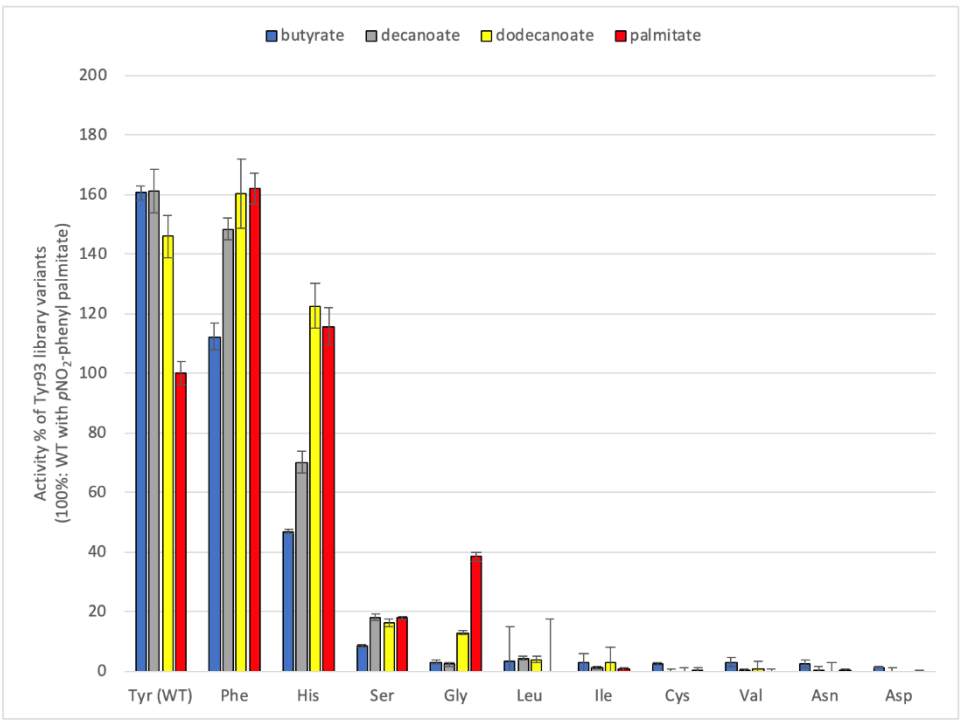


Figure 4.5. – Hydrolytic activity of Tyr93 library variants with *p*-NO_2-phenyl fatty acids

Assays were performed in triplicate with clarified E. coli lysates. Activity is reported relative to WT Cal-A with p-NO_2-phenyl-palmitate. Its specific activity (S.A) = 0.4 U/mg, is set as 100%.

Among the 53 variants from randomized libraries identified as discriminative when screened against triglycerides, 7 showed discrimination in the *p*-NO_2-phenyl-fatty acid assays. We quantified their activity towards these industrially preferred benchmarking compounds (Figure 4.6). Consistent with their lack of activity on the olive oil solid medium, the variants all show selective hydrolysis of short-chain *p*-NO_2-phenyl-fatty acid substrates. Their activity in the lysate is in the same range as, or greater than, the wild-type towards *p*-NO_2-phenyl-butyrate and they are poorly active or inactive towards all longer chain substrates. We thus demonstrate that these variants exhibiting high activity only toward short-chain triglycerides (solid medium assay) show the same selectivity toward short-chain *p*-NO_2-phenyl-fatty acids.

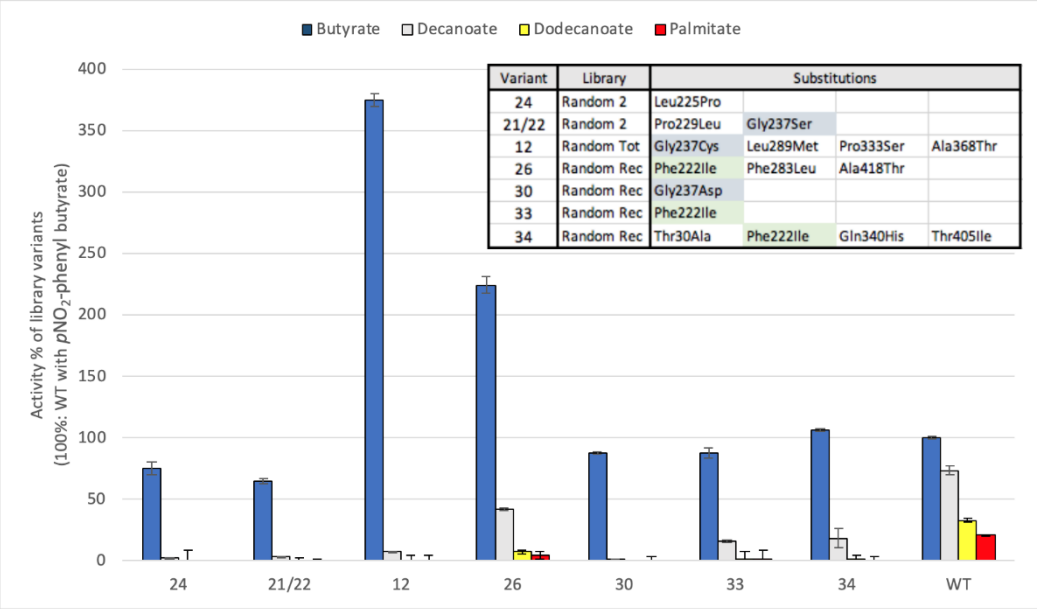


Figure 4.6. – Hydrolytic activity of discriminative variants from randomized libraries with *p*-NO_2-phenyl fatty acids.

Assays were performed in triplicate with clarified E. coli lysates. Activity is reported relative to WT Cal-A with p-NO_2-phenyl-butyrate. Its specific activity (S.A.) = 0.4 U/mg, is set as 100%. Substitutions included in each variant are inset and recurring mutations are highlighted. Variant number is shown as per S1-S3 Tables.

Among these variants, substitutions Gly237Asp and Phe222Ile of the 217-245 helix-loop-helix motif appear as point mutations, confirming their role in improved discrimination (Figure A2.2 in annex 2). Interestingly, position 237 was previously reported to be important for discrimination of short and long-chain fatty acids according to analysis of 4 different substitutions (Ala, Leu, Tyr, Val) (23). Here, three distinct discriminative variants carry Gly237 substitutions, as identified in the triglyceride screening (Asp, Cys, Ser) that are an addition to those that were previously characterized (23). To verify their effect in isolation, we broadened the deconvolution at this position (Figure 4.7). When Gly237 is substituted with any amino acid, strong short-chain discrimination is established; Cal-A loses its ability to process longer substrates. It thus appears that any steric hindrance at position 237 precludes binding of fatty acid esters longer than butyrate, confirming and extending prior observations of the importance of this residue in chain-length discrimination (23).

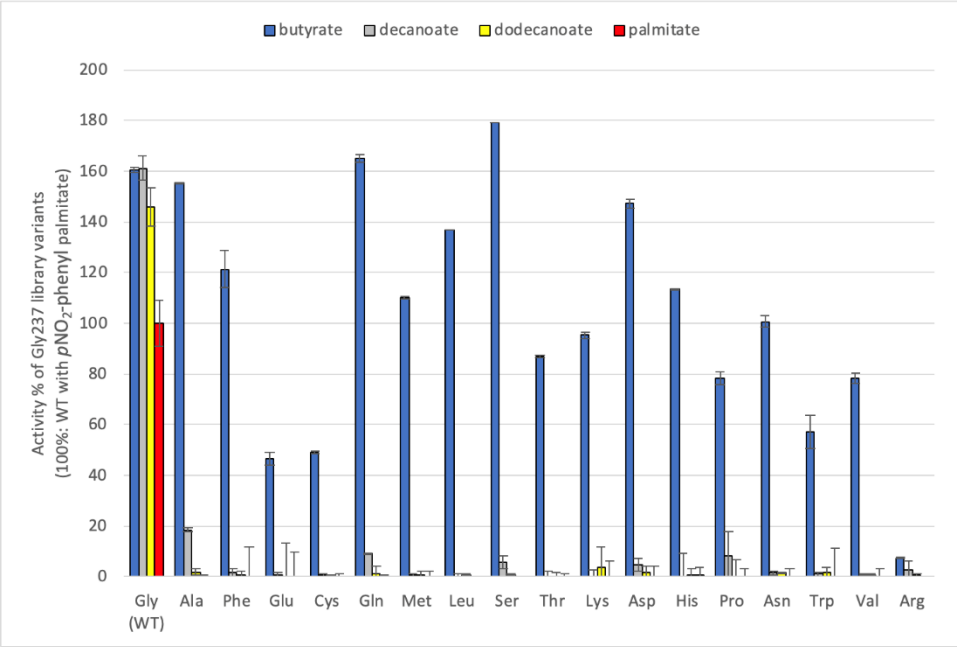


Figure 4.7. – Hydrolytic activity of Gly237 variants with *p*-NO_2-phenyl fatty acids.

Assays were performed in triplicate with clarified E. coli lysates. Activity is reported relative to WT Cal-A with p-NO_2-phenyl-palmitate. Its specific activity (S.A) = 0.4 U/mg, is set as 100%.

4.4. Conclusions

Through the application of our Golden Gate-based strategy for enzyme engineering, we generated focused randomized libraries covering specific parts of Cal-A lipase to scan for variants improved in the discriminative hydrolysis of short- *vs* long-chain triglycerides. This approach allowed introducing the entire mutational load in selected regions, without putting a strain on distal parts of the protein, which is not readily achievable with traditional randomization methods. Furthermore, automated visualization of activity screening results on the structure allowed straightforward identification of a key functional hot-spot.

By those means, we have identified the 217-245 helix-loop-helix region in Part 2 of Cal-A as being fundamental for triglyceride recognition. While it was not possible to dock a triglyceride in Cal-A without assuming extensive conformational changes to fit it in the tunnel (11), our findings confirm that Part 2 is a key to triglyceride recognition. We also identified several previously

unreported point substitutions that improve discrimination both for short and long-chain triglycerides. Although there was an incomplete correlation between hydrolysis of triglycerides and *p*-NO_2-phenyl fatty acids, limited analysis using the latter has provided some quantitative insights on improved selectivity.

4.5. Materials and methods

4.5.1. Materials, strains, vectors and culture conditions

Unless otherwise stated, DNA primers and chemicals were purchased from Sigma-Aldrich. The rectangular Nunc OmniTray dishes were purchased from Thermo Fisher Scientific. Restriction enzymes were obtained from New England Biolabs, while the TAKARA ligase was from Clontech. Protein markers were purchased from Thermofisher and New England Biolabs. Phusion Green High-Fidelity DNA Polymerase and the PureLink PCR Purification Kit (Invitrogen) were also from Thermo Fisher Scientific. Taq polymerase was purchased from Biobasics. The QuikChange Lightning Site-Directed Mutagenesis Kit and the GeneMorph II Random Mutagenesis Kit from Agilent Technologies were used for mutagenesis. The Monarch Plasmid Miniprep Kit and the Monarch DNA Gel Extraction Kit were purchased from New England Biolabs. A Beckman Coulter Biomek NX^p Robot was used to perform the automated triglyceride screening.

Cal-A parts were provided in mother vectors (pM269; chloramphenicol resistant) and codon optimised by ATUM (https://www.atum.bio, formerly DNA2.0 California, USA) as were the linearized daughter vectors (pD441pelB, pD441OmpA; kanamycin resistant). 'Mother vector' refers to a plasmid carrying a part, and 'daughter vector' is the expression plasmid. Sequencing of the DNA was performed by either the Genomic Platform of IRIC (Institute for Research in Immunology and Cancerology), Université de Montréal, or the Centre d'Innovation Génome Québec at McGill University (QC, Canada).

The $CaCl_2$ method was used for the preparation of chemically competent *E. coli* BL21 (DE3). Unless otherwise specified, transformed *E. coli* strains were cultured on LB agar and in LB broth containing kanamycin (50 µg/mL) and chloramphenicol (35 µg/mL) or ampicillin (100 µg/mL),

depending on the resistance marker, overnight (for 16 hours) at 37°C, with shaking at 250 rpm when appropriate.

4.2.1. Assembly of plasmids and library generation

Restriction digestion of daughter plasmids and wild-type or mutated parts was performed using SapI or BsaI restriction enzymes. Mutant libraries were obtained as described in previous work (11). Design and assembly *in silico* of all genetic constructs and analysis of the sequencing results were carried out using the SnapGene software (GSL Biotech; available at snapgene.com).

4.2.2. Screening of Cal-A variants through automation by liquid handling robot

Screening with triglycerides was carried out on tributyrin (short-chain triglyceride) and olive oil (long-chain triglyceride) plates as described previously (11). A halo of clearance around a colony grown on tributyrin-containing medium indicates activity toward the short-chain substrate while a halo of fluorescence around a colony on olive oil/rhodamine-containing medium indicates activity toward long-chain substrate. Wild-type Cal-A activity was assigned a value of 3 and haloes of similar size and intensity (either degree of clearance for tributyrin, or fluorescence intensity for olive oil/rhodamine) upon visual inspection of triplicates were assigned that value. Larger and more intense haloes were clearly distinguished and were assigned the maximal value of 4. Less active variants generated smaller and less intense haloes that assigned the value of 2 or 1 (smallest/faintest haloes).

4.2.3. Computational method for phenotypic characterization

To facilitate the establishment and analysis of genotype/phenotype relationships in our dataset, a Python script was developed. DNA sequencing results were classified by phenotype according to the results of screening with triglycerides. One phenotype qualified the level of activity observed toward a given triglyceride substrate (halo formation). The second phenotype identified short- or long-chain discrimination based on the relative activity with each of the triglyceride substrates; mutated residues that contributed to both phenotypes were assigned to a third discriminative category.

To simplify 3D analysis, each phenotype was assigned a numerical value that was used to modify the B-factor column in 2veo pdb file using the data2bfactor.py script (http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/data2bfactor.py) in the molecular graphics program PyMOL. Coloring by B-factor number revealed the localization of the residues contributing to each phenotype. The script is available upon request.

4.2.4. Screening Cal-A variants with p-NO_2-phenyl esters

For the liquid screening using *p*-NO_2-phenyl esters, the variants and the negative control (*E. coli* BL21(DE3) transformed with pD441pelB that does not contain the Cal-A gene) were expressed in 10 mL ZYP-5052 auto-inducing medium, described by Studier (39), supplied with kanamycin, at 30°C over 24 hours. An overnight culture in LB grown at 37 °C was used as inoculum (100 µL inoculum for 10 mL of culture). Conditions were determined by a Design of Experiment performed considering different temperatures, volumes and inoculum conditions (11). The cultures were pelleted by microcentrifuging at 3000 rpm over 10 min and lysed by resuspending in 1 mL of 50 mM Tris, pH 8.0, sonicating until clear and microcentrifuging at 14000 rpm for 10 min. The resulting supernatant was assayed. Spectrophotometric activity measurements were based on the substrate-dependent absorbance change of *p*-NO_2-phenyl esters of different chain lengths at 405 nm. Assays were routinely done in 200 µL at 45°C, with measurement over 30 min using a *BMG FLUOstar* 96-well plate spectrophotometer.

Unless otherwise stated, the reaction mixture contained *p*-NO_2-phenyl-fatty acid (0.5 mM), 1% Triton x 100, 50 mM Tris, pH 8.0. A Bradford assay was performed on the lysates. One unit of Cal-A corresponded to the amount of enzyme required to produce 1 µmol of *p*-NO_2-phenolate per min at 45°C.

The specific activity was calculated according to the following formula:

$$S.A.\left(\frac{U}{mg}\right) = \frac{\Delta 450}{l \times \varepsilon pNO2phenolate} \times \frac{Vt}{Ve} \times \frac{1}{[Ce]}$$

where: Δ450 = absorbance slope (min^{-1}); l = cuvette length (cm); ε *p*-NO_2-phenolate = 18.1 mM^{-1} cm^{-1}; Vt = total reaction volume in mL; Ve = total enzyme volume in mL; [Ce] = concentration of total protein in lysate in mg/mL.

4.6. Acknowledgments

The authors acknowledge Paul Mugford (DSM) and Pierre Lepage (McGill Innovation Center) for fruitful discussions.

4.7. References

1. Denard CA, Ren H, Zhao H, 2015. Improving and repurposing biocatalysts via directed evolution. Curr Opin Chem Biol. 2015;25:55-64.

2. Porter JL, Rusli RA, Ollis DL. Directed evolution of enzymes for industrial biocatalysis. Chembiochem : a European journal of chemical biology. 2015;17(3):197-203.

3. Turner NJ. Directed evolution drives the next generation of biocatalysts. Nature chemical biology. 2009;5(8):568-74.

4. Li Y, Cirino PC. Recent advances in engineering proteins for biocatalysis. Biotechnol Bioeng. 2014;111(7):1273-87.

5. Packer MS, Liu DR. Methods for the directed evolution of proteins. Nat Rev Genet. 2015;16(7):379-94.

6. Wong TS, Roccatano D, Schwaneberg U. Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. Environ Microbiol. 2007;9(11):2645-59.

7. Ruff AJ, Dennig A, Schwaneberg U. To get what we aim for - progress in diversity generation methods. FEBS J. 2013;280(13):2961-78.

8. Gillam EMJ, Copp JN, Ackerley D, editors. Directed evolution library creation. New York, NY: Springer; 2014.

9. Davids T, Schmidt M, Böttcher D, Bornscheuer UT. Strategies for the discovery and engineering of enzymes for biocatalysis. Curr Opin Chem Biol. 2013;17(2):215-20.

10. Cozens C, Pinheiro VB. Darwin Assembly: fast, efficient, multi-site bespoke mutagenesis. Nucleic Acids Res. 2018;46(8):e51.

11. Quaglia D, Ebert MCCJC, Mugford PF, Pelletier JN. Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening. PLoS ONE. 2017;12(2):e0171741.

12. Arnold FH. Combinatorial and computational challenges for biocatalyst design. Nature. 2001;409:1-5.

13. Chica RA, Doucet N, Pelletier JN. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol. 2005;16(4):378-84.

14. Damborsky J, Brezovsky J. Computational tools for designing and engineering biocatalysts. Curr Opin Chem Biol. 2009;13(1):26-34.

15. Scrutton NS. Speeding up enzyme engineering computationally. IUCrJ. 2017;4(1):5-6.

16. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PloS one. 2008;3(11):e3647.

17. Engler C, Gruetzner R, Kandzia R, Marillonnet S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PloS one. 2009;4(5):e5553.

18. Sarrion-Perdigones A, Falconi EE, Zandalinas SI, Juárez P, Fernández-del-Carmen A, Granell A, et al. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS ONE. 2011;6(7):e21622.

19. Reetz MT. The importance of additive and non-additive mutational effects in protein engineering. Angew Chem Int Ed. 2013;52(10):2658-66.

20. He Y, Li J, Kodali S, Chen B, Guo Z. Rationale behind the near-ideal catalysis of Candida antarctica lipase A (CAL-A) for highly concentrating ω-3 polyunsaturated fatty acids into monoacylglycerols. Food Chem. 2017;219:230-9.

21. Akanbi TO, Barrow CJ. Candida antarctica lipase A effectively concentrates DHA from fish and thraustochytrid oils. Food Chem. 2017;229:509-16.

22. Brundiek HB, Evitt AS, Kourist R. Creation of a lipase highly selective for trans fatty acids by protein engineering. Angew Chem Int Ed. 2012;51:412-4.

23. Brundiek H, Padhi SK, Kourist R, Evitt A. Altering the scissile fatty acid binding site of Candida antarctica lipase A by protein engineering for the selective hydrolysis of medium chain fatty acids. Eur J Lipid Sci Technol. 2012;114:1148-53.

24. Nyyssölä A, Miettinen H, Kontkanen H, Lille M, Partanen R, Rokka S, et al. Treatment of milk fat with sn-2 specific Pseudozyma antarctica lipase A for targeted hydrolysis of saturated medium and long-chain fatty acids. Int Dairy J. 2015;41:16-22.

25. Miettinen H, Nyyssölä A, Rokka S, Kontkanen H, Kruus K. Screening of microbes for lipases specific for saturated medium and long-chain fatty acids of milk fat. Int Dairy J. 2013;32(2):61-7.

26. Borrelli GM, Trono D. Recombinant lipases and phospholipases and their use as biocatalysts for industrial applications. International Journal of Molecular Sciences. 2015;16(9):20774-840.

