The Engineering of Tannase for Industrial Application

By

Deandre Chevannes

A thesis submitted to the Faculty of Science in the Department of Chemistry and Biomolecular Sciences

In conformity with the requirements for

the degree of Master's

University of Ottawa

Ottawa, Ontario, Canada

(April 2024)

Abstract

Enzymes are effective and efficient biocatalysts that when applied can greatly improve production outputs in industrial settings and create more environmentally friendly processes when compared with chemical options. However, the application of enzymes is often hindered due their functionality having evolved naturally against moderate environmental conditions. This generally results in them being relatively unstable and unfit for usage under industrial conditions. This study applied a combination of computational prediction and experimentation with the aim of optimizing the enzyme tannase from *Lactobacillus plantarum* (*Lp*Tan) for industrial usage. Two different methods were explored. The first method described in detail in Chapter 2 of this thesis applied genetic code expansion in the active site of the enzyme to improve catalytic function. In particular, the impact of replacing histidine 451 in the esterase's catalytic triad with its non-canonical counterpart 3-methyl-histidine (NmH) was evaluated. A decreased reliance on the Asp of the catalytic triad, while shortening the distance between the His and the Ser was hypothesized to create a more efficient catalytic dyad. However, when the activity of the derived NmH451 tannase was compared to wild type, it showed a decrease in activity. The second method described in detail in Chapter 3 of this thesis applied combinatorial multi-site-directed mutagenesis toward improving thermal stability and total turnover number of the enzyme. Protein Repair One Stop Shop (PROSS) was used to generate stabilizing mutation predictions for a 'flapless (FLT)' version of *Lp*Tan, including a list of 62 possibilities that were narrowed down to 8 through a set of selection criteria. These mutations included Q63T, A65T, A184Y, A257D, V276Y, T321G, G421D and G439D. Combinatorial screening of the impact of these site changes on functionality has, to date, yielded 4 variants that were characterized for catalytic efficiency (k_{cat}/K_m), melting temperature (T_m), inactivation constant (k_{inact}) and total turnover numbers

(TTN; k_{cat}/k_{inact}). Variant P8E5, with 6 of 8 site variations showed the highest significant T_m value, by an increase of 6.5 ˚C compared to FLT. In conclusion, while the genetic code expansion strategy was not successful, the combinatorial PROSS-predicted site-mutagenesis method did yield a more thermostable variant of the FLT, highlighting the value of computational prediction. Overall, we provide evidence that a tannase enzyme can be stabilized for potential industrial applications.

Acknowledgements

When I first started my master's degree, I was unsure of myself and my abilities. I was placed in a new environment and a new field and had no idea where to begin. As time went on though, I gained new experiences and met so many amazing people that helped me along this fun journey. As the saying goes “Take chances, make mistakes and get messy!” – Miss Frizzle from the Magic School Bus

The first person I would like to thank personally, is my amazing supervisor, Michele Loewen. The invaluable experience I got from working in her lab and her unwavering support in my endeavours. I am truly fortunate to have Michele as a mentor and I am grateful for the countless hours of insight discussions, constructive feedback and encouragement that helped shaped these past two years.

Next, I would like to thank Tony St-Jacques and Kelly Robinson. Tony taught me how to work independently and think for myself, and to always go deeper. This in turn sparked my curiosity and helped me develop my scientific thinking. To Kelly I would like to thank you for training me in many different biological lab techniques which has helped me to expand my knowledge and experience base greatly.

I would also like to thank the rest of the team including Fang, Chengsong, Trent, Tanya, Luana, Yutong, Joanna, Ben, Ricardo, Emma, Josh and Tom. You all were a great group to work with and I’ve really enjoyed our time together.

Finally, I would like to make a special thanks to the multichannel pipette, without which I would probably have taken much, much longer to finish with the bench work aspect of my thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

*Lp*Tan	*Lactobacillus plantarum* Tannase
ABHD	α/β- hydrolase domain
GCE	Genetic Code Expansion
NmH	3-Methyl Histidine
PyIRS	Pyrrolysl tRNA synthetase
ncAA	Non-canonical amino acid
OTS	Orthogonal Translational Systems
aaRS	Aminoacyl-tRNA synthetase
LB	Luria Broth
IPTG	Isopropyl β-d-1-thiogalactopyranoside
PMSF	Phenylmethylsulphonyl fluoride
DTT	Dithiothreitol
Ni-NTA	Nickel Nitriloacetic acid
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
BCA	Bicinchoninic acid
PCR	Polymerase Chain Reaction
SEC	Size Exclusion Chromatography
MeG	Methyl Gallate
FLT	Flapless Tannase
PROSS	Protein Repair One Stop Shop
T_m	Melting Temperature
k_{inact}	Rate of inactivation
TTN	Total Turnover Number
R.e.u	Rosetta Energy Units

Chapter 1: Literature Review

1.1 Overview of Literature Review

Arising from sequences comprised of only 20 different types of amino acids, enzymes are extremely useful biocatalytic proteins that contribute to the metabolic functions within all living organisms. These biocatalysts are also now used in a variety of different industries based on their ability to increase the rate of reactions. The incorporation of enzymes into industrial processes is hypothesized to reduce chemical pollutants created from secondary reactions and increase production output. The enzyme explored in this study is tannin acyl hydrolase or tannase (E.C. 3.1.1.20) from *Lactobacillus plantarum* (*Lp*Tan). Tannase is an enzyme that is expressed in a variety of different organisms and used in the breakdown of naturally occurring tannins to access their carbon.[1] Tannins are large polyphenols found commonly in plants, primarily acting as a storage form of secondary metabolites used in defense responses. Based on its ability to breakdown tannins, there has been growing interest in tannase for industrial applications.

1.2 Biocatalysts in Industry

Chemical reactions are important to industry, contributing to the production of useful compounds or the degradation of a harmful ones.[2] Catalysts are small molecules, proteins or ions used to