27. Ericsson DJ, Kasrayan A, Johansson P, Bergfors T, Sandström AG, Bäckvall J, et al. X-ray structure of Candida antarctica lipase A shows a novel lid structure and a likely mode of interfacial activation. J MolBiol. 2008;376:109-19.

28. De María PD, Carboni-Oerlemans C, Tuin B. Biotechnological applications of Candida antarctica lipase A: State-of-the-art. J Mol Catal B: Enzym. 2005;37:36-46.

29. Ericsson DJ, Kasrayan A, Johansson P. X-ray structure of Candida antarctica lipase A shows a novel lid structure and a likely mode of interfacial activation. J MolBiol. 2008;375:109-19.

30. Zorn K, Oroz-Guinea I, Brundiek H, Dörr M, Bornscheuer UT. Alteration of chain length selectivity of Candida antarctica Lipase A by semi-rational design for the enrichment of erucic and gondoic fatty acids. Adv Synth Catal. 2018;360(21):4115-31.

31. Boskou D. Olive oil: Chemistry and technology. Second ed. Champaign, Illinois: AOCS Press; 2006.

32. Divakar K, Gautam P. A multisubstrate assay for lipases/esterases: Assessing acyl chain length selectivity by reverse-phase high-performance liquid chromatography. Anal Biochem. 2014;448(1):38-40.

33. Joerger RD, Haas MJ. Alteration of chain length selectivity of a Rhizopus delemar lipase through site-directed mutagenesis. Lipids. 1994;29(6):377-84.

34. Stoytcheva M, Montero G, Zlatev R, Leon JA, Gochev V. Analytical methods for lipases activity determination: A Review. Curr Anal Chem. 2012;8(3):400-7.

35. Kouker G, Jaeger KE. Specific and sensitive plate assay for bacterial lipases. Appl Environ Microbiol. 1987;53(1):211-3.

36. Lawrence RC, Fryer TF, Reiter B. Rapid method for the quantitative estimation of microbial lipases. Nat Rev Drug Discov. 1967;213(5082):1264-5.

37. Sandstrom AG, Wikmark Y, Engström K, Nyhlén J, Bäckvall J-E. Combinatorial reshaping of the Candida antarctica lipase A substrate pocket for enantioselectivity using an extremely condensed library. Proc Nat Acad Sci USA. 2012;109(1):78-83.

38. Tanaka S-i, Takahashi T, Koide A, Iwamoto R, Koikeda S, Koide S. Monobody-mediated alteration of lipase substrate specificity. ACS Chem Biol. 2018;13(6):1487-92.
39. Studier F. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif. 2005;41:207-34.

Chapter 5 – Epistatic interactions in engineered Cal-A lipases modulate chain-length discrimination via tunnel reshaping

Preface to Chapter 5

In order for directed evolution to improve our understanding of a specific enzyme system, characterization of the effects of mutations is needed. In this study, chain-length selective variants identified in Chapter 4 have been further characterized to understand the basis for their observed phenotypes, revealing an epistatic interaction at the bottom of the acyl-binding tunnel. As in Chapter 4, the variants were assayed for activity *in vivo* on agar plates containing emulsified triglycerides. To broaden our analysis, medium chain-length triglycerides were included here. In addition, the variants were assayed *in vitro* for activity towards a set of *p*-NO_2-phenyl esters and triglycerides of different chain length. Rationalization of the observed phenotypes was done *in silico* by identifying the main tunnels and performing flexible substrate docking.

This is an invited research article for the Protein Chemistry and Enzymology section in the journal Frontiers in Molecular Biosciences, to be submitted in february 2021. This study was initially designed by myself with the help of postdoctoral fellow Dr. Daniela Quaglia under the supervision of Prof. Joelle Pelletier. Qualitative plate assays were done by Dr. Quaglia and myself. Quantitative *p*-NO_2-phenyl ester assays were done by myself. Quantitative triglyceride assays and flexible docking are a result of a short internship in Prof. Manuel Ferrer's research group in the Institute of Catalysis, CSIC, Spain. Quantitative triglyceride and flexible docking were done by myself with the guidance of Dr. Cristina Coscolín, postdoctoral Fellow in the Ferrer group. I wrote the original draft which has been revised by Dr. Quaglia, Prof. Pelletier and Prof. Ferrer prior to submitting this thesis.

Manuscript in preparation - Article 4. Epistatic interactions in engineered Cal-A lipases modulate chain-length discrimination via tunnel reshaping

Lorea Alejaldre[1,2], Daniela Quaglia[3,4], Cristina Coscolín[5], Manuel Ferrer[5], Joelle N. Pelletier[1,2,3]

[1] Biochemistry Department, Université de Montréal, Canada

[2] PROTEO, The Québec Network for Research on Protein Function, Engineering and Applications, Canada

[3]Chemistry Department and Center in Green Chemistry and Catalysis (CGCC), Université de Montréal, Canada

[4] School of Chemistry, University of Nottingham, United Kingdom

[5]Institute of Catalysis, Consejo Superior de Investigaciones Científicas, Spain

***Correspondence:**

Joelle Pelletier

joelle.pelletier@umontreal.ca

5.1 Abstract

Enzyme engineering is a two-in-one tool: while it can be used to modify the enzyme towards a new function, it also paves the way to the discovery of key functional features. However, the efficient exploration of sequence space during the process of enzyme engineering is challenged by epistatic interactions between mutations that are difficult to predict from structure alone. The more engineering an enzyme has gone through, the more we know its key epistatic interactions and mechanisms by which successful mutations were obtained. This mechanistic understanding can help to improve future engineering endeavors by identifying key regions or mechanisms operating. Cal-A lipase enjoys properties that make it attractive for biotechnological applications yet it has been subjected to few directed evolution studies. As a result of its uncommon structure, information about triglyceride binding, conformational dynamics and epistasis is scarce and thus these variables are difficult to incorporate in enzyme engineering workflows. Following on the report of a key region conferring chain-length selectivity in Cal-A variants, we report *in vivo* and *in vitro* characterization of their deconvolution. Our study reveals that epistatic interactions at the end of the putative acyl-binding tunnel of a short-chain selective variant are responsible for this phenotype. Computational tunnel analysis and flexible docking reveal that these mutations re-shape the acyl-binding tunnel by increasing its curvature and thus decreasing its effective length. This epistatic mechanism for short-chain selectivity is less straightforward than previously reported mutations in which this phenotype was achieved in Cal-A by sterically blocking near the entrance of the acyl-binding tunnel. The difficulty in predicting such epistatic effects *a priori* underlines the importance of further exploring non-additive effects in Cal-A lipase.

5.2 Introduction

Enzymes are used daily, whether for the synthesis of valuable compounds in the pharmaceutical industry or for low-temperature stain removal in our own homes (1). Their biotechnological interest is such that it is estimated that the global enzyme market will be worth 10 billion USD by 2024(2). Their use and expected growth for applications in diverse industries is due to our ability to identify and tailor enzymatic activities to societal needs. In this respect, enzyme engineering has become a powerful tool modify enzymes and to inform about the properties of the

engineered enzyme (3-5). Rationalizing which regions or residues hold the greatest impact in modifying enzymatic function has helped to understand structure-function relationships that are not apparent when observing a structure (6-8). Integral enzyme properties such as protein dynamics, stability, allosterism or epistasis remain undetermined for many enzymes and are often revealed through *in vitro* evolution (7, 9-15). This information increasingly serves to guide enzyme engineering (16-21), justifying the need to better characterize new properties of enzymes throughout the process of enzyme engineering.

Here we characterize the effects of individual and combined mutations previously identified as modulating chain-length recognition in the enzyme *Candida antarctica* lipase A (Cal-A) (22). Several features of Cal-A lipase remain underexplored despite its potential industrial applications in the food, energy or pharmaceutical industries (23, 24). Noteworthy are its high thermal and solvent resistance, substrate and activity promiscuity, preference for hydrolysis of triglycerides at the sn2 position and trans fatty acid selectivity (23-25). In addition, it shows short-chain fatty acid selectivity (26), despite a putative acyl-binding tunnel that could bind up to 25 methylene units (27). Its crystal structure was resolved in 2008(27) and only a handful of directed evolution experiments have been performed on Cal-A lipase (23, 24). Moreover, its low similarity to other known lipases prevents making inferences about its conformational dynamics, activation or substrate binding, justifying deeper investigation into effects of mutations (28).

Chain-length selectivity is highly sought in the food industry to enhance the composition of fatty acids of precise lengths that offer health benefits or certain flavors (29-31). Selectivity for fatty acids shorter than C6 has been achieved in several lipases, including Cal-A, by blocking the acyl-binding tunnel sterically (32-34). Achieving selectivity towards fatty acids of lengths greater than C6 has proven to be more challenging since activity for shorter chain-lengths is preserved (22, 35). However, activity towards long-chain fatty acids can be increased by widening the entrance to the acyl-binding tunnel (35, 36). Other mechanisms to achieve either chain-length selectivity described in other lipases include conformational changes and lid modifications that often include epistatic effects (37-41).

In Cal-A lipase A, a prior study created region-focused libraries to identify variants with enhanced chain-length selectivity (22, 42). This semi-rational approach to directed evolution confirmed the existence of a key region for substrate recognition in the putative acyl-binding tunnel of Cal-A (43). Other studies have described short-chain selective Cal-A variants, mainly by sterically blocking the putative acyl-binding tunnel (34). To our knowledge, there is a single report of an epistatic interaction in Cal-A between two mutations at the bottom of the putative acyl-binding tunnel that confer short-chain selectivity (11).

To gain further insight into the mechanisms whereby substitutions modulate chain-length selectivity in Cal-A and to identify potential epistatic effects, we deconvoluted two chain-length selective variants isolated from a prior study that were defined by more than one mutation (22). Activity was determined *in vivo* and *in vitro* using natural triglyceride substrates and synthetic, activated *p*-NO_2-phenyl esters of varying chain lengths to determine the contribution of each mutation to the phenotype. Epistatic effects between mutations were confirmed, confirming the existence of other epistatic interactions in the putative acyl-binding tunnel. Tunnel analysis and flexible docking simulations are consistent with chain-length selectivity in Cal-A being a consequence of conformational changes in the acyl-binding tunnel or their accessibility. These events could not have been predicted from the 3D structure alone; the ready availability of a medium-throughput screening assay allowed for their experimental determination.

5.3Results

5.3.1 Deconvolution of variants for investigation of chain-length selectivity

Cal-A lipase, despite its long putative acyl-binding tunnel (27), shows higher activity towards short-chain fatty acids, more precisely chain lengths of C3-C4 (25, 26). Previously, to identify chain-length selective Cal-A variants, hydrolysis of the triglyceride glyceryl tributyrate (C4:0) and the long-chain triglyceride-enriched olive oil (C18:1) were assessed (22, 42). For these triglycerides, Cal-A wild-type shows ~70% lower activity towards olive oil (C18:1) than glyceryl tributyrate (C4:0) (26).

In order to decipher the mechanisms at play in Cal-A variants selective for substrates of different chain lengths, a short-chain and a long-chain discriminative variant previously obtained through focused mutagenesis or whole-gene random mutagenesis were selected (22). Both include mutations that have not been the focus of prior studies (34) (11). These variants were deconvoluted in this study to assess the impact of their individual mutations on ester hydrolysis.

Short-chain selective Cal-A variant R2-27 includes mutations M248L and N294I whereas long-chain selective AR15 includes mutations R262H, L274W and G432D (Figure 5.1). Short-chain selectivity has previously been achieved by individual or combined mutations in residues T221, G237, V238 and V290 that prevent binding of long fatty acids in the acyl-binding tunnel by blocking the entrance to the putative acyl-binding tunnel (11, 34). However, in variant R2-27, the mechanism providing short-chain selectivity should differ as mutations are found at the bottom of the acyl-binding tunnel (Figure 5.1).

The mechanism for long-chain selectivity is more elusive. In a previous study, all 13 long-chain discriminative variants of Cal-A identified retained at least moderate capacity to hydrolyze the short-chain triglyceride glyceryl tributyrate (C4:0) (22). In Cal-A, mutation Y93G at the entrance of the acyl-binding tunnel has been shown to confer long-chain selectivity by increasing activity towards long-chain triglycerides (22). The variant AR15 has mutation L274W close to the active site, as well as the flap G432D that could be modifying the active site cavity or substrate recognition (Figure 5.1). In other lipases, lid modifications have been shown to confer long-chain selectivity (38, 40, 41). However, this mechanism has not been described in Cal-A lipase. It should be noted that AR15, although it is one of the best long-chain selective variants isolated in the previous study (22), shows weak selectivity. However, understanding the implication of each mutation in this variant could help to enhance this property in future engineering efforts.

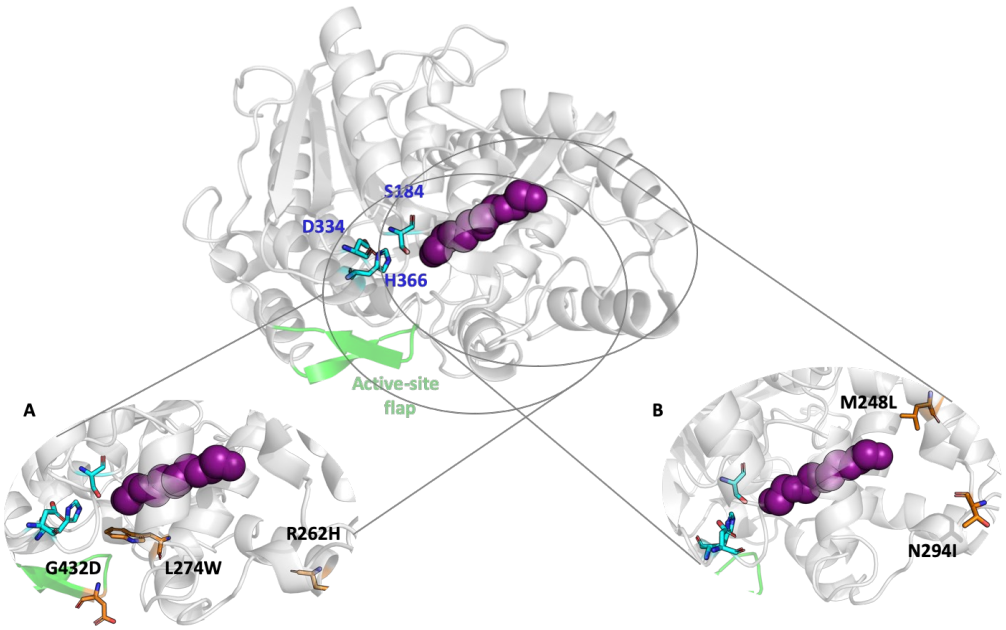


Figure 5.1. – Structure of Cal-A lipase (PDB ID: 2veo) with the catalytic triad in cyan sticks and the active site flap in green.

The putative acyl-binding tunnel is represented by a co-crystallized PEG molecule (purple). A: Zoom highlighting modelled mutations into the wild-type structure (PDB ID:2VEO) R262H, L274W and G432D in long-chain selective variant AR15. B: Zoom highlighting modelled mutations into the wild-type structure (PDB ID:2VEO) M248L and N294I at the bottom of the putative acyl-binding tunnel in short-chain selective variant R2-27.

5.3.2 Characterization of deconvoluted variants

Short-chain selective Cal-A variant R2-27 was deconvoluted into individually substituted variants M248L and N294I, and long-chain selective AR15 was deconvoluted into individually substituted variants R262H, L274W and G432D. All variants were expressed in *E. coli* and were assayed either as colonies on substrate-containing solid medium or as cell lysate in liquid reaction mixes. Natural triglyceride substrates were assayed in solid and liquid medium; they best represent the potential for the variants to hydrolyze complex substrates with more than one fatty acid chain. *p*-NO_2-Phenyl esters were assayed in liquid form. They are used as industry standards for the ease of the colorimetric assay. *p*-NO_2-Phenyl esters may represent the potential for variants to hydrolyze single-chain esters yet caution should be used in interpreting hydrolysis of those activated, aromatic esters (22, 34).

A) Qualitative assay for hydrolysis of triglycerides on agar plates

A qualitative *in vivo* assay was first performed to assess chain-length selectivity in a high-throughput manner. The assay consisted in plating *E. coli* expressing the Cal-A variants on ZY autoinduction agar plates containing emulsions of glyceryl tributyrate (C4:0), glyceryl trioctanoate (C8:0), coconut oil (C12:0) or olive oil (C18:1) and assessing halo size/fluorescence upon triglyceride hydrolysis (Figure 5.2). Comparison of halo sizes over the course of colony growth provides a rough estimate of activity (22, 42) and, at a glance, of the impact of individual mutations relative to the combined mutations.

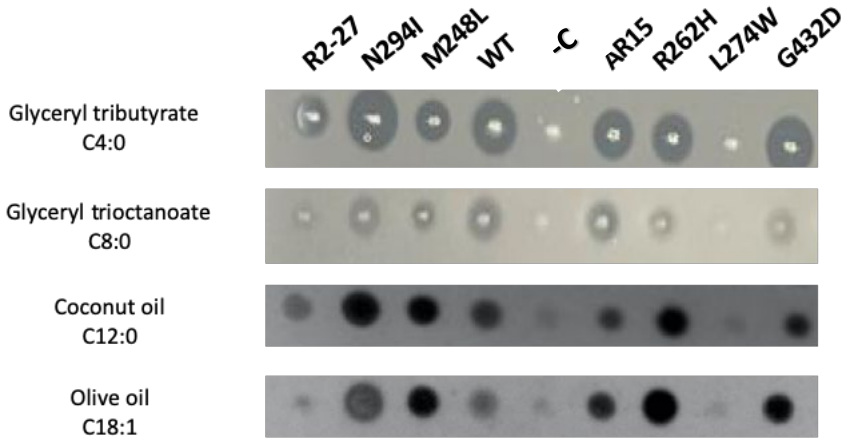


Figure 5.2. – Representative results of the qualitative screening of R2-27, AR15 and their deconvoluted mutants in agar emulsified with glyceryl tributyrate (C4:0), glyceryl trioctanoate (C8:0), coconut oil (C12:0) and olive oil (C18:1).

Hydrolysis of triglycerides produces a halo of clearance for glyceryl tributyrate and trioctanoate and a halo of fluorescence from the pH-sensing rhodamine included with coconut oil and olive oil. '-C' is the negative control for Cal-A expression.

The individual mutations M248L and N294I each result in a ratio of short-chain triglyceride to long-chain triglyceride hydrolysis similar to the wild-type (WT) Cal-A, contrary to R2-27, which has decreased activity towards fatty acids > C8 (Figure 5.2). This demonstrates that the increase in short-chain selectivity of R2-27 is due to the combination of M248L and N294I. Interestingly, mutation M248L appears to give the opposite phenotype: it displays greater long-chain hydrolysis than the WT. This suggests that sign epistasis, where the combination of mutations

flips the phenotypic effect of the single mutants, may underlie the interaction between mutations M248L and N294I that results in increased short-chain selectivity.

In the case of AR15, although its selectivity phenotype is weak, it shows higher activity towards long-chain olive oil than does the WT. The deconvolution suggests that both mutations R262H and G432D contribute to the improved long-chain selectivity of AR15. It should be noted that the single mutant L274W shows no activity for all substrates. As a decreased activity is not observed in the triple mutant AR15, this suggests that mutations R262H and/or G432D could stabilize or enhance function individually or in combination with L274W.

B) Quantitative assay for hydrolysis of single chain esters in solution

A quantitative assay of chain-length selectivity was performed using *p*-NO_2-phenyl esters, the gold standard lipase assay. The highly activated substrates and colorimetric detection of the aromatic leaving group ensure sensitive detection of low catalytic activities (44). Clarified lysates were tested against *p*-NO_2-phenyl esters of chain lengths ranging between C4:0 to C16:0. The results correlated well with the qualitative triglycerides screening assay above, despite the extensively differing chemical nature of the substrates (Figure 5.3; Supplemental Figure A3.1). The long-chain selective variant AR15 and each of its deconvoluted mutants display higher activity toward all single-chain esters assayed in *E. coli* lysate than does WT Cal-A, in addition to increased hydrolysis of the long-chain *p*-NO_2-phenyl palmitate (C16:0). In this assay, only the deconvolution mutant G432D displays clear, long-chain selectivity.

Conversely, short-chain selective R2-27 retained a WT level of activity towards *p*-NO_2-phenyl butyrate (C4:0); its activity for *p*-NO_2-phenyl esters longer than C8 decreased, thereby improving its short-chain selectivity (Figure 5.3). Again, neither of the single substituted variants M248L or N294I showed a similar phenotype, pointing to epistasis between the two substitutions in variant R2-27.

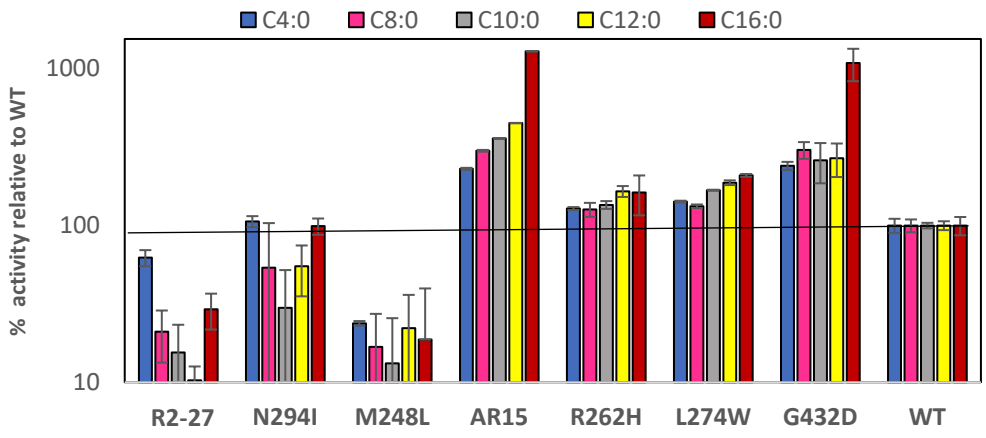


Figure 5.3. – Quantitative *p*-NO_2-phenyl ester assay for clarified lysates of R2-27, AR15 and their deconvoluted mutants.

Values are normalized against WT activity for each substrate; note the logarithmic scale of the y-axis.

C) Quantitative assay for hydrolysis of triglycerides in solution

Quantification of lipase activity using triglycerides has the main drawback of lower solubility and lower sensitivity than *p*-NO_2-phenyl esters. However, triglyceride activity results are more representative of potential applications using triglycerides or their derivatives (23, 24, 45). Here, triglyceride solubility is enhanced with the use of gum Arabic and colorimetric assessment of hydrolysis is enabled with inclusion of phenol red as a pH indicator (46). Quantitative analysis with triglycerides of chain lengths C4:0 and C8:0, and olive oil (predominantly C18:1), further confirmed the chain-length selectivity pattern observed for R2-27 (Figure 5.4; Supplemental Figure A3.2). In this experiment, single mutants N294I and M248L do show weak improvements in the selectivity profile compared to wild-type (Figure 5.4) which were not evident in the *in vivo* assay with triglycerides.

In accordance to the weak chain-length selectivity observed in the *in vivo* assay, and contrary to the *p*-NO_2-phenyl esters assays, AR15 barely shows improvements in its discrimination profile. In fact, based on this quantitative triglyceride assay, this variant would not be considered chain-selective. In addition, variants AR15 and G432D also do not have the enhanced activity observed

with p-NO_2-phenolates. Furthermore, the selectivity profile of G432D seems inversed to what it was observed in the *p*-NO_2-phenyl esters assays (Figure 5.4, Supplemental Figure A3.2). However, none of the single mutants composing variant AR15 show any significant improvements in the discriminative profile.

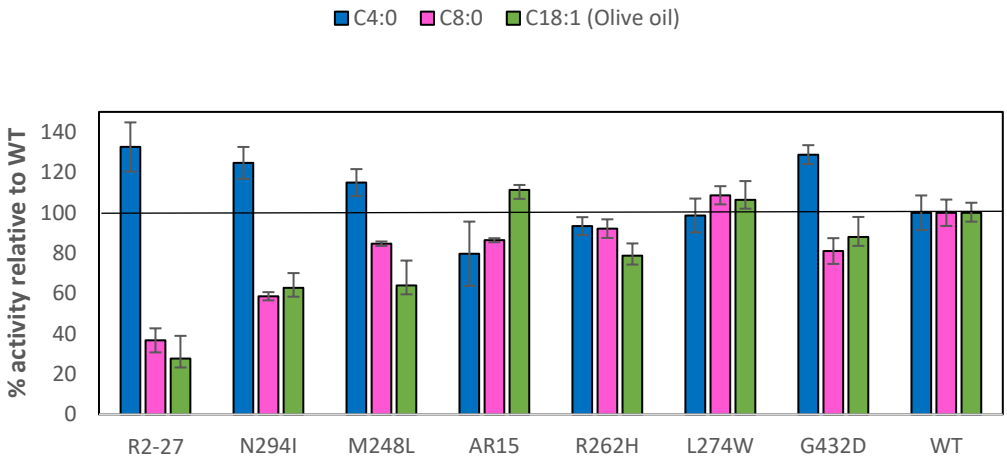


Figure 5.4. – Quantitative triglyceride assay for clarified lysates of R2-27, AR15 and its deconvoluted variants.

Values are normalized against WT activity for each substrate.

5.3.3 Structural analysis to decipher chain-length selectivity

To gain insight into the molecular basis of the chain-length selectivity observed in Cal-A and its variants R2-27 and AR15, they were subjected to computational tunnel analysis and to flexible docking. Tunnel analysis was done by computationally exploring the cavities identified starting from the active site in the absence of ligand with the software CAVER. This resulted in the identification of several cavities and characteristics such as their length, width or curvature (Supplemental Tables A3.1-4). Tunnel analysis was done for the crystallographic structure, as well as the structures of Cal-A wild type and its *in silico*-generated mutants that were relaxed by conformational sampling. Substrate docking with PELE permitted the exploration of different

protein and substrate conformations. The initial docking pose was generated with SwissDock by directing the ligand within 5 Å of the catalytic serine.

Docking of *p*-NO_2-phenyl octanoate into the relaxed structures that include the relevant mutations showed binding in tunnels that had been predicted in the absence of ligand (Figure 5.5). Specifically, *p*-NO_2-phenyl octanoate bound to the putative acyl-binding tunnel where the PEG molecule co-crystallized (27). In variant AR15, docking of *p*-NO_2-phenyl octanoate also occurred in an alternative, predicted tunnel (Figure 5.5). This alternative tunnel is present in the crystallographic WT structure prior to relaxation but is not present in the relaxed WT or R2-27 structures (Supplemental Figure A3.3; Supplementary Tables A3.1-4).

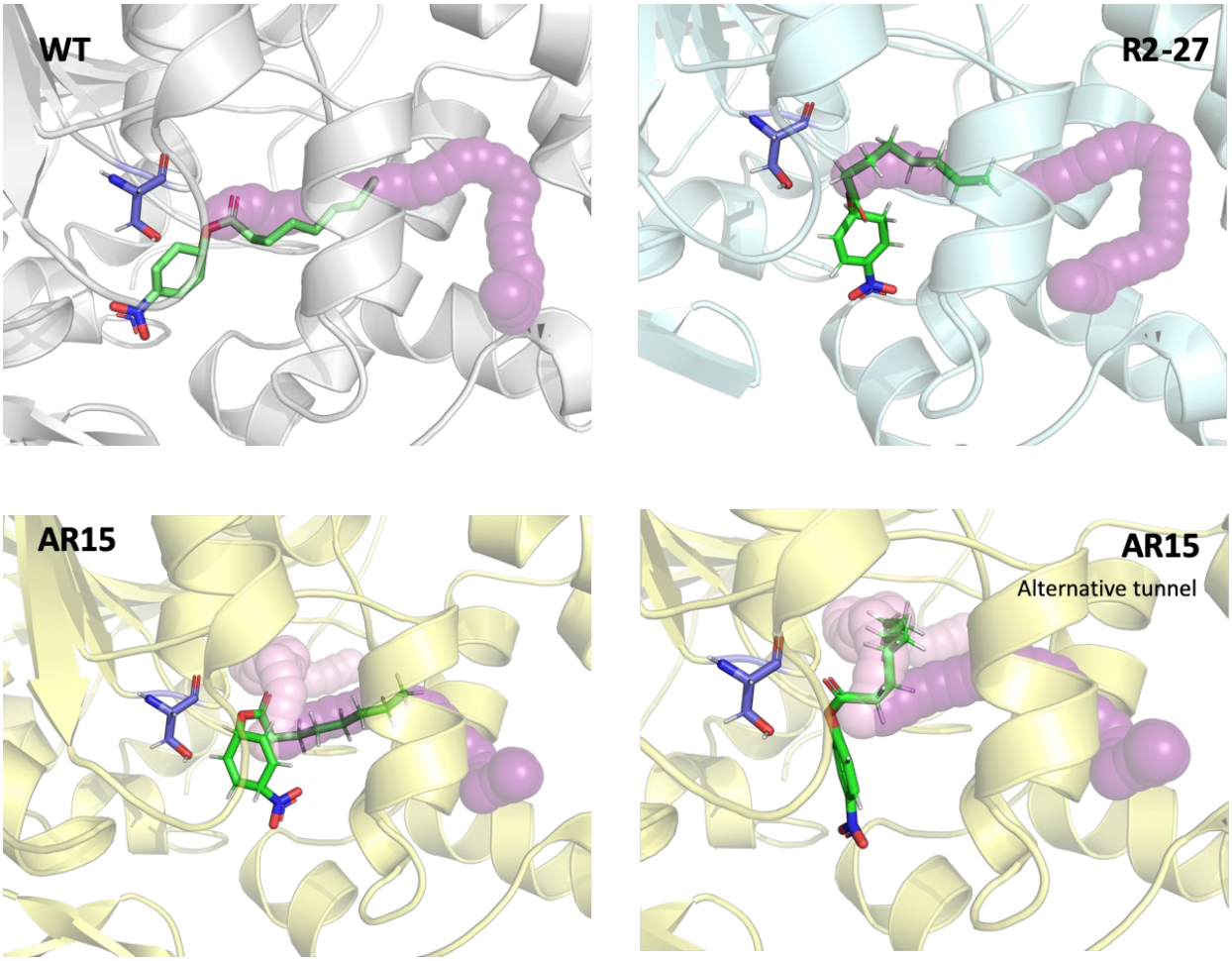


Figure 5.5. – Docking of *p*-NO_2-phenyloctanoate into WT Cal-A, R2-27 and AR15 superimposed with tunnels mapped in the relaxed structures in the absence of ligand.

Our tunnel analysis results are consistent with a previous report that Cal-A WT has a long acyl-binding tunnel with a kink towards the bottom (27). Interestingly, this curvature is more marked in the double mutant R2-27 (Table 5.1). This results in the tunnel becoming effectively shorter than in WT Cal-A and is consistent with the observed changes in chain-length selectivity. This would be an alternative mechanism to sterically blocking the acyl-binding tunnel to obtain short-chain selectivity in Cal-A (34). In contrast, the alternative tunnel in AR15 presents a length comparable to the main tunnel of Cal-A WT, with lower curvature (Table 5.1). Whereas AR15 does not show consistent chain-length discrimination, there are definite changes in activity towards *p*-NO_2-phenyl esters of fatty acid. Further engineering exploiting this alternative tunnel could lead to even greater discrimination.

Overall, the observed changes in the tunnel conformations would not have been easily predicted from modelling mutations into the crystallographic structure, emphasizing the need to thoroughly explore epistasis and its functional effects in Cal-A lipase. Prior identification of these mutations by directed evolution was key to detect novel chain-length selective variants.

Table 5.1. – Tunnel analysis of most probable tunnels.

Curvature is defined as the ratio between tunnel length and distance. Where distance is defined by CAVER as the shortest possible distance between the calculation starting point and the tunnel ending point. Bottleneck radius is the maximal probe size that can fit in the narrowest part of the tunnel.

	Length (Å)	Curvature	Bottleneck radius (Å)	Distance (Å)
WT	27.3	1.9	0.82	14.4
AR15	22.1	1.6	0.62	13.8
AR15 alternative	26.5	1.5	0.75	17.7
R2-27	29.5	2.3	0.61	12.8

5.4 Discussion

Cal-A lipase has been successfully engineered in previous reports to modify chain length-selectivity for biotechnological purposes (11, 23, 24, 30, 34). Mutations that confer short-chain triglyceride selectivity were found predominantly in a helix-loop-helix between residues S217-S245, near the entrance of the acyl-binding tunnel, when Cal-A was subjected to targeted random

mutagenesis (22). Other reports have also shown short-chain fatty acid selective mutations at the bottom of the acyl-binding tunnel (11, 30) (Figure 5.6).

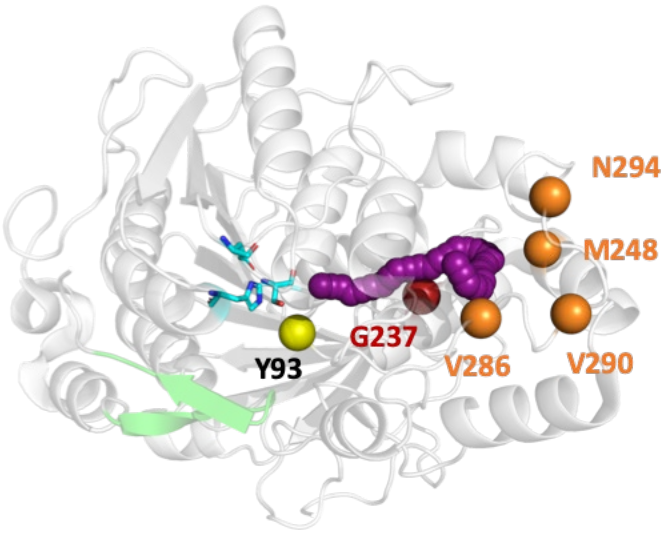


Figure 5.6. – Cal-A wild-type (PDB ID: 2veo) highlighting residues in which mutations that lead to chain-length selectivity have been identified.

Y93G in yellow confers long-chain selectivity, whereas the rest of highlighted residues confer short-chain selectivity. Mutations in G237 (red) confer short-chain selectivity by blocking the acyl-binding tunnel sterically. Mutations in residues highlighted in orange, confer short-chain selectivity through epistatic interactions. The catalytic triad is shown as cyan sticks and the small flap is in green.

As in other lipases, the relative position in the tunnel allows to modulate chain length-selectivity by narrowing the acyl-binding tunnel or blocking it sterically, allowing to create selectivity for increasingly shorter chain-lengths (32, 33, 35). In Cal-A in particular, mutations in residue G237 at the entrance of the tunnel prevent hydrolysis of fatty-acids longer than C6(34) whereas mutations in residues V290 and V286 at the bottom of the tunnel decrease hydrolysis of fatty acids longer than C18(11) (Figure 5.6). Here, we have characterized variant R2-27 which contains mutations N294I and M248L at the end of the acyl-binding tunnel. These mutations decrease hydrolysis of fatty acids longer than C8 in all of the activity assays performed in this study (Figures 5.2-5.4). This trend is shown in the double mutant but not when the mutations are present individually. Tunnel analysis reveals that the putative acyl-binding tunnel is effectively shorter due to its enhanced curvature (Table 5.1). This effect is difficult to predict from the three-

dimensional structure as these residues are ~14 Å away from each other and the incorporated substitutions do not reflect dramatic changes in the side-chain size that could approximate these regions. It is likely that more voluminous mutations in residues N294 and M248 could lead to higher chain-length selectivity. Indeed, a prior report has shown that introduction of bulky residues at the end of the tunnel decreased activity towards fatty acids longer than C18 due to shortening of the acyl-binding tunnel (11). It should be noted that Mg^{2+} has been proposed to bind near residue N294, which could have indirect effects on the reaction rate when mutating this residue (47). The interaction of mutations N294I and M248L at the end of the tunnel conferring short chain-length selectivity constitutes another example of non-additive interactions occurring in Cal-A lipase apart from mutations V290W and V286Q (11, 30) (Figure 5.6). These effects on activity might due to conformational dynamics as it has been reported for other lipases (48-50). However, to date a molecular dynamic characterization of Cal-A lipase to identify long-range networks is lacking. Further engineering efforts targeting this region through iterative saturation mutagenesis could aid in further modifying chain-length selectivity for biotechnological purposes.

Compared to other lipases, Cal-A has a characteristic long acyl-binding tunnel that could accommodate up to 20-25 methylene units (27) However, it is more proficient towards the hydrolysis of short-chain fatty acids (22,25,26). Contrary to short-chain discrimination, no hot-spot for long-chain fatty acid selectivity has been previously identified in Cal-A lipase (22). Long-chain fatty acid selective variants generally maintain activity towards short-chain fatty acids, making mechanism of long-chain selectivity more difficult to characterize (22). Cal-A compared to other lipases has a characteristic long acyl-binding tunnel that could accommodate up to 20-25 methylene units (27). However, it is more proficient towards the hydrolysis of short-chain fatty acids (22, 25, 26). As it has been described for other lipases, activity towards long-chain fatty acids can be increased by widening the entrance to the long acyl-binding tunnel (22, 35, 36). This is also the case for Cal-A where mutation Y93G, in the oxyanion hole near the catalytic triad, has been shown to increase long-chain fatty acid hydrolysis (22) (Figure 5.6). Other more complex mechanisms such as mutation-induced conformational changes that increase the active site volume, improve lid flexibility or stabilize open conformation have also been shown to favor long-

chain fatty acid selectivity in other lipases (37, 40, 41). In addition, lid swapping in other lipases has been shown to shift chain-length selectivity (38, 39, 51). In Cal-A, site-directed saturation mutagenesis at G431, near the flap, did not result in any chain-length discrimination (22). In this report we analyze the triple mutant AR15 which contains mutations R262H, L274W and G432D which had been shown to be discriminative in a qualitative assay with emulsified triglycerides (22), as confirmed in this study (Figure 5.2). Whereas quantitative assays with *p*-NO_2-phenyl esters were consistent with that phenotype (Figure 5.3), quantitative assays with triglycerides of varying chain-lengths did not reveal a marked discriminative profile of this variant. This variant showed the stabilization of an alternative acyl-binding tunnel through docking and tunnel analysis that might be involved in substrate recognition and binding. Performing further rounds of directed evolution could enhance the discriminative profile of AR15 toward *p*-NO_2-phenyl esters and triglycerides.

Overall, we report an epistatic interaction in a novel chain-length discriminative variant (R2-27) of Cal-A lipase that was first identified through focused random mutagenesis (22). Variant R2-27 displays a short-chain selective phenotype consistent with strangling the acyl-binding tunnel, even though the mutations are not particularly bulky. Interestingly, these epistatic mutations occur in the same region as previously identified chain-length selective synergistic mutations (52). This suggests that residue interactions in this region may be characterized by complex, non-additive effects. In addition, we report the stabilization of an alternative acyl-binding tunnel that might be implicated in substrate binding and recognition. Several interrogations remain in Cal-A lipase such as substrate recognition, conformational dynamics and the predominance of epistasis. Further studies addressing these questions will aid to accelerate engineering of novel variants.

5.5 Materials and methods

5.5.1 Mutagenesis deconvolution

The Cal-A construct used as a template for mutagenesis was codon-optimized for expression in *E. coli*. The pelB signal peptide was *N*-terminally fused for periplasmic export and a poly-hexa-histidine tag was *C*-terminally fused. The construct was cloned into plasmid pD441 as previously

reported (42). Cal-A variants with single amino acid substitutions M248L, N294I, L274W, R262H and G432D were created through whole-plasmid site-directed mutagenesis(53) using primers (Alpha DNA) M248L-Fwd (5'-GCACGACTTGCGGCCAGCAAAAGCCGCGACC-3'), M248L-Rev (5'-CGATGAAAGATTCCAAGTCCGGGTGCGCC-3'), N294I-Fwd (5'-GGTCAATGATACCATCTTGCTGAATGAAGC-3'), N294I-Rev (5'-GCTTCATTCAGCAAGATGGTATCATTGACC-3'), L274W-Fwd (5'-GGTCGCGGCTTTTGCTGGCCGCAAGTCGTGC-3'), L274W-Rev (5'-GCACGACTTGCGGCCAGCAAAAGCCGCGACC-3'), R262H-Fwd (5'-CGCAAAGGGTCAACATACGCTGAAGCAAATCCG-3'), R262H-Rev (5'-CGGATTTGCTTCAGCGTATGTTGACCCTTTGCG-3'), G432D-Fwd (5'-CAGAGCGCGTTTGATAAGCCGTTTGGTCCG-3') and G432D-Rev (5'-CGGACCAAACGGCTTATCAAACGCGCTCTG-3'). Phusion polymerase (Thermo Fisher) was used following the manufacturer's guidelines using ~100 ng of pD441-pelB-Cal-A as template in a 50 µL total volume. Adding a final concentration of 3% DMSO improved the reaction efficiency for mutations introduced in high GC-rich regions. After completion, the products were digested with DpnI (NEB) for 1h at 37°C and 5 µL were transformed into chemically competent *E. coli* BL21 (DE3) prepared following the $CaCl_2$ protocol (54). The transformed cells were plated onto Luria-Bertani agar-plates containing 50 µg/mL kanamycin as a selective marker. Plasmid DNA from colonies obtained after overnight incubation at 37°C was isolated using the Monarch Plasmid miniprep kit (NEB) and sequenced at the Genome Quebec Innovation Center (McGill University, Canada) using primers pFWD (5'-TTACGAGCTTCATGCACAG-3') and pRVS (5'- TGGTAGTGTGGGGACTC-3').

5.5.2 Protein expression and cell lysis

E. coli BL21(DE3) cell cultures transformed with the relevant Cal-A genes were grown to saturation at 37°C in Luria-Bertani medium supplemented with 50 µg/mL kanamycin. For quantitative activity assays using *p*-NO_2-phenyl esters, 100 µL of saturated cultures were used to inoculate 10 mL ZYP-5052 of auto-inducing medium (55) with kanamycin and grown at 22°C. After 24h, cultures were pelleted and resuspended in 1 mL of 50 mM Tris pH 8.0 as previously described (22). Following 20 s sonication pulses until clear, the lysate was centrifuged and the supernatant assayed.

For triglyceride lipase activity assays, the expression and lysis were optimized to increase activity. In this case, Luria-Bertani medium was used instead of auto-inducing medium and cells were grown at 37°C for 1-2h until an OD_{600} = 0.6 and induced with 0.5 mM IPTG overnight at 16°C. Cell cultures were pelleted at 3000 rpm and resuspended in 40 mM HEPES buffer at pH 7.0 to wash the pellet. The washed pellet was then stored at -20°C or directly resuspended in 250-500 µL 40 mM HEPES buffer at pH 7.0. Cell lysis was done by incubating at 37°C for 1h with 0.1 mg/mL polymyxin B solution.

Protein quantification of protein lysates was done using the Bradford Protein Assay Kit (Bio-Rad).

5.5.3 Lipase activity assays

A) Qualitative assay for hydrolysis of triglycerides on agar plates

The qualitative plate assays containing emulsions of glyceryl tributyrate, glyceryl trioctanoate, coconut oil or olive oil were as previously described with a few modifications (22). Briefly, 2 mL of the appropriate triglyceride were added to 50 mL of distilled water with 5 g of gum arabic. This solution was emulsified and added to 200 mL of auto-inducing medium with 2% agar (55) to obtain a final concentration of 1% v/v triglyceride and cast into rectangular Nunc OmniTray dishes (Thermo Fisher). Glyceryl tributyrate-containing plates are opaque and lipase activity is detected as a halo of clearance around the colony. Rhodamine B (0.001% w/v) was included in coconut oil and olive oil plates to detect lipase activity as a halo of fluorescence around the colony. *E. coli* lysates expressing Cal-A variants were spotted onto the plates. Following 18-24h incubation at 30°C, each variant was assigned an arbitrary activity value ranging between 1 and 4 relative to wild-type activity (which was assigned a value of 3). *E. coli* expressing TEM-1 ß-lactamase was used as a negative control for lipase activity.

B) Quantitative assay for hydrolysis of single chain esters in solution

The quantitative assay using *p*-NO_2-phenyl esters of fatty acids with chain lengths C4:0 to C16:0 was done as previously described (22). Briefly, *E. coli* lysates were assayed in a total reaction volume of 200 µL containing 0.5 mM *p*-NO_2-phenyl fatty acid and 0.1% Triton in 50 mM TrisHCl pH 8.0. Lipase activity was monitored as the increase in absorbance of *p*-NO_2-phenolate at 405

nm (ε_{405}=18.1 $mM^{-1}cm^{-1}$) upon its release from the corresponding fatty acid. The assay was performed in triplicate in 96-well plates at 45°C using a BMG FLUOstar 96-well plate spectrophotometer.

C) Quantitative assay for hydrolysis of triglycerides in solution

The quantitative assay using triglycerides of chain lengths C3:0-C8:0 and olive oil (mostly C18:1) was done using a previously described colorimetric assay with a few modifications (25, 56). Glyceryl tripropionate, tributyrate, trihexanoate, trioctanoate and olive oil were made up to 96 mM in acetonitrile, except for olive oil which was sonicated into 10% DMSO and 10% gum arabic. The reaction mixture was prepared by adding 2 µL of 96 mM triglyceride to 40 µL of 0.45 mM Phenol Red in 5 mM *N*-(2-hydroxyethyl) piperazine-*N*'-(3-propanesulfonic acid (EPPS) buffer at pH 8.0. Upon addition of 2 µL of polymyxin B-produced-lysates to the reaction mixture, fatty acid hydrolysis was monitored at 550 nm (ε_{550}=8.5 $mM^{-1}cm^{-1}$) upon color change of phenol red due to a change in pH. The assay was performed in triplicate in 384-well plates at 45°C in a Synergy HT Multi-Mode Microplate reader (Biotek).

In the quantitative assays, the inactive catalytic acid variant D334Y was used as a negative control, as well as a blank containing no lysate.

5.5.4 Protein docking

The different variants subjected to docking were generated *in silico* in PyMOL using the crystal structure 2veo corresponding to wild-type Cal-A. The structures were subsequently prepared in Maestro using Schrodinger's Preparation Wizard and eliminating solvent molecules and a co-crystallized PEG molecule. The prepared molecules were subjected to accurate docking using the online server SwissDock (http://www.swissdock.ch/docking). The coordinates chosen for the accurate docking correspond to Ser184. Flexibility of the side-chains was allowed within 5 Å of any atom of the ligand. Once a docking pose was identified, the protein with the ligand was subjected to binding refinement to obtain docking poses where the ester bond was >3.5 Å from the catalytic S184. Binding refinement was done using the online server for the Protein Energy Landscape Exploration algorithm (PELE) (https://pele.bsc.es/). PELE allows for sampling of the

conformational space by allowing protein and ligand perturbations (57). Analysis of the obtained poses was done using PyMOL and Maestro.

5.5.5 Protein tunnel calculations

Protein structures were prepared as for protein docking in the absence of ligand and solvent. The structures were relaxed using the protein motion algorithm in the PELE online server. Tunnels were calculated using CAVER 3.0 command-line version (58). The catalytic S184 was put as the starting point residue for tunnel calculations with a probe of 0.6 Å. A maximal distance of 5 Å from the starting point was allowed.

5.6 References

1. Nestl BM, Hammer SC, Nebel BA, Hauer B. New generation of biocatalysts for organic synthesis. Angew Chem Int Ed Engl. 2014;53(12):3070-95.
2. Abdelraheem EMM, Busch H, Hanefeld U, Tonin F. Biocatalysis explained: from pharmaceutical to bulk chemical production. React Chem Eng. 2019;4(11):1878-94.
3. Jackel C, Hilvert D. Biocatalysts by evolution. Curr Opin Biotechnol. 2010;21(6):753-9.
4. Mazurenko S, Prokop Z, Damborsky J. Machine learning in enzyme engineering. ACS Catal. 2019;10(2):1210-23.
5. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Engineering the third wave of biocatalysis. Nature. 2012;485(7397):185-94.
6. Maria-Solano MA, Serrano-Hervás E, Romero-Rivera A, Iglesias-Fernández J, Osuna S. Role of conformational dynamics in the evolution of novel enzyme function. Chem Commun. 2018;54(50):6622-34.
7. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, et al. The role of protein dynamics in the evolution of new enzyme function. Nat Chem Biol. 2016;12(11):944-50.
8. Wilding M, Hong N, Spence M, Buckle AM, Jackson CJ. Protein engineering: the potential of remote mutations. Biochem Soc Trans. 2019;47(2):701-11.
9. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. Proc Natl Acad Sci U S A. 2006;103(15):5869-74.

10. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. Nature. 2006;444(7121):nature05385.

11. Zorn K, Oroz-Guinea I, Brundiek H, Dorr M, Bornscheuer UT. Alteration of chain length selectivity of Candida antarctica lipase A by semi-rational design for the enrichment of erucic and gondoic fatty acids. Adv Synth Catal. 2018;360(21):4115-31.

12. Reetz MT. The importance of additive and non-additive mutational effects in protein engineering. Angew Chem Int Ed Engl. 2013;52(10):2658-66.

13. Tomatis PE, Fabiane SM, Simona F, Carloni P, Sutton BJ, Vila AJ. Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. Proc Natl Acad Sci U S A. 2008;105(52):20605-10.

14. Gobeil SMC, Clouthier CM, Park J, Gagne D, Berghuis AM, Doucet N, et al. Maintenance of native-like protein dynamics may not be required for engineering functional proteins. Chem Biol. 2014;21(10):1330-40.

15. Sato TK, Tremaine M, Parreiras LS, Hebert AS, Myers KS, Higbee AJ, et al. Directed evolution reveals unexpected epistatic interactions that alter metabolic regulation and enable anaerobic xylose use by Saccharomyces cerevisiae. PLoS Genet. 2016;12(10):e1006372.

16. Osuna S, Jimenez-Oses G, Noey EL, Houk KN. Molecular dynamics explorations of active site structure in designed and evolved enzymes. Acc Chem Res. 2015;48(4):1080-9.

17. Otten R, Liu L, Kenner LR, Clarkson MW, Mavor D, Tawfik DS, et al. Rescue of conformational dynamics in enzyme catalysis by directed evolution. Nat Commun. 2018;9(1):1314.

18. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. Biophysical principles predict fitness landscapes of drug resistance. Proc Natl Acad Sci U S A. 2016;113(11):E1470-8.

19. Zoi I, Antoniou D, Schwartz SD. Incorporating fast protein dynamics into enzyme design: A proposed mutant aromatic amine dehydrogenase. J Phys Chem B. 2017;121(30):7290-8.

20. Vicente AI, Viña-Gonzalez J, Mateljak I, Monza E, Lucas F, Guallar V, et al. Enhancing thermostability by modifying flexible surface loops in an evolved high-redox potential laccase. AlChE J. 2020;66(3).

21. Zimmerman MI, Hart KM, Sibbald CA, Frederick TE, Jimah JR, Knoverek CR, et al. Prediction of new stabilizing mutations based on mechanistic insights from Markov state models. ACS Cent Sci. 2017;3(12):1311-21.

22. Quaglia D, Alejaldre L, Ouadhi S, Rousseau O, Pelletier JN. Holistic engineering of Cal-A lipase chain-length selectivity identifies triglyceride binding hot-spot. PLoS One. 2019;14(1):e0210100.

23. Monteiro RRC, Virgen-Ortiz JJ, Berenguer-Murcia Á, Da Rocha TN, Dos Santos JCS, Alcántara AR, et al. Biotechnological relevance of the lipase A from Candida antarctica. Catal Today. 2020.

24. Domínguez De María P, Carboni-Oerlemans C, Tuin B, Bargeman G, Van Der Meer A, Van Gemert R. Biotechnological applications of Candida antarctica lipase A: State-of-the-art. J Mol Catal B: Enzym. 2005;37(1-6):36-46.

25. Martínez-Martínez M, Coscolín C, Santiago G, Chow J, Stogios PJ, Bargiela R, et al. Determinants and prediction of esterase substrate promiscuity patterns. ACS Chem Biol. 2018;13(1):225-34.

26. Robles-Machuca M, Del Campo MM, Camacho-Ruiz MA, Ordaz E, Zamora-Gonzalez EO, Muller-Santos M, et al. Comparative features between recombinant lipases CALA-like from U. maydis and CALA from C. antarctica in thermal stability and selectivity. Biotechnol Lett. 2019;41(2):241-52.

27. Ericsson DJ, Kasrayan A, Johansson P, Bergfors T, Sandström AG, Bäckvall J, et al. X-ray structure of Candida antarctica lipase A shows a novel lid structure and a likely mode of interfacial activation. J Mol Biol. 2008;376:109-19.

28. Widmann M, Juhl PB, Pleiss J. Structural classification by the lipase engineering database: a case study of Candida antarctica lipase A. BMC Genomics. 2010;11:123.

29. Nyyssölä A, Miettinen H, Kontkanen H, Lille M, Partanen R, Rokka S, et al. Treatment of milk fat with sn-2 specific Pseudozyma antarctica lipase A for targeted hydrolysis of saturated medium and long-chain fatty acids. Int Dairy J. 2015;41:16-22.

30. Zorn K, Oroz-Guinea I, Bornscheuer UT. Strategies for enriching erucic acid from Crambe abyssinica oil by improved Candida antarctica lipase A variants. Process Biochem. 2019;79:65-73.

31. SÁ AGA, Meneses ACd, Araújo PHHd, Oliveira Dd. A review on enzymatic synthesis of aromatic esters used as flavor ingredients for food, cosmetics and pharmaceuticals industries. Trends Food Sci Technol. 2017;69:95-105.

32. Schmitt J, Brocca S, Schmid RD, Pleiss J. Blocking the tunnel: engineering of Candida rugosa lipase mutants with short chain length specificity. Protein Eng. 2002;15(7):595-601.

33. Klein RR, King G, Moreau RA, Haas MJ. Altered acyl chain length specificity of Rhizopus delemar lipase through mutagenesis and molecular modeling. Lipids. 1997;32(2):123-30.

34. Brundiek H, Padhi SK, Kourist R, Evitt A, Bornscheuer UT. Altering the scissile fatty acid binding site ofCandida antarcticalipase A by protein engineering for the selective hydrolysis of medium chain fatty acids. Eur J Lipid Sci Technol. 2012;114(10):1148-53.

35. Yang J, Koga Y, Nakano H, Yamane T. Modifying the chain-length selectivity of the lipase from Burkholderia cepacia KWI-56 through in vitro combinatorial mutagenesis in the substrate-binding site. Protein Eng Des Sel. 2002;15(2):147-52.

36. Panizza P, Cesarini S, Diaz P, Rodriguez Giordano S. Saturation mutagenesis in selected amino acids to shift Pseudomonas sp. acidic lipase Lip I.3 substrate specificity and activity. Chem Commun (Camb). 2015;51(7):1330-3.

37. Gaskin DJ, Romojaro A, Turner NA, Jenkins J, Vulfson EN. Alteration of lipase chain length specificity in the hydrolysis of esters by random mutagenesis. Biotechnol Bioeng. 2001;73(6):433-41.

38. Brocca S, Secundo F, Ossola M, Alberghina L, Carrea G, Lotti M. Sequence of the lid affects activity and specificity of Candida rugosa lipase isoenzymes. Protein Sci. 2003;12(10):2312-9.

39. Yu XW, Zhu SS, Xiao R, Xu Y. Conversion of a Rhizopus chinensis lipase into an esterase by lid swapping. J Lipid Res. 2014;55(6):1044-51.

40. Syal P, Verma VV, Gupta R. Targeted mutations and MD simulations of a methanol-stable lipase YLIP9 from Yarrowia lipolytica MSR80 to develop a biodiesel enzyme. Int J Biol Macromol. 2017;104(Pt A):78-88.

41. Santarossa G, Lafranconi PG, Alquati C, DeGioia L, Alberghina L, Fantucci P, et al. Mutations in the "lid" region affect chain length specificity and thermostability of a Pseudomonas fragi lipase. FEBS Lett. 2005;579(11):2383-6.

42. Quaglia D, Ebert MCCJC, Mugford PF, Pelletier JN. Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening. PLoS ONE. 2017;12(2):e0171741.

43. Brundiek HB, Evitt AS, Kourist R. Creation of a lipase highly selective for trans fatty acids by protein engineering. Angew Chem Int Ed. 2012;51:412-4.

44. Stoytcheva M, Montero G, Zlatev R, Leon JA, Gochev V. Analytical methods for lipases activity determination: A review. Curr Anal Chem. 2012;8(3):400-7.

45. Houde A, Kademi A, Leblanc D. Lipases and their industrial applications. Appl Biochem Biotechnol. 2004;118(1):155-70.

46. Reyes-Duarte D, Coscolin C, Martinez-Martinez M, Ferrer M, Garcia-Arellano H. Functional-based screening methods for detecting esterase and lipase activity against multiple substrates. Methods Mol Biol. 2018;1835:109-17.

47. Sirén S, Dahlström KM, Puttreddy R, Rissanen K, Salminen TA, Scheinin M, et al. Candida antarctica Lipase A-based enantiorecognition of a highly strained 4-dibenzocyclooctynol (dibo) used for PET imaging. Molecules. 2020;25(4):879.

48. Tyukhtenko S, Rajarshi G, Karageorgos I, Zvonok N, Gallagher ES, Huang H, et al. Effects of distal mutations on the structure, dynamics and catalysis of human monoacylglycerol lipase. Sci Rep. 2018;8(1).

49. Reetz MT, Puls M, Carballeira JD, Vogel A, Jaeger K-E, Eggert T, et al. Learning from directed evolution: Further lessons from theoretical investigations into cooperative mutations in lipase enantioselectivity. ChemBioChem. 2007;8(1):106-12.

50. Xia Q, Ding Y. Thermostability of Lipase A and dynamic communication based on residue interaction network. Protein Peptide Lett. 2019;26(9):702-16.

51. Karkhane AA, Yakhchali B, Jazii FR, Bambai B. The effect of substitution of Phe181 and Phe182 with Ala on activity, substrate specificity and stabilization of substrate at the active site of Bacillus thermocatenulatus lipase. J Mol Catal B: Enzym. 2009;61(3):162-7.

52. Oroz-Guinea I, Zorn K, Bornscheuer T. Enhancement of lipase CAL-A selectivity by protein engineering for the hydrolysis of erucic acid for crambe oil. Eur J Lipid Sci Technol. 2019.

53. Laible M, Boonrod K. Homemade site directed mutagenesis of whole plasmids. J Vis Exp. 2009(27).

54. Sambrook J, Russell DW. Preparation and transformation of competent E. coli using calcium chloride. CSH Protoc. 2006;2006(1).

55. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005;41(1):207-34.

56. Janes LE, Löwendahl AC, Kazlauskas RJ. Quantitative screening of hydrolase libraries using pH indicators: Identifying active and enantioselective hydrolases. Chemistry – A European Journal. 1998;4(11):2324-31.

57. Madadkar-Sobhani A, Guallar V. PELE web server: atomistic study of biomolecular systems at your fingertips. Nucleic Acids Res. 2013;41(Web Server issue):W322-8.

58. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol. 2012;8(10):e1002708.

Chapter 6 – Discussion, conclusions and perspectives

The general framework of this thesis was to contribute to understand essential features of proteins in order to enhance predictability in enzyme engineering. The effective search of the sequence space to identify the best variant has several bottlenecks, both technical and conceptual, as highlighted in the Chapters 2, 3 and 4. To surmount some of those limitations, successful engineering efforts typically use enzymes that exhibit high stability, tolerance to new mutations and substrate promiscuity (21). Here, we used two model systems that possess those features to gain knowledge that can be applied in future enzyme engineering quests in these or related systems. Below, I will present the global conclusions, contributions, limitations and future work related to each system used and will discuss the perspectives that arise from this work.

6.1. Model 1: Dynamic ß-lactamase variants

In Chapter 3, the evolution of dynamic variants of TEM-1 ß-lactamase through random mutagenesis and by introducing specific epistatic mutations has revealed yet another feature of the robustness of this enzyme: the tolerance to slow-timescale dynamics in the evolution towards new function. Based on previous studies using TEM-1 ß-lactamase, we had initially hypothesized that differences in protein dynamics would alter the course of protein evolution. One previous study had pointed to increased protein dynamics in ancestral and promiscuous ß-lactamases, followed by a rigidification and specialization of the modern TEM-1 variant (93). Another study had revealed an evolutionary dead-end due to increased dynamics in the omega-loop region of TEM-1 ß-lactamase (78). Whereas these prior studies specified the dynamic regions, they did not specify the particular timescales studied. However, as mentioned in Chapters 1 and 3, protein dynamics are complex: they exist in different regions of a protein and at broadly differing timescales.

The strength of our model lies in the prior characterization of the dynamic landscape from *ps* to *ms* of two dynamic ß-lactamase chimeras presenting new motions in different regions (80). Certainly, the main limitation that arises from our model system is that these differences in

dynamics are a consequence of sequence differences that could also have an effect on further protein evolution. This limitation cannot be circumvented. Interestingly, Chapter 3 demonstrated that the extent of the sequence differences did not correlate with differences in the evolution towards new function: the outcome of cTEM-17m (with 17 mutations) was more similar to TEM-1 than was cTEM-2m (with 2 mutations).

In Chapter 3, it was shown that the patterns of protein motions specific to ß-lactamase variants cTEM-17m and cTEM-2m allowed evolution along an evolutionary path that is known to the more rigid TEM-1 ß-lactamase. Positive epistasis was maintained despite large variations in initial protein dynamics at the ms to μs timescale, showing a rather neutral effect in evolution of new protein function. Although it is plausible that motions at this timescale in other regions are not tolerated, the work presented in this thesis is a stepping stone in dissecting the relevance of protein dynamics on evolution in TEM-1 ß-lactamase. Further work in this direction could explore protein motions in different regions on the same timescale, or protein motions at different timescales, to have a complete picture of the relevant motions at the outset of evolution in this system. Although this is conceptually straightforward, it is experimentally complex since it is not yet possible to engineer new motions into proteins in a directed, predictable manner.

As discussed in Chapter 1, distinct motional events occur at different timescales. For example, side-chain rotations occur at the *ps* to *ns* timescales whereas loop rearrangements occur at the *ms* to *μs* timescale (79). Combined, motions at all timescales are relevant to the maintenance of the native structure. These motional events concur with relevant aspects of function such as ligand binding or substrate turnover that occur at the *ms* to *μs* timescale. It is probable that the modest impact on function of increased dynamics in variants cTEM-2m and cTEM-17m is due to the preservation of key motions necessary for catalysis. Nonetheless, a limitation in the field of protein dynamics has been monitoring motions during the catalytic cycle to decipher the key motions associated to catalysis. This is also a limitation to the prior characterization of the dynamic variants used in this thesis as their dynamic landscape was determined for the ligand-free enzyme (80), thereby setting a boundary to our interpretation of the structure/function/dynamics relation in TEM-1 ß-lactamase.

The structural determination of the inhibitor-bound and -free enzyme has been used in some studies to circumvent this limitation (78, 81). Recently, hydrogen/deuterium-exchange mass spectrometry has allowed to characterize dynamics of TEM-1 ß-lactamase at the millisecond timescale during catalysis (141). This has shown that motions in conserved regions such as the Ser70-X-X-Lys73 motif are associated with ligand-binding, in residues 240-247 to acylation or in residues 250-257 for deacylation of cephalosporins. Additional studies of the dynamic variants cTEM-2m and cTEM-17m using this methodology could reveal whether these key motions are conserved in the variants in the presence of the ligand.

Another layer of complexity to protein dynamics is whether relevant conformational changes occur prior to binding a ligand (conformational selection) or after binding (induced fit). The structural characterization of ligand-free enzymes has made conformational selection the most accepted model because the dynamic nature of enzymes has been revealed (99, 177). A greater extent or distribution of protein dynamics signifies a larger conformational exploration of enzymes. However, it has been shown that both mechanisms can coexist in several proteins (178, 179) and it has been proposed that one model might be favored over the other depending on volume of the ligand (180). Prior studies in TEM-1 ß-lactamase also support coexistence of conformational selection and induced fit. MD simulations have supported the induced-fit model upon binding of ampicillin and amoxicillin (181) whereas NMR spectroscopy has shown a correlation of certain motions of the ligand-free protein to catalytic efficiency towards benzylpenicillin (131). Whether one mechanism dominates over the other is an ongoing debate that is not the object of this thesis. However, determining the relevant motions occurring during catalysis for different ligands in the dynamic variants could help to dissect which mechanism is prevalent for each ligand. This information will help to further decipher the key motions that need to be preserved to maintain a specific function and are thus essential to evolve new functions.

In the context of evolving new protein function, the conformational selection model has been proposed as the most feasible (96, 99). In TEM-1 ß-lactamase, conformational selection has been demonstrated to occur in evolution towards hydrolysis of cefotaxime, the same substrate used to evolve the dynamic ß-lactamases in Chapter 3. Hart, Ho (140) showed that cefotaximase activity could be evolved by introducing mutations that stabilized pre-existent conformers in wild-

type TEM-1 that conferred cefotaximase activity. The evolved cefotaximase conformers were already present, albeit as minor conformers, in the unevolved enzyme. The results shown in Chapter 3 using Protein Energy Landscape Exploration (PELE) simulations in the presence of cefotaxime also point to the stabilization of cefotaximase conformers upon the introduction of the synergistic mutations E104K/G238S. These results remove the limitation of not knowing the dynamics of variants cTEM-2m and cTEM-17m in the presence of the ligand. Whether other mechanisms are implicated in evolution towards further ligands is a subject of future work. Perhaps the new motions present in cTEM-2m and cTEM-17m are neutral in the evolution of cefotaximase activity but will modulate evolution towards other ligands.

Protein epistasis is a mechanism by which protein dynamics can modulate the course of evolution. An additional contribution extracted from Chapter 3 is the observation that increased protein dynamics on the *ms* to *μs* timescale in cTEM-2m and cTEM-17m conserved the synergistic effect of mutations E104K/G238S. This result, taken with the predominance of mutations that appear in TEM-1 evolution, shows that similar mutational paths can be taken despite these differences in dynamics. Nonetheless, it is plausible that not all epistatic interactions would be maintained in the dynamic variants. Future work could address this by introducing other known epistatic interactions, such as A237T/R164S, A237T/E240K or R164S/G238S(126), or by conducting a thorough study of pairwise epistasis using whole-gene saturation mutagenesis in localized regions(40).

The analysis of protein dynamics during the course of directed evolution has provided with insights into its interplay with epistasis to modulate evolutionary trajectories (58, 99). Campbell, Kaltenbach (54) analyzed the conformational changes that occurred during the evolution of a phosphodiesterase to an aryl esterase and back to phosphodiesterase. Analysis of X-ray crystal structures showed that the evolution of new function happened through conformational changes associated to epistatic interactions. They also showed that the stabilized conformations were pre-existing, further confirming the conformational selection hypothesis in evolution. In Chapter 3, the dynamics of the evolved dynamic variants were not thoroughly characterized since the question was whether dynamic features of the evolutionary start-points modulated the course of evolution. Nonetheless, PELE simulation results in Chapter 3 give us some insights into dynamics

of the evolved variants. They suggest that despite the increased dynamics in the start-points, i.e. variants cTEM-2m and cTEM-17m, the Ω-loop is stabilized by mutations E104K/G238S.

Overall, the results presented in Chapter 3 demonstrate the tolerance of motions at the *ms* to *μs* timescale in the evolution of TEM-1 ß-lactamase towards cefotaximase activity. As discussed here, protein dynamics are an inherent and complex trait and further work is needed to extend our knowledge on the relationship between protein dynamics and evolution. TEM-1 ß-lactamase variants have served as an excellent model to study this relationship due to the extensive data available: TEM-1 is a gold standard in molecular evolution and has been thoroughly characterized structurally and kinetically. While protein engineering of TEM-1 is not a goal in the biotechnological industry, the conclusions drawn from this and similar studies will help to identify relevant features for optimal start-points in enzyme evolution.

6.2. Model 2: Cal-A lipase

Unlike TEM-1 ß-lactamase and despite its biotechnological interest, Cal-A lipase remains poorly characterized. In Chapters 4 and 5, the use of focused mutagenesis and deconvolution in Cal-A lipase has improved our knowledge of substrate binding and recognition. In particular, residues 217-245 in the tunnel region were identified as a hot-spot for triglyceride recognition in Chapter 4. The Cal-A variants we obtained were screened on agar plates containing emulsions of a short-chain or a long-chain triglyceride. This allowed the rapid identification of variants that hydrolyzed one substrate preferentially over the other. The main limitation of that study was, in fact, the qualitative nature of the screening. Nonetheless, as pointed out in Chapter 4, the qualitative assay allowed identification of highly chain-selective variants that had been previously reported in other studies as well as new ones. Therefore, the qualitative nature of the assay did not hinder directed evolution toward new selectivity. In addition, the qualitative assay has the advantage of using natural triglyceride substrates that are relevant in industrial settings, rather than using substrate analogs that may make the course of evolution deviate from the true objective.

The triglycerides used in Chapter 4 were glyceryl tributyrate (C4:0) and olive oil. The latter contains triglycerides predominantly constituted by oleic acid (C18:1). These two triglycerides have been used extensively to determine chain-length selectivity in lipases (182, 183). It should

be noted that although olive oil is a heterogeneous mix of other long-chain fatty acids, mainly in triglyceride form, it can be easily emulsified for the agar screening, contrary to the homogeneous triglycerides of the same chain-length. The other long-chain fatty acids contained in olive oil are linoleic acid (C18:2), palmitic acid (C16:0), stearic acid (C18:0), linolenic acid (C18:3) or arachidic (C20:0) (184). A limitation that arises from using olive oil is that oleic acid is a *cis* unsaturated acid, and Cal-A shows preference towards trans fatty-acids (149). Therefore, using olive oil means selecting for long-chain hydrolysis and *cis* selectivity simultaneously. To address this limitation, an additional assay using a triglyceride containing saturated fatty acids of intermediate chain-length, glyceryl trioctanoate (C8:0) and coconut oil (C12:0), were included when characterizing chain-length selective variants in Chapter 5. The emulsion of triglycerides containing fatty acids of higher chain-lengths for the agar screening was not accomplished due to their strong propensity to separate.

The most common approach for the quantitative screening of lipases is to use *p*-NO_2-phenyl esters as substrates. However, the quantification of chain-length selectivity by that approach has been limited by the chemical nature of *p*-NO_2-phenyl esters. Although of *p*-NO_2-phenyl esters are the gold-standard for industrial determination of lipase activity, their main disadvantages to quantify chain-length selectivity toward natural triglycerides have been described in Chapters 4 and 5. The limitations resulting from the *p*-NO_2-phenyl ester assay were addressed in Chapter 5 by using a high-throughput triglyceride hydrolysis assay. The main disadvantage in our case was the low activity of Cal-A which were due to low expression yields. Low expression of Cal-A was due to its expression in *E. coli*. This has made protein purification difficult and resulted in low activity in the quantitative triglyceride assays. *E. coli* was chosen as a host because although Cal-A has a eukaryotic origin, previous research had shown that *E. coli* could be used for directed evolution and expression of Cal-A (147). However, further characterization of Cal-A variants would benefit from expression in yeast hosts such as *S. cerevisiae* or *P. pastoris*.

Contrary to the ß-lactamase system, no thorough study characterizing the prevalence of epistasis in Cal-A lipase has been done to date. The epistatic interaction at the end of the acyl-binding tunnel described in Chapter 5 reveals how little is known about substrate binding or key structural features in Cal-A. Whereas most directed evolution studies done in Cal-A have a specific

biotechnological goal in mind, preliminary studies to decipher the evolvability of Cal-A would be of great use. It is plausible that the fitness landscape of Cal-A is too constrained by epistasis (Figure 1.3) and that reaching new activities would be challenging. In the absence of a thorough characterization of epistasis, it is encouraging to have observed that one round of directed evolution in Chapter 4 sufficed to obtain chain-length selective variants. Further rounds of evolution could yield variants with higher selectivity. Alternatively, iterative saturation mutagenesis in the acyl-binding tunnel could be an interesting tool to obtain additional variants while exploring epistatic interactions.

Overall, the results presented in Chapters 4 and 5 reveal relevant aspects for the evolution of chain-length selectivity for an enzyme that remains largely uncharacterized. These aspects are the existence of a triglyceride binding hot-spot and of an epistatic interaction at the end of the acyl-binding tunnel. Contrary to TEM-1 ß-lactamase, Cal-A lipase is poorly characterized yet has direct biotechnological applications. The results presented here highlight the need to further characterize this system to evolve it more effectively towards new industrial applications.

6.3. Considerations that are common to model 1 and model 2

An important aspect that emerges from both model systems used is that the method chosen to screen for activity has a great influence on the inferences of the study; as the common saying goes enzyme engineering, “You get what you screen for”. Two significant parameters of activity assays that can be extracted from this thesis are: the substrate chosen and enzyme purity.

When planning a directed evolution experiment, two paths can be chosen: either several substrates are used to investigate promiscuity or a single substrate is used to evolve selectivity. Both paths were used to evolve Cal-A in this thesis, whereas only the latter was applied for the dynamic variants of TEM-1. The substrate(s) chosen will have a defined chemical nature, such as hydrophobicity and three fatty acid chains in triglycerides or a ß-lactam ring and a bulky side-chain in cephalosporins. This is translated in selection towards improved binding or catalytic turnover for substrates with specific features. Extrapolations to other substrates can be made when the rate-limiting step of the reaction, chemical nature or binding mechanism of the substrate are similar.

For example, in TEM-1 ß-lactamases the rate-limiting step of cephalosporin hydrolysis is the acylation whereas for penicillin hydrolysis it is deacylation. Induced-fit has been proposed to occur in the binding of some penicillins but not of cephalosporins. Therefore, results will likely differ in terms of epistasis and dynamics when using penicillins *vs* cephalosporins. Overall, results for one type of substrate cannot generally predict activity for another substrate. However, the more knowledge there is about the reaction mechanism and substrate recognition, the better the quality of extrapolations to other substrates.

The second parameter to be considered is enzyme purity. This can broadly span *in vivo* overexpression and clarified cell lysates all the way to pure enzyme for *in vitro* assays. Enzymes present in the cellular milieu interact with numerous proteins that may favour protein stability or on the contrary favour degradation. Chaperones are a great example, since these proteins aid in protein folding and can be even used in evolution to diminish the effects of destabilizing mutations (62). Following the example of chaperones, the destabilizing effects of a mutation might be apparent in *in vitro* assays using purified enzyme but not in *in vivo* assays due to the buffering effect of chaperones. As a result, an epistatic interaction with this mutation will only be apparent *in vitro*. In another example, the catalytic improvement of a certain mutation might be apparent *in vitro* but not sufficient to procure phenotypic effects *in vivo*, as observed for cTEM-2m variant E104K/G238S in Chapter 3. In that case, if the activity had been measured exclusively *in vivo*, the epistatic effect of mutations E104K and G238S would have been disregarded.

Other interactions and other conditions of the cellular milieu can make protein stability and function higher *in vivo* or in cellular lysates than in the purified enzyme. As an example, antibiotic resistance profiles of metallo-ß-lactamases have been shown to correlate to activity values measured in cell lysates but not using purified enzyme (95). The cellular milieu provides conditions such as protein interactions, cofactors or a buffered medium that best suit the protein functionality. *In vivo* or cell lysate assays offer the additional advantage of not requiring lengthy purification. These reasons have made *in vivo* assays or cell extract assays attractive to accelerate protein discovery and its application in industry (185). This supports the use of *in vivo* assays in Chapter 4 and 5 of this thesis to measure Cal-A activity. However, further characterization of protein epistasis in Cal-A will need the study of purified variants. As illustrated above and

throughout this thesis, it is important to consider the question being asked and the ultimate goal in order to establish the appropriate level of enzyme purity to use in the assays: *in vivo*, cell lysates or purified enzyme.

An additional parameter to consider in both models is the choice of library creation methods. Random mutagenesis in dynamic ß-lactamases allowed exploration of evolutionary paths in Chapter 3 whereas focused, random mutagenesis in Cal-A lipase has helped uncover a key region for triglyceride recognition in Chapter 4. These mutational choices were deemed to be adequate to explore the different questions that each system posed. However, as reviewed in Chapter 2, other choices could have been made or can be done to further explore these models or other questions. Here, I briefly mention two promising methods that could be applied to further decipher the prevalence and relevance of epistasis in the models studied. A recent published book chapter has proposed a pipeline to analyse interaction networks in small libraries (53). In this method, the statistical analysis of a small combinatorial library of 6 residues yielding 64 variants served to describe high-order epistatic networks. Another method that could be of use for both model systems is deep mutational scanning to assess the sequence variations tolerated at each position (186). This could serve to explore whether the same mutational changes are tolerated in the dynamic ß-lactamases and other mutational hot-spots determinant for function in Cal-A.

Overall, the experimental design chosen contributed to expand our knowledge in both TEM-1 and Cal-A. Nonetheless, the results obtained from this thesis can be expanded using alternative experimental designs.

6.4. Perspectives

The inherent structural complexity of enzymes adds a great challenge to enzyme engineering. Successfully predicting which mutations will promote a new function needs prior knowledge of structure-function-evolution relationships. Directed evolution schemes often depict a mutated gene and the selection of a mutated protein variant. Yet a more accurate representation would include the structural modifications that are associated with the changes in function. By thoroughly characterizing which conformations lead to a certain function in directed evolution, we can start to dissect the most relevant aspects in enzyme engineering. However, this can only

be achieved once we possess extensive knowledge on the reaction mechanism, key structural features and prior successful mutations for individual systems. Some structure-function aspects can be acquired while evolving an enzyme towards a new function. However, careful structural characterization requires expertise of numerous methods such as X-ray crystallography, NMR and MD simulations that require development to be conducted with much higher throughput. Even when there is a careful structural characterization, epistatic interactions that guide evolution are challenging to predict. Ultimately, an optimal dissection of the key features to engineer in enzymes needs a multidisciplinary approach that also includes molecular evolution and enzymology.

Increasing predictability in enzyme engineering is a difficult task that is being improved with methodological and technological development. As shown in Chapter 2, library creation methods continue to be developed to better address current challenges in enzyme engineering such as the relevance of protein epistasis in evolution. Results such as those shown in Chapters 3, 4 and 5 allow to identify key regions and features to evolve new functions. Whereas these results belong to individual enzyme systems, common denominators will continue to be extracted with additional studies of these and other enzymes. Studies of individual and dissected systems are needed in order to identify key features. In this sense, the use of artificial intelligence to predict improved variants from current experimental knowledge is gaining popularity (24, 26). Computational development will also aid in performing detailed structural studies in a high-throughput fashion (25, 187-190). Altogether, multidisciplinary efforts are in place to increase predictability in enzyme engineering one step at a time.

Bibliographic references

1. Benner SA. Defining life. Astrobiology. 2010;10(10):1021-30.
2. Yarus M. Getting past the RNA world: The initial darwinian ancestor. Cold Spring Harb Perspect Biol. 2011;3(4):a003590-a.
3. Adamski P, Eleveld M, Sood A, Kun Á, Szilágyi A, Czárán T, et al. From self-replication to replicator systems en route to de novo life. Nat Rev Chem. 2020;4(8):386-403.
4. Edwards DR, Lohman DC, Wolfenden R. Catalytic proficiency: the extreme case of S-O cleaving sulfatases. J Am Chem Soc. 2012;134(1):525-31.
5. Albery WJ, Knowles JR. Evolution of enzyme function and the development of catalytic efficiency. Biochemistry. 1976;15(25):5631-40.
6. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry. 2011;50(21):4402-10.
7. Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci U S A. 1993;90(12):5618-22.
8. Ali M, Ishqi HM, Husain Q. Enzyme engineering: Reshaping the biocatalytic functions. Biotechnology and Bioengineering. 2020;117(6):1877-94.
9. Wilding M, Hong N, Spence M, Buckle AM, Jackson CJ. Protein engineering: the potential of remote mutations. Biochem Soc Trans. 2019;47(2):701-11.
10. Turner NJ. Directed evolution drives the next generation of biocatalysts. Nat Chem Biol. 2009;5(8):567-73.
11. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Engineering the third wave of biocatalysis. Nature. 2012;485(7397):185-94.
12. Li Y, Cirino PC. Recent advances in engineering proteins for biocatalysis. Biotechnology and Bioengineering. 2014;111(7):1273-87.
13. Jackel C, Hilvert D. Biocatalysts by evolution. Curr Opin Biotechnol. 2010;21(6):753-9.

14. Davids T, Schmidt M, Böttcher D, Bornscheuer UT. Strategies for the discovery and engineering of enzymes for biocatalysis. Curr Opin Chem Biol. 2013;17(2):215-20.

15. O'Reilly E, Ryan J. Biocatalytic cascades go viral. Science. 2019;366(6470):1199-200.

16. Baker M. Protein engineering: navigating between chance and reason. Nat Methods. 2011;8(8):623-6.

17. Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016;537(7620):320-7.

18. Chica RA, Doucet N, Pelletier JN. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol. 2005;16(4):378-84.

19. Ruff AJ, Dennig A, Schwaneberg U. To get what we aim for - progress in diversity generation methods. FEBS J. 2013;280(13):2961-78.

20. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH. Evolving strategies for enzyme engineering. Curr Opin Struct Biol. 2005;15(4):447-52.

21. Trudeau DL, Tawfik DS. Protein engineers turned evolutionists-the quest for the optimal starting point. Curr Opin Biotechnol. 2019;60:46-52.

22. Swint-Kruse L. Using evolution to guide protein engineering: The Devil IS in the details. Biophys J. 2016;111(1):10-8.

23. Yang G, Miton CM, Tokuriki N. A mechanistic view of enzyme evolution. Protein Science. 2020;29(8):1724-47.

24. Siedhoff NE, Schwaneberg U, Davari MD. Chapter Twelve - Machine learning-assisted enzyme engineering. In: Tawfik DS, editor. Methods in Enzymology. 643: Academic Press; 2020. p. 281-315.

25. Ebert MC, Pelletier JN. Computational tools for enzyme improvement: why everyone can - and should - use them. Curr Opin Chem Biol. 2017;37:89-96.

26. Mazurenko S, Prokop Z, Damborsky J. Machine learning in enzyme engineering. ACS Catal. 2019;10(2):1210-23.

27. Risso VA, Martinez-Rodriguez S, Candel AM, Kruger DM, Pantoja-Uceda D, Ortega-Munoz M, et al. De novo active sites for resurrected precambrian enzymes. Nat Commun. 2017;8:16113.

28. Alonso S, Santiago G, Cea-Rama I, Fernandez-Lopez L, Coscolín C, Modregger J, et al. Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. Nat Catal. 2020;3(3):319-28.

29. Rasila TS, Pajunen MI, Savilahti H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. Anal Biochem. 2009;388(1):71-80.

30. Xiao H, Bao Z, Zhao H. High throughput screening and selection methods for directed enzyme evolution. Ind Eng Chem Res. 2015;54(16):4011-20.

31. Toth-Petroczy A, Tawfik DS. The robustness and innovability of protein folds. Curr Opin Struct Biol. 2014;26:131-8.

32. Bunzel HA, Garrabou X, Pott M, Hilvert D. Speeding up enzyme discovery and engineering with ultrahigh-throughput methods. Curr Opin Struct Biol. 2018;48:149-56.

33. Leemhuis H, Kelly RM, Dijkhuizen L. Directed evolution of enzymes: Library screening strategies. IUBMB Life. 2009;61(3):222-8.

34. Aita T, Uchiyama H, Inaoka T, Nakajima M, Kokubo T, Husimi Y. Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin. Biopolymers. 2000;54(1):64-79.

35. Kaltenbach M, Tokuriki N. Dynamics and constraints of enzyme evolution. J Exp Zool B Mol Dev Evol. 2014;322(7):468-87.

36. Reetz MT. The importance of additive and non-additive mutational effects in protein engineering. Angew Chem Int Ed Engl. 2013;52(10):2658-66.

37. Miton CM, Tokuriki N. How mutational epistasis impairs predictability in protein evolution and design: How Epistasis Impairs Predictability in Enzyme Evolution. Protein Science. 2016;25(7):1260-72.

38. Goldsmith M, Tawfik DS. Enzyme engineering: reaching the maximal catalytic efficiency peak. Curr Opin Struct Biol. 2017;47:140-50.

39. Starr TN, Thornton JW. Epistasis in protein evolution. Protein Science. 2016;25(7):1204-18.

40. Olson CA, Wu Nicholas C, Sun R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Current Biology. 2014;24(22):2643-51.

41. Otwinowski J, McCandlish DM, Plotkin JB. Inferring the shape of global epistasis. Proc Natl Acad Sci U S A. 2018;115(32):E7550-E8.

42. Steinberg B, Ostermeier M. Shifting fitness and epistatic landscapes reflect trade-offs along an evolutionary pathway. Journal of Molecular Biology. 2016;428(13):2730-43.

43. Weinreich DM, Lan Y, Jaffe J, Heckendorn RB. The influence of higher-order epistasis on biological fitness landscape topography. J Stat Phys. 2018;172(1):208-25.

44. Sailer ZR, Harms MJ. High-order epistasis shapes evolutionary trajectories. PLoS Comput Biol. 2017;13(5):e1005541.

45. Tamer YT, Gaszek IK, Abdizadeh H, Batur TA, Reynolds KA, Atilgan AR, et al. High-Order epistasis in catalytic power of dihydrofolate reductase gives rise to a rugged fitness landscape in the presence of trimethoprim selection. Mol Biol Evol. 2019;36(7):1533-50.

46. Yang G, Anderson DW, Baier F, Dohmen E, Hong N, Carr PD, et al. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. Nat Chem Biol. 2019;15(11):1120-8.

47. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature. 2007;445(7126):383-6.

48. Weinreich DM. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 2006;312(5770):111-4.

49. Salverda MLM, Dellus E, Gorter FA, Debets AJM, Oost Jvd, Hoekstra RF, et al. Initial mutations direct alternative pathways of protein evolution. PLOS Genetics. 2011;7(3):e1001321.

50. Acevedo-Rocha CG, Hoebenreich S, Reetz MT. Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. Methods Mol Biol. 2014;1179:103-28.

51. Araya CL, Fowler DM. Deep mutational scanning: assessing protein function on a massive scale. Trends Biotechnol. 2011;29(9):435-42.

52. Mehlhoff JD, Ostermeier M. Biological fitness landscapes by deep mutational scanning. Methods Enzymol. 2020;643:203-24.

53. Miton CM, Chen JZ, Ost K, Anderson DW, Tokuriki N. Statistical analysis of mutational epistasis to reveal intramolecular interaction networks in proteins. Methods Enzymol. 2020;643:243-80.

54. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, et al. The role of protein dynamics in the evolution of new enzyme function. Nature Chemical Biology. 2016;12(11):944-50.

55. Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. Nat Commun. 2012;3:1257.

56. Poelwijk FJ, Socolich M, Ranganathan R. Learning the pattern of epistasis linking genotype and phenotype in a protein. Nat Commun. 2019;10(1):4213.

57. Petrovic D, Risso VA, Kamerlin SCL, Sanchez-Ruiz JM. Conformational dynamics and enzyme evolution. J R Soc Interface. 2018;15(144).

58. Johansson KE, Lindorff-Larsen K. Structural heterogeneity and dynamics in protein evolution and design. Curr Opin Struct Biol. 2018;48:157-63.

59. Campitelli P, Modi T, Kumar S, Ozkan SB. The role of conformational dynamics and allostery in modulating protein evolution. Annu Rev Biophys. 2020;49:267-88.

60. Crean RM, Gardner JM, Kamerlin SCL. Harnessing conformational plasticity to generate designer enzymes. J Am Chem Soc. 2020;142(26):11324-42.

61. Faber MS, Wrenbeck EE, Azouz LR, Steiner PJ, Whitehead TA. Impact of in vivo protein folding probability on local fitness landscapes. Mol Biol Evol. 2019;36(12):2764-77.

62. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. Current Opinion in Structural Biology. 2009;19(5):596-604.

63. Knies J, Cai F, Weinreich DM. Enzyme efficiency but not thermostability drives cefotaxime resistance evolution in TEM-1 β-lactamase. Molecular Biology and Evolution. 2017:msx053.

64. Pucci F, Bernaerts KV, Kwasigroch JM, Rooman M. Quantification of biases in predictions of protein stability changes upon mutations. Bioinformatics. 2018;34(21):3659-65.

65. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007;2(9):2212-21.

66. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. The stability effects of protein mutations appear to be universally distributed. J Mol Biol. 2007;369(5):1318-32.

67. Sanchez IE, Tejero J, Gomez-Moreno C, Medina M, Serrano L. Point mutations in protein globular domains: contributions from function, stability and misfolding. J Mol Biol. 2006;363(2):422-32.

68. Tokuriki N, Stricher F, Serrano L, Tawfik DS. How protein stability and new functions trade off. PLOS Computational Biology. 2008;4(2):e1000002.

69. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. Proc Natl Acad Sci U S A. 2006;103(15):5869-74.

70. Wang X, Minasov G, Shoichet BK. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. Journal of Molecular Biology. 2002;320(1):85-95.

71. Takano K, Aoi A, Koga Y, Kanaya S. Evolvability of thermophilic proteins from archaea and bacteria. Biochemistry. 2013;52(28):4774-80.

72. Finch AJ, Kim JR. Thermophilic proteins as versatile scaffolds for protein engineering. Microorganisms. 2018;6(4).

73. Thomas A, Cutlan R, Finnigan W, van der Giezen M, Harmer N. Highly thermostable carboxylic acid reductases generated by ancestral sequence reconstruction. Commun Biol. 2019;2:429.

74. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. J Am Chem Soc. 2013;135(8):2899-902.

75. Dasmeh P, Serohijos AWR. Estimating the contribution of folding stability to nonspecific epistasis in protein evolution. Proteins. 2018;86(12):1242-50.

76. Shah P, McCandlish DM, Plotkin JB. Contingency and entrenchment in protein evolution under purifying selection. Proc Natl Acad Sci U S A. 2015;112(25):E3226-35.

77. Yu H, Dalby PA. Coupled molecular dynamics mediate long- and short-range epistasis between mutations that affect stability and aggregation kinetics. Proc Natl Acad Sci U S A. 2018;115(47):E11043-E52.

78. Dellus-Gur E, Elias M, Caselli E, Prati F, Salverda MLM, de Visser JAGM, et al. Negative epistasis and evolvability in TEM-1 β-lactamase—The thin line between an enzyme's conformational freedom and disorder. Journal of Molecular Biology. 2015;427(14):2396-409.

79. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature. 2007;450(7172):964-72.

80. Gobeil SMC, Ebert MCCJC, Park J, Gagné D, Doucet N, Berghuis AM, et al. The structural dynamics of engineered β-lactamases vary broadly on three timescales yet sustain native function. Sci Rep. 2019;9(1):6656.

81. Pandya MJ, Schiffers S, Hounslow AM, Baxter NJ, Williamson MP. Why the energy landscape of barnase Is hierarchical. Front Mol Biosci. 2018;5:115.

82. Duff MR, Jr., Borreguero JM, Cuneo MJ, Ramanathan A, He J, Kamath G, et al. Modulating enzyme activity by altering protein dynamics with solvent. Biochemistry. 2018;57(29):4263-75.

83. Liu Y, Bahar I. Sequence evolution correlates with structural dynamics. Mol Biol Evol. 2012;29(9):2253-63.

84. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature. 2007;450(7171):913-6.

85. Hensen U, Meyer T, Haas J, Rex R, Vriend G, Grubmuller H. Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. PLoS One. 2012;7(5):e33931.

86. Gardner JM, Biler M, Risso VA, Sanchez-Ruiz JM, Kamerlin SCL. Manipulating conformational dynamics to repurpose ancient proteins for modern catalytic functions. ACS Catal. 2020;10(9):4863-70.

87. Pabis A, Risso VA, Sanchez-Ruiz JM, Kamerlin SC. Cooperativity and flexibility in enzyme evolution. Curr Opin Struct Biol. 2018;48:83-92.

88. Campbell EC, Correy GJ, Mabbitt PD, Buckle AM, Tokuriki N, Jackson CJ. Laboratory evolution of protein conformational dynamics. Current Opinion in Structural Biology. 2018;50:49-57.

89. Henderson R, Edwards RJ, Mansouri K, Janowska K, Stalls V, Gobeil SMC, et al. Controlling the SARS-CoV-2 spike glycoprotein conformation. Nature Structural & Molecular Biology. 2020;27(10):925-33.

90. Lu M, Uchil PD, Li W, Zheng D, Terry DS, Gorman J, et al. Real-Time conformational dynamics of SARS-CoV-2 spikes on virus particles. Cell Host & Microbe. 2020;28(6):880-91.e8.

91. Boehr DD, Dyson HJ, Wright PE. An NMR perspective on enzyme dynamics. Chem Rev. 2006;106(8):3055-79.

92. Ortega G, Pons M, Millet O. Chapter Six - Protein functional dynamics in multiple timescales as studied by NMR spectroscopy. In: Karabencheva-Christova T, editor. Advances in Protein Chemistry and Structural Biology. 92: Academic Press; 2013. p. 219-51.

93. Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. Mol Biol Evol. 2015;32(1):132-43.

94. Skopalík J, Anzenbacher P, Otyepka M. Flexibility of human cytochromes P450: Molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. J Phys Chem B. 2008;112(27):8165-73.

95. Meini MR, Tomatis PE, Weinreich DM, Vila AJ. Quantitative description of a protein fitness landscape based on molecular features. Mol Biol Evol. 2015;32(7):1774-87.

96. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science. 2009;324(5924):203-7.

97. Gonzalez MM, Abriata LA, Tomatis PE, Vila AJ. Optimization of conformational dynamics in an epistatic evolutionary trajectory. Mol Biol Evol. 2016;33(7):1768-76.

98. Otten R, Liu L, Kenner LR, Clarkson MW, Mavor D, Tawfik DS, et al. Rescue of conformational dynamics in enzyme catalysis by directed evolution. Nat Commun. 2018;9(1):1314.

99. Maria-Solano MA, Serrano-Hervás E, Romero-Rivera A, Iglesias-Fernández J, Osuna S. Role of conformational dynamics in the evolution of novel enzyme function. Chemical Communications. 2018;54(50):6622-34.

100. Bonomi M, Heller GT, Camilloni C, Vendruscolo M. Principles of protein structural ensemble determination. Curr Opin Struct Biol. 2017;42:106-16.

101. Ward AB, Sali A, Wilson IA. Biochemistry. Integrative structural biology. Science. 2013;339(6122):913-5.

102. Davey JA, Damry AM, Goto NK, Chica RA. Rational design of proteins that exchange on functional timescales. Nature Chemical Biology. 2017;13(12):1280-5.

103. Zoi I, Antoniou D, Schwartz SD. Incorporating fast protein dynamics into enzyme design: A proposed mutant aromatic amine dehydrogenase. J Phys Chem B. 2017;121(30):7290-8.

104. Firnberg E, Labonte JW, Gray JJ, Ostermeier M. A comprehensive, high-resolution map of a gene's fitness landscape. Molecular Biology and Evolution. 2014;31(6):1581-92.

105. Bershtein S, Tawfik DS. Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. Molecular Biology and Evolution. 2008;25(11):2311-8.

106. Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, et al. Capturing the mutational landscape of the beta-lactamase TEM-1. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(32):13067-72.

107. Barlow M, Hall BG. Predicting evolutionary potential: In vitro evolution accurately reproduces natural evolution of the TEM β-lactamase. Genetics. 2002;160(3):823-32.

108. Meyer MM, Hochrein L, Arnold FH. Structure-guided SCHEMA recombination of distantly related beta-lactamases. Protein Eng Des Sel. 2006;19(12):563-70.

109. Orencia MC, Yoon JS, Ness JE, Stemmer WPC, Stevens RC. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nature Structural Biology. 2001;8(3):238-42.

110. Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. J Mol Biol. 2013;425(14):2609-21.

111. Bush K. Past and present perspectives on β-Lactamases. Antimicrobial Agents and Chemotherapy. 2018;62(10).

112. Suay-Garcia B, Perez-Gracia MT. Present and future of carbapenem-resistant enterobacteriaceae (CRE) infections. Antibiotics (Basel). 2019;8(3).

113. D'Angelo RG, Johnson JK, Bork JT, Heil EL. Treatment options for extended-spectrum beta-lactamase (ESBL) and AmpC-producing bacteria. Expert Opin Pharmacother. 2016;17(7):953-67.

114. Bonomo RA. β-Lactamases: A focus on current challenges. Cold Spring Harb Perspect Med. 2017;7(1).

115. Ambler RP. The structure of beta-lactamases. Philos Trans R Soc Lond B Biol Sci. 1980;289(1036):321-31.

116. Bush K. Proliferation and significance of clinically relevant beta-lactamases. Ann N Y Acad Sci. 2013;1277:84-90.

117. Gobeil SMC, Clouthier CM, Park J, Gagne D, Berghuis AM, Doucet N, et al. Maintenance of native-like protein dynamics may not be required for engineering functional proteins. Chem Biol. 2014;21(10):1330-40.

118. Majiduddin FK, Materon IC, Palzkill TG. Molecular analysis of beta-lactamase structure and function. Int J Med Microbiol. 2002;292(2):127-37.

119. Drawz SM, Bonomo RA. Three Decades of β-Lactamase Inhibitors. Clinical Microbiology Reviews. 2010;23(1):160-201.

120. Clouthier CM, Morin S, Gobeil SMC, Doucet N, Blanchet J, Nguyen E, et al. Chimeric β-Lactamases: Global conservation of parental function and fast time-scale dynamics with increased slow motions. PLoS One. 2012;7(12):e52283.

121. Herzberg O, Moult J. Bacterial resistance to beta-lactam antibiotics: crystal structure of beta-lactamase from Staphylococcus aureus PC1 at 2.5 A resolution. Science. 1987;236(4802):694-701.

122. Strynadka NC, Adachi H, Jensen SE, Johns K, Sielecki A, Betzel C, et al. Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 A resolution. Nature. 1992;359(6397):700-5.

123. Escobar WA, Tan AK, Fink AL. Site-directed mutagenesis of beta-lactamase leading to accumulation of a catalytic intermediate. Biochemistry. 1991;30(44):10783-7.

124. Delaire M, Lenfant F, Labia R, Masson JM. Site-directed mutagenesis on TEM-1 beta-lactamase: role of Glu166 in catalysis and substrate binding. Protein Eng. 1991;4(7):805-10.

125. Adachi H, Ohta T, Matsuzawa H. Site-directed mutants, at position 166, of RTEM-1 beta-lactamase that form a stable acyl-enzyme intermediate with penicillin. J Biol Chem. 1991;266(5):3186-91.

126. Palzkill T. Structural and mechanistic basis for extended-spectrum drug-resistance ,utations in altering the specificity of TEM, CTX-M, and KPC β-lactamases. Front Mol Biosci. 2018;5.

127. Atanasov BP, Mustafi D, Makinen MW. Protonation of the beta-lactam nitrogen is the trigger event in the catalytic action of class A beta-lactamases. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(7):3160-5.

128. Minasov G, Wang X, Shoichet BK. An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. J Am Chem Soc. 2002;124(19):5333-40.

129. Meroueh SO, Fisher JF, Schlegel HB, Mobashery S. Ab initio QM/MM study of class A beta-lactamase acylation: dual participation of Glu166 and Lys73 in a concerted base promotion of Ser70. J Am Chem Soc. 2005;127(44):15397-407.

130. Hermann JC, Hensen C, Ridder L, Mulholland AJ, Holtje HD. Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A beta-lactamase with benzylpenicillin. J Am Chem Soc. 2005;127(12):4454-65.

131. Doucet N, Savard PY, Pelletier JN, Gagne SM. NMR Investigation of Tyr105 mutants in TEM-1 beta-lactamase: Dynamics are correlated with function. Journal of Biological Chemistry. 2007;282(29):21448-59.

132. Doucet N. Site-saturation mutagenesis of Tyr-105 reveals its importance in substrate stabilization and discrimination in TEM-1 -lactamase. Journal of Biological Chemistry. 2004;279(44):46295-303.

133. Dubus A, Wilkin JM, Raquet X, Normark S, Frere JM. Catalytic mechanism of active-site serine beta-lactamases: role of the conserved hydroxy group of the Lys-Thr(Ser)-Gly triad. Biochem J. 1994;301 (Pt 2):485-94.

134. Imtiaz U, Manavathu EK, Lerner SA, Mobashery S. Critical hydrogen bonding by serine 235 for cephalosporinase activity of TEM-1 beta-lactamase. Antimicrob Agents Chemother. 1993;37(11):2438-42.

135. Doucet N. Semi-random mutagenesis and dynamic analysis of TEM-1 β-lactamase from Escherichia coli. [PhD thesis]: Universite de Montreal; 2006.

136. Salverda MLM, De Visser JAGM, Barlow M. Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. FEMS Microbiology Reviews. 2010;34(6):1015-36.

137. Bowman GR, Bolin ER, Hart KM, Maguire BC, Marqusee S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(9):2734-9.

138. Modi T, Ozkan SB. Mutations utilize dynamic allostery to confer resistance in TEM-1 β-lactamase. Int J Mol Sci. 2018;19(12):3808.

139. Kather I, Jakob RP, Dobbek H, Schmid FX. Increased folding stability of TEM-1 beta-lactamase by in vitro selection. J Mol Biol. 2008;383(1):238-51.

140. Hart KM, Ho CMW, Dutta S, Gross ML, Bowman GR. Modelling proteins' hidden conformations to predict antibiotic resistance. Nat Commun. 2016;7:12965.

141. Knox R, Lento C, Wilson DJ. Mapping conformational dynamics to individual steps in the TEM-1 β-lactamase catalytic mechanism. Journal of Molecular Biology. 2018;430(18):3311-22.

142. Kapoor M, Gupta MN. Lipase promiscuity and its biochemical applications. Process Biochemistry. 2012;47(4):555-69.

143. Houde A, Kademi A, Leblanc D. Lipases and their industrial applications. Applied Biochemistry and Biotechnology. 2004;118(1):155-70.

144. Monteiro RRC, Virgen-Ortiz JJ, Berenguer-Murcia Á, Da Rocha TN, Dos Santos JCS, Alcántara AR, et al. Biotechnological relevance of the lipase A from Candida antarctica. Catalysis Today. 2020.

145. Widmann M, Juhl PB, Pleiss J. Structural classification by the lipase engineering database: a case study of Candida antarctica lipase A. BMC Genomics. 2010;11:123.

146. Kirk O, Christensen MW. Lipases from Candida antarctica: Unique biocatalysts from a unique origin. Organic Process Research & Development. 2002;6(4):446-51.

147. Brundiek H, Padhi SK, Kourist R, Evitt A, Bornscheuer UT. Altering the scissile fatty acid binding site of Candida antarctica lipase A by protein engineering for the selective hydrolysis of medium chain fatty acids. European Journal of Lipid Science and Technology. 2012;114(10):1148-53.

148. Pleiss J, Fischer M, Peiker M, Thiele C, Schmid RD. Lipase engineering database: Understanding and exploiting sequence–structure–function relationships. Journal of Molecular Catalysis B: Enzymatic. 2000;10(5):491-508.

149. Ericsson DJ, Kasrayan A, Johansson P, Bergfors T, Sandström AG, Bäckvall J-E, et al. X-ray Structure of Candida antarctica Lipase A Shows a Novel Lid Structure and a Likely Mode of Interfacial Activation. Journal of Molecular Biology. 2008;376(1):109-19.

150. Verger R. 'Interfacial activation' of lipases: facts and artifacts. Trends in Biotechnology. 1997;15(1):32-8.

151. Khan FI, Lan D, Durrani R, Huan W, Zhao Z, Wang Y. The lid domain in lipases: Structural and functional determinant of enzymatic properties. Front Bioeng Biotechnol. 2017;5:16.

152. Wikmark Y, Engelmark Cassimjee K, Lihammar R, Backvall JE. Removing the active-site flap in lipase A from Candida antarctica produces a functional enzyme without interfacial activation. Chembiochem. 2016;17(2):141-5.

153. Kasrayan A, Bocola M, Sandstrom AG, Laven G, Backvall JE. Prediction of the Candida antarctica lipase A protein structure by comparative modeling and site-directed mutagenesis. Chembiochem. 2007;8(12):1409-15.

154. Rotticci D, Norin T, Hult K, Martinelle M. An active-site titration method for lipases. Biochim Biophys Acta. 2000;1483(1):132-40.

155. Robles-Machuca M, Del Campo MM, Camacho-Ruiz MA, Ordaz E, Zamora-Gonzalez EO, Muller-Santos M, et al. Comparative features between recombinant lipases CALA-like from U. maydis and CALA from C. antarctica in thermal stability and selectivity. Biotechnol Lett. 2019;41(2):241-52.

156. Liljeblad A, Kallio P, Vainio M, Niemi J, Kanerva LT. Formation and hydrolysis of amide bonds by lipase A from Candida antarctica; exceptional features. Org Biomol Chem. 2010;8(4):886-95.

157. Gedey S, Liljeblad A, Lázár L, Fülöp F, Kanerva LT. Preparation of highly enantiopure β-amino esters by Candida antarctica lipase A. Tetrahedron: Asymmetry. 2001;12(1):105-10.

158. Martínez-Martínez M, Coscolín C, Santiago G, Chow J, Stogios PJ, Bargiela R, et al. Determinants and prediction of esterase substrate promiscuity patterns. ACS Chem Biol. 2018;13(1):225-34.

159. Zorn K, Oroz-Guinea I, Brundiek H, Dorr M, Bornscheuer UT. Alteration of chain length selectivity of Candida antarctica lipase A by semi-rational design for the enrichment of erucic and gondoic fatty acids. Adv Synth Catal. 2018;360(21):4115-31.

160. Brundiek HB, Evitt AS, Kourist R, Bornscheuer UT. Creation of a lipase highly selective for trans fatty acids by protein engineering. Angew Chem Int Ed Engl. 2012;51(2):412-4.

161. Domínguez De María P, Carboni-Oerlemans C, Tuin B, Bargeman G, Van Der Meer A, Van Gemert R. Biotechnological applications of Candida antarctica lipase A: State-of-the-art. Journal of Molecular Catalysis B: Enzymatic. 2005;37(1-6):36-46.

162. Nyyssölä A, Miettinen H, Kontkanen H, Lille M, Partanen R, Rokka S, et al. Treatment of milk fat with sn-2 specific Pseudozyma antarctica lipase A for targeted hydrolysis of saturated medium and long-chain fatty acids. International Dairy Journal. 2015;41:16-22.

163. Glod D. Modification of fatty acid selectivity of Candida antarctica lipase A by error-prone PCR. Biotechnol Lett. 2017;39(5):767-73.

164. He Y, Li J, Kodali S, Chen B, Guo Z. Rationale behind the near-ideal catalysis of Candida antarctica lipase A (CAL-A) for highly concentrating omega-3 polyunsaturated fatty acids into monoacylglycerols. Food Chem. 2017;219:230-9.

165. Svendsen A, Vind J, Pathar S, Borch K, inventors; Novozymes, assignee. Lipolytic enzyme variants. patent WO/2008/040738. 2008.

166. Svendsen A, Pathar SA, Egel-Mitani M, Borch K, Clausen IG, Hansen MT, inventors; Novo Nordisk, assignee. C. Antarctica lipase and lipase variants. patent WO9401541 1993.

167. Bornsheuer U, Brundiek H, Evitt A, Saß S, Bönisch F, Kourist R, inventors; Frito-Lay trading company, assignee. Lipase variants having increased enzyme specificity or enhanced trans-selectivity and methods of use. patent WO/2012/146935. 2012.

168. Panizza P, Cesarini S, Diaz P, Rodriguez Giordano S. Saturation mutagenesis in selected amino acids to shift Pseudomonas sp. acidic lipase Lip I.3 substrate specificity and activity. Chem Commun (Camb). 2015;51(7):1330-3.

169. Yang J, Koga Y, Nakano H, Yamane T. Modifying the chain-length selectivity of the lipase from Burkholderia cepacia KWI-56 through in vitro combinatorial mutagenesis in the substrate-binding site. Protein Eng Des Sel. 2002;15(2):147-52.

170. Gaskin DJ, Romojaro A, Turner NA, Jenkins J, Vulfson EN. Alteration of lipase chain length specificity in the hydrolysis of esters by random mutagenesis. Biotechnol Bioeng. 2001;73(6):433-41.

171. Syal P, Verma VV, Gupta R. Targeted mutations and MD simulations of a methanol-stable lipase YLIP9 from Yarrowia lipolytica MSR80 to develop a biodiesel enzyme. Int J Biol Macromol. 2017;104(Pt A):78-88.

172. Santarossa G, Lafranconi PG, Alquati C, DeGioia L, Alberghina L, Fantucci P, et al. Mutations in the "lid" region affect chain length specificity and thermostability of a Pseudomonas fragi lipase. FEBS Lett. 2005;579(11):2383-6.

173. Brocca S, Secundo F, Ossola M, Alberghina L, Carrea G, Lotti M. Sequence of the lid affects activity and specificity of Candida rugosa lipase isoenzymes. Protein Sci. 2003;12(10):2312-9.

174. Yu XW, Zhu SS, Xiao R, Xu Y. Conversion of a Rhizopus chinensis lipase into an esterase by lid swapping. J Lipid Res. 2014;55(6):1044-51.

175. Stoytcheva M, Montero G, Zlatev R, Leon JA, Gochev V. Analytical methods for lipases activity determination: A review. Curr Anal Chem. 2012;8(3):400-7.

176. Reyes-Duarte D, Coscolin C, Martinez-Martinez M, Ferrer M, Garcia-Arellano H. Functional-based screening methods for detecting esterase and lipase activity against multiple substrates. Methods Mol Biol. 2018;1835:109-17.

177. Vogt AD, Pozzi N, Chen Z, Di Cera E. Essential role of conformational selection in ligand binding. Biophys Chem. 2014;186:13-21.

178. Michel D. Conformational selection or induced fit? New insights from old principles. Biochimie. 2016;128-129:48-54.

179. Morando MA, Saladino G, D'Amelio N, Pucheta-Martinez E, Lovera S, Lelli M, et al. Conformational selection and induced fit mechanisms in the binding of an anticancer drug to the c-Src kinase. Sci Rep. 2016;6:24439.

180. Okazaki K, Takada S. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. Proc Natl Acad Sci U S A. 2008;105(32):11182-7.

181. Pimenta AC, Martins JM, Fernandes R, Moreira IS. Ligand-induced structural changes in TEM-1 probed by molecular dynamics and relative binding free energy calculations. J Chem Inf Model. 2013;53(10):2648-58.

182. Desnuelle P, Savary P. Specificities of Lipases. J Lipid Res. 1963;4:369-84.

183. Klein RR, King G, Moreau RA, Haas MJ. Altered acyl chain length specificity of Rhizopus delemar lipase through mutagenesis and molecular modeling. Lipids. 1997;32(2):123-30.

184. Ben Ayed R, Ennouri K, Ercişli S, Ben Hlima H, Hanana M, Smaoui S, et al. First study of correlation between oleic acid content and SAD gene polymorphism in olive oil samples through statistical and bayesian modeling analyses. Lipids Health Dis. 2018;17(1).

185. de Carvalho CC. Whole cell biocatalysts: essential workers from Nature to the industry. Microb Biotechnol. 2017;10(2):250-63.

186. Shin H, Cho BK. Rational protein engineering guided by deep mutational scanning. Int J Mol Sci. 2015;16(9):23094-110.

187. Singh A. Deep learning 3D structures. Nature Methods. 2020;17(3):249-.

188. Lindorff-Larsen K, Maragakis P, Piana S, Shaw DE. Picosecond to millisecond structural dynamics in human ubiquitin. J Phys Chem B. 2016;120(33):8313-20.

189. Duan L, Guo X, Cong Y, Feng G, Li Y, Zhang JZH. Accelerated molecular dynamics simulation for helical proteins folding in explicit water. Frontiers in Chemistry. 2019;7.

190. Martinez-Rosell G, Giorgino T, Harvey MJ, de Fabritiis G. Drug discovery and molecular dynamics: Methods, applications and perspective beyond the second timescale. Curr Top Med Chem. 2017;17(23):2617-25.

ANNEX 1 – Supplementary Material to Chapter 3

Table A1.1. – Salt bridges identified by salt bridge plugin in VMD

TEM-1	TEM-1 E104K/G238S	cTEM-2m	cTEM-2m E104K/G238S	cTEM-17m	cTEM-17m E104K/G238S
D163-R161	D163-R161	D115-HIS112	D115-HIS112	D115-R94	D115-R94
D176-R178	D176-R178	D163-R161	D163-R161	D115-HIS112	D115-HIS112
D179-R164	D179-R164	D176-R178	D176-R178	D157-R153	D157-R153
D214-K234	D214-K234	D179-R164	D179-R164	D163-R161	D163-R161
D233-R222	D233-R222	D214-K234	D214-K234	D176-R178	D176-R178
D254-K256	D273-R277	D233-R222	D233-R222	D179-R164	D179-R164
D273-R277	D34-K31	D273-R277	D273-R277	D214-R222	D214-R222
D34-K31	D37-K33	D35-K32	D35-K32	D233-R222	D233-R222
D37-K33	D49-K54	E121-R94	E166-K73	D233-K215	D233-K215
D49-K54	E121-R94	E166-K73	E171-R164	D254-K256	D254-K256
E121-R94	E166-K73	E171-R164	E171-R241	D273-R277	D273-R277
E166-K73	E171-R164	E171-R241	E177-R65	D35-K32	D35-K32
E171-R164	E171-R178	E177-R65	E197-R204	E110-K111	E110-K111
E171-R178	E177-R65	E197-R204	E240-R241	E166-K73	E121-HIS112
E177-R65	E197-R204	E240-R241	E240-K104	E171-R164	E166-K73
E197-R204	E240-K104	E274-R277	E274-R277	E240-R241	E171-R164
E281-R277	E281-R277	E281-R277	E281-K32	E37-R61	E240-R241
E281-K31	E281-K31	E281-K32	E37-R61	E37-K34	E240-K104
E36-R61	E36-R61	E37-R61	E37-K34	E48-R259	E37-R61
E47-R259	E47-R259	E37-K34	E48-R259	E64-R43	E37-K34
E64-R42	E64-R42	E48-R259	E64-R43	E64-R61	E48-R259
E64-R61	E64-R61	E64-R43	E64-R61	E89-R83	E64-R43
E89-R83	E89-R83	E64-R61	E89-R83	E89-R93	E64-R61
E89-R93	E89-R93	E89-R83	E89-R93		E89-R83
		E89-R93			E89-R93

Table A1.2. – Mutations identified in clones prior to selection at each generation of directed molecular evolution

	cTEM-2m	cTEM-17m
Generation 1	A9V, K111M, N136S, K146*	WT
	Y46C, P62P, T140I	WT
	R43*, V108L	I127L
	I47L, S124N, T128S	V89I, P175A
	WT	F81I, V89E, L160I
	L57L	R60H, C132G, L145I
	L57L, K215I	F15L, Q97H
	K234L	L57P, E62D, I50F
	WT	N79S, F81L, L111*
	L40S	F81I, H105Y
		A11T
		I163N
		WT
		WT
		WT
		WT
Generation 2	WT	L57L, P62P
	D115G, N170I	A215A
	F17S,	E104D, D163D

	WT	P62P, L162L
	F17S, G116S, T133I	L162L, T182T
	Q24*, A232A, R128H	G54G, A125A, L207S
	C77S, E104V, T118A, E178D, A184A	E64*, K192E
	E28*, T48T, P55L, R66C	E64*, V84L
	D35N, L102M, G116S	D168N, I282I
	E28*. T133P	R275R
	G156D, P167L, R178C, R241H	
	D232D, A270T	
	T18S	
Generation 3	L40F, I56F	A36A, P62P, A135V
	L30P, L40F, C77R	E64*, A125A,
	I7F, L31L, L40F, G45R	S59C, E64*, M117L, R164R
	K2R, L40F, G116S	A15T, E64*, A125A
	T29T, L91V, G116S	E64*, S70S, R164R
	L40F, I95L, G116S, T160T, R178C	A15T, P62P, I95V,
	G54C, G116S, T160T	E64*, A135V, T141R, R164R
	L40F, G116S, T160T, P167L, R178C	P62P, T118T
	L40F, C77*, P107P, K111N, G116S, T160T, P167L, R178C	Q39L, E64*

	K2R, L40F, G116S, T160T, P167L, R178C R81R, S98Y, E104V, T114A, T118A, S124N. T160T, P167L, R178C L12L, G14D, L40F, G116S, T160T, P167L, R178C, S223P	T21S, N52S, E64*, A135V, R164R

*represents stop codons

ANNEX 2 – Supplementary Material to Chapter 4

Table A2.1. – **Activity for short-chain and long-chain discriminative Cal-A variants selected from library Random 2 during screening against triglyceride substrates**

Hydrolytic activity of variants towards the short-chain triglyceride tributyrin and the long-chain substrate olive oil was categorized as very low (1), low (2), medium (3) or high (4). The value (0) indicates no activity detected towards that substrate. Wild-type Cal-A activity value is 3. The color code is identical to Figure 4.3. Twenty-one discriminative variants were identified upon screening library 2: eighteen variants showing short-chain discrimination (Table A2.1A) and three variants showing long-chain discrimination (Table A2.1B).

A. Random 2 library variants that discriminate for short-chain fatty acids

Variant	Activity		Residue		
	Short-chain	Long-chain		WT	Mut
101	1	0	217	S	I
			244	A	T
			315	A	T
106	1	0	236	A	V
			331	A	T
110	1	0	243	L	P
			300	P	S
			327	F	L
111	1	0	236	A	T
			298	E	K
			307	Q	H
			319	V	E
112	2	0	227	G	S
			232	G	C
			265	K	T
114	1	0	221	T	R
			291	N	Y
117	3	0	225	L	P
			221	T	K

118	1	0	234	A	T
			271	G	S
			291	N	K
121	4	0	229	P	L
			237	G	S
122	4	0	229	P	L
			237	G	S
124	3	0	228	G	S
			258	A	S
126	4	3	244	A	V
127	4	0	248	M	L
			294	N	I
130	4	2	223	T	S
			244	A	V
			263	T	A
			285	N	I
			290	V	D
			306	K	N
131	1	0	228	G	V
			231	A	T
133	3	1	235	L	P
135	3	0	237	G	C
			251	F	Y
			275	P	Q
			335	E	D
			342	A	V
136	3	0	218	A	V
			240	G	V

B. Random 2 library variants that discriminate for long-chain fatty acids

Variant	Activity		Residue		
	Short-chain	Long-chain		WT	Mut
116	1	2	238	V	L
			262	R	H
			265	K	Q
			293	T	S
			325	P	T
132	1	2	232	G	C
134	1	2	254	A	E
			324	F	C

Table A2.2. – **Activity for discriminative variants selected from library Random Rec during screening against triglyceride substrates**

Hydrolytic activity of variants towards the short-chain triglyceride tributyrin and the long-chain substrate olive oil was categorized as very low (1), low (2), medium (3) or high (4). The value (0) indicates no activity detected towards that substrate. Wild-type Cal-A activity value is 3. The color code is identical to Figure 4.3 . Eighteen discriminative variants were identified upon screening library Random Rec: twelve variants showing short-chain discrimination and six variants showing long-chain discrimination.

A. Random Rec library variants that discriminate for short-chain fatty acids

Variant	Activity		Residue		
	Short-chain	Long-chain		WT	Mut
22	4	0	302	A	S
			334	D	Y
			342	A	E
23	3	2	250	S	F
			335	E	V
			342	A	V
			412	V	I

24	3	2	312	Q	K
			435	F	L
26	4	0	222	F	I
			283	F	L
			418	A	T
30	4	0	237	G	D
32	3	1	247	D	G
			249	E	G
			305	L	V
33	4	0	222	F	I
34	4	0	30	T	A
			222	F	I
			340	Q	H
			405	T	I
37	3	0	233	F	I
			287	F	L
39	3	2	338	P	Q
42	3	2	377	S	R
45	3	2	338	P	T

B. Random Rec library variants that discriminate for long-chain fatty acids

Variant	**Activity**		**Residue**		
	Short-chain	Long-chain		WT	Mut
21	2	3	84	P	Q
27	2	3	359	S	I
28	3	4	24	T	I
			220	D	N
			341	P	S
			402	A	E
29	2	3	228	G	S
			402	A	V

			407	P	A
31	0	1	65	Q	L
			444*	A	T
44	3	4	255	R	S
			257	N	Y

*Residue 444 is not represented in the 3D structure because it is not resolved in the original 2VEO PDB file (far C-term).

Table A2.3. – **Activity for discriminative variants selected from library Random Tot during screening against triglyceride substrates**

Hydrolytic activity of variants towards the short-chain triglyceride tributyrin and the long-chain substrate olive oil was categorized as very low (1), low (2), medium (3) or high (4). The value (0) indicates no activity detected towards that substrate. Wild-type Cal-A activity value is 3. The color code is identical to Figure 4.3. Fourteen discriminative variants were identified upon screening library Random Tot: ten variants showing short-chain discrimination and four variants showing long-chain discrimination.

A. Random Tot library variants that discriminate for short-chain fatty acids

Variant	**Activity**		**Residue**		
	Short-chain	Long-chain		WT	Mut
3	3	0	240	G	C
5	3	1	306	K	R
			336	I	F
			382	K	N
			433	K	R
6	3	1	181	E	G
			313	A	T
			402	A	V
9	2	0	128	G	D
			136	Y	N
			144	G	S
			255	R	H

10	2	0	232	G	C
			428	Q	L
12	3	0	237	G	C
			289	L	M
			333	P	S
			369	A	T
13	3	2	253	E	G
			268	R	S
14	2	0	290	V	D
			424	L	M
17	3	1	29	G	V
			41	K	M
			211	H	Q
			251	F	Y
			284	L	S
18	3	0	190	T	I
			226	N	S
			406	T	S
			410	D	E

B. Random Tot library variants that discriminate for long-chain fatty acids

Variant	Activity		Residue		
	Short-chain	Long-chain		WT	Mut
4	0	1	22	Y	C
			27	N	S
			307	Q	R
			344	T	A
			371	I	T
11	2	3	136	Y	F
			289	L	V
			357	N	D
15	3	4	262	R	H
			274	L	W
			432	G	D
20	2	3	224	F	Y
			339	Y	H
			372	F	L
			396	T	I

Table A2.4. – **Residues that appear in both short-chain and long-chain discriminative variants in Random 2 and Random Rec and Tot libraries**

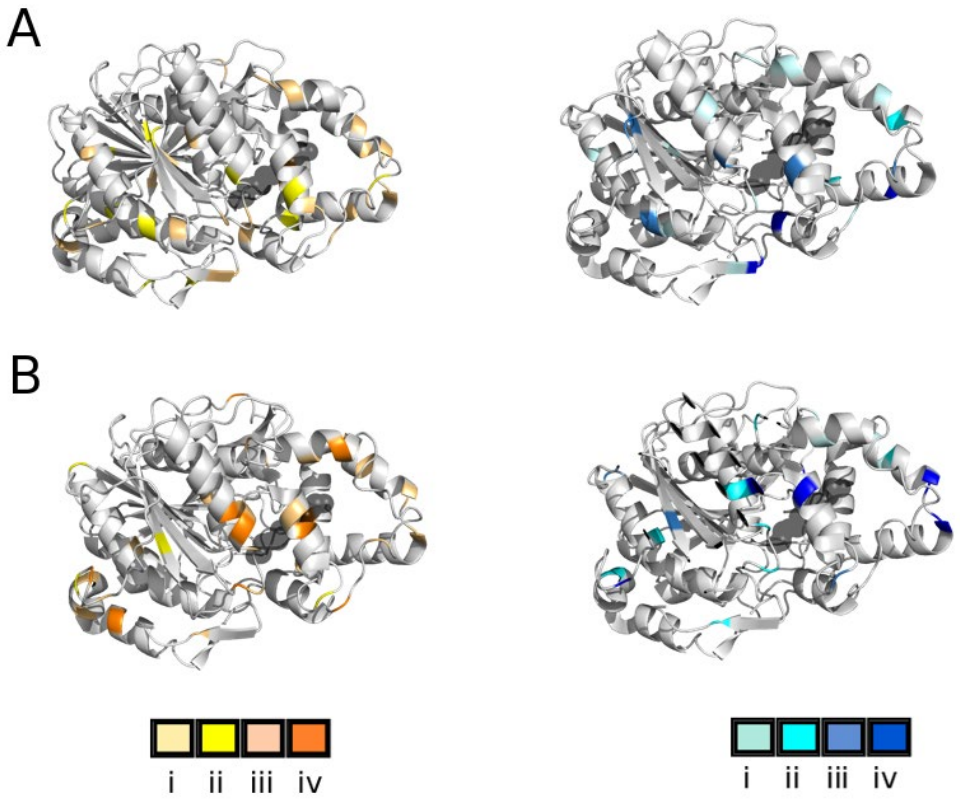
A. Random 2 library:

Residue	WT	Mut	Variant
Short-chain discrimination			
232	G	C	112
265	K	T	112
Long-chain discrimination			
232	G	C	132
265	K	Q	116

B. Random Rec and Random Tot libraries:

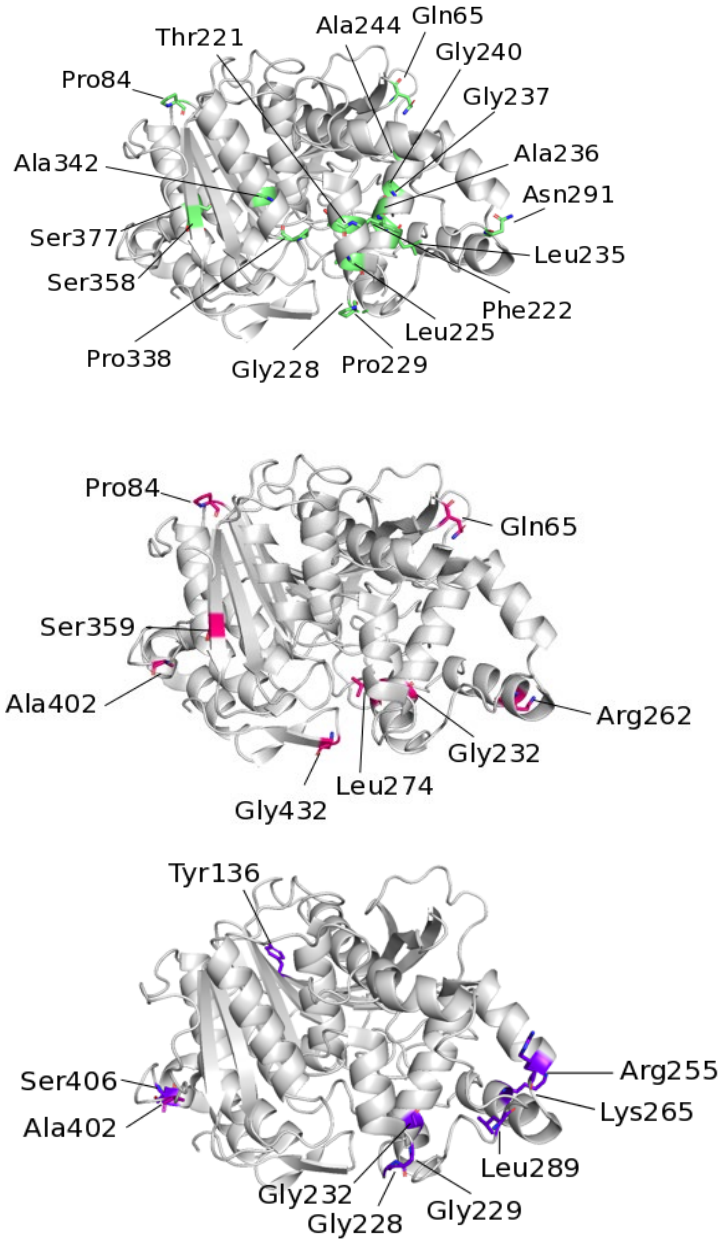
Residue	WT	Mut	Variant
Short-chain discrimination			
136	Y	N	9
289	L	M	12
402	A	V	6
255	R	H	9
Long-chain discrimination			
136	Y	F	11
289	L	V	11
402	A	E	28
402	A	V	29
255	R	S	44

Figure A2.1. – Activity for discriminative variants selected from libraries Random Tot and Random Rec towards short- and long-chain triglycerides



The residues that were substituted in the selected, discriminative variants are colored according to the activity of the variant towards each substrate. A: Discriminative variants selected from library Random Tot. B: Discriminative variants selected from library Random Rec. Left panels, short-chain activity (hydrolysis of tributyrin): gradient from light yellow (low activity) to orange (high activity). Right panels, long-chain activity (hydrolysis of olive oil): gradient from light to dark blue. Wild-type level activity corresponds to shade iii. Where no activity was detected towards a substrate, no color was assigned. Where more than one variant was mutated at the same position, the position is colored according to the variant having the highest activity. A PEG molecule is shown in black spheres, crystallized inside the putative tunnel (1).

Figure A2.2. – Identification of key residues belonging to discriminative variants, classified according to their discriminative nature



Residues highlighted in green were substituted in randomized variants that discriminate towards short-chain fatty acids, in magenta towards long-chain fatty acids and, in purple, residues that were substituted both in variants that discriminate for short-chain or long-chain fatty acids.

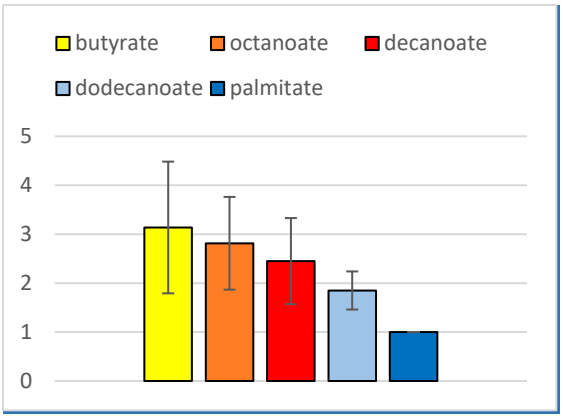


Figure A2.3. – Hydrolytic activity of wild-type Cal-A with *p*-NO_2-phenyl fatty acids

Assays were performed in triplicate with clarified E. coli lysates. Activity is normalized to that of wild-type Cal-A with p-NO_2-phenyl-palmitate (= 1, S.A. = 0.4 U/mg).

References to annex 2

1. Ericsson DJ, Kasrayan A, Johansson P. X-ray structure of *Candida antarctica* lipase A shows a novel lid structure and a likely mode of interfacial activation. *J Mol. Biol*. 2008; 375: 109-19. doi: 10.1016/j.jmb.2007.10.079.

ANNEX 3 – Supplementary Material to Chapter 5

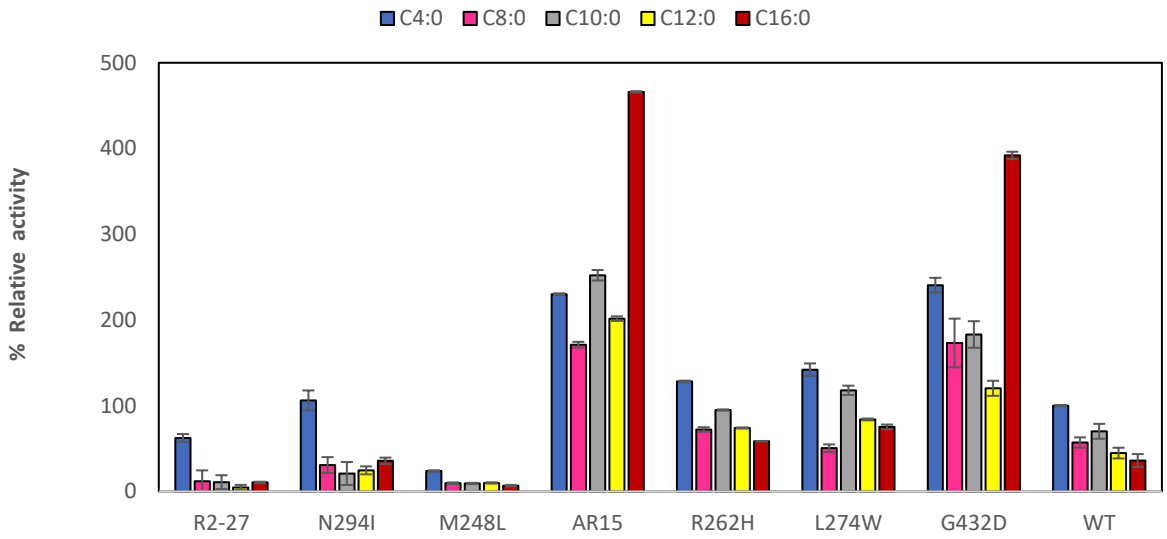


Figure A3.1. – Quantitative *p*-NO_2-phenyl ester assay for clarified lysates of R2-27, AR15 and their deconvoluted mutants

Values are normalized against wild-type activity for p-NO_2-phenyl ester butyrate.

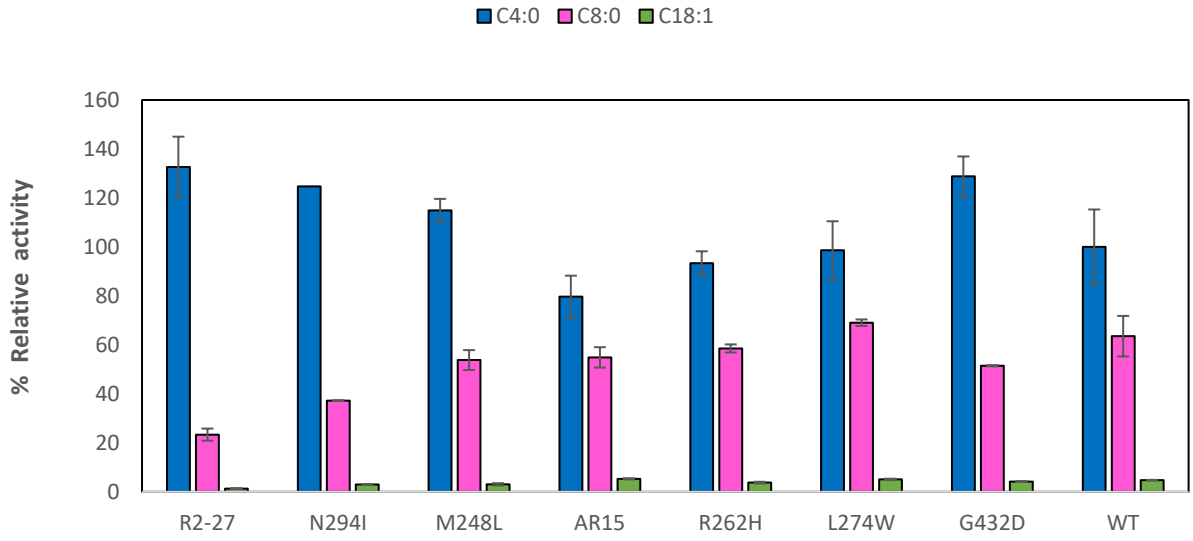


Figure A3.2. – Quantitative triglyceride assay for clarified lysates of R2-27, AR15 and their deconvoluted variants

Values are normalized against wild-type activity for glyceryl tributyrate.

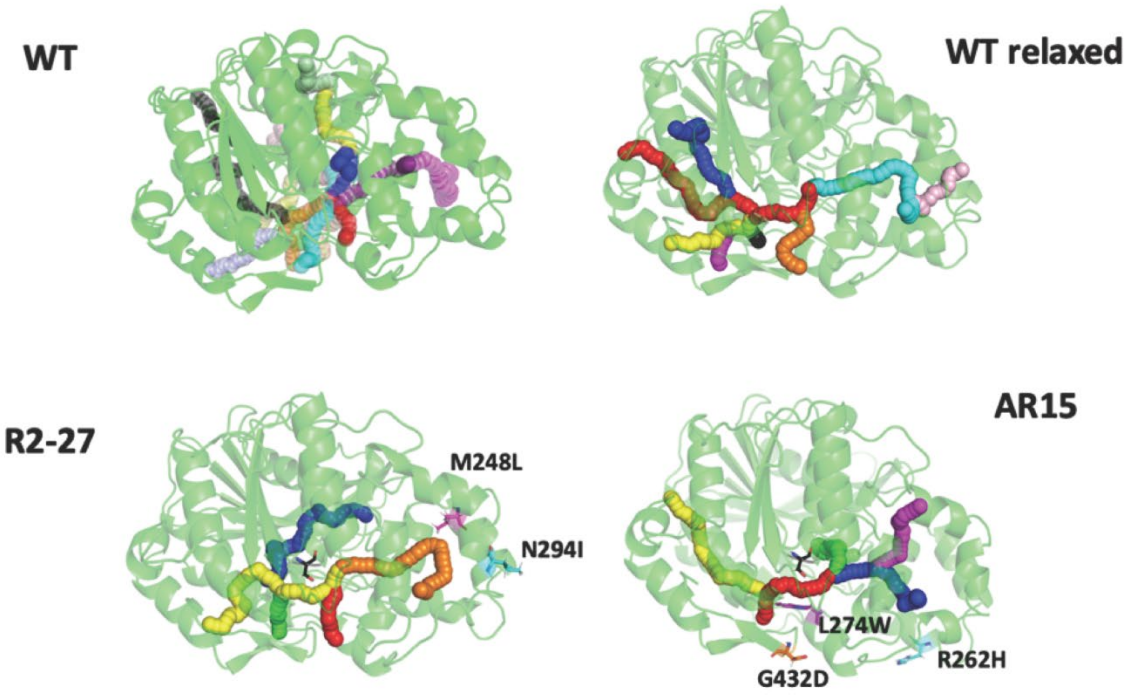


Figure A3.3. – Tunnels observed in WT Cal-A lipase, in its relaxed structure and in discriminative mutants R2-27 and AR15 as determined in simulations using CAVER 3.0

Mutations in variants R2-27 and AR15 are labelled. Each of the identified tunnels using CAVER 3.0 are colored differently.

Table 5.1. – Tunnel results using CAVER 3.0 for Cal-A wild type (PDB ID: 2veo)

The bottleneck radius is defined by CAVER as the narrowest part of a given tunnel. Whereas the cost is a measure of the performance of the algorithm, the lower the better. The throughput is related to the cost e^{-cost}, length is defined as the distance from the calculation starting point to the end of the tunnel along the tunnel axis and the curvature is length/distance where distance is the shortest distance to the end of the tunnel from the starting point.

WT- 2veo

Tunnel cluster	Tunnel	Throughput	Cost	Bottleneck radius	Length	Curvature
1	1	0.39	0.95	0.68	9.77	1.14
2	2	0.26	1.33	0.62	21.89	1.43
3	3	0.23	1.48	0.62	17.89	1.72
4	4	0.22	1.53	0.72	21.26	1.33
5	5	0.22	1.53	0.70	38.39	1.84
6	6	0.19	1.66	0.62	14.45	1.26
7	7	0.13	2.03	0.68	25.22	1.44
8	8	0.11	2.18	0.62	18.39	1.31
9	9	0.07	2.69	0.60	20.00	1.71
10	10	0.07	2.72	0.62	35.07	1.65
11	11	0.05	2.98	0.60	27.81	1.27
12	12	0.04	3.28	0.61	33.64	1.66
13	13	0.01	4.25	0.65	43.88	1.43
14	14	0.00	6.14	0.60	59.11	2.28

Table 5.2. – Tunnel results using CAVER 3.0 for the relaxed structure of Cal-A wild type

The putative acyl-binding tunnel corresponds to the tunnel cluster 3.

WT relaxed

Tunnel cluster	Tunnel	Throughput	Cost	Bottleneck radius	Length	Curvature
1	1	0.32	1.15	0.86	19.74	1.51
2	2	0.15	1.87	0.60	17.15	1.46
3	3	0.14	1.94	0.82	27.32	1.87
4	4	0.14	1.98	0.61	25.72	1.43
5	5	0.07	2.73	0.62	32.52	1.47
6	6	0.05	2.91	0.60	41.08	1.87

7	7	0.05	3.07	0.68	37.02	1.59
8	8	0.03	3.45	0.68	42.93	1.52

Table A3.1. – **Tunnel results using CAVER 3.0 for the relaxed structure of the double mutant R2-27**

The putative acyl-binding tunnel corresponds to the tunnel cluster 3.

R2-27 relaxed

Tunnel cluster	Tunnel	Throughput	Cost	Bottleneck radius	Length	Curvature
1	1	0.28	1.27	0.65	12.13	1.17
2	2	0.15	1.89	0.74	20.10	1.38
3	3	0.09	2.40	0.61	29.49	2.28
4	4	0.04	3.22	0.63	32.43	1.59
5	5	0.01	5.08	0.61	49.35	2.48

Table A3.2. – **Tunnel results using CAVER 3.0 for the relaxed structure of the triple mutant AR15**

The putative acyl-binding tunnel corresponds to the tunnel cluster 4 and the alternative tunnel to cluster 2.

AR15 relaxed

Tunnel cluster	Tunnel	Throughput	Cost	Bottleneck radius	Length	Curvature
1	1	0.16	1.84	0.61	18.63	1.40
2	2	0.16	1.85	0.75	26.54	1.50
3	3	0.14	2.00	0.60	23.44	1.36
4	4	0.09	2.44	0.62	22.15	1.57
5	5	0.02	3.82	0.61	40.11	1.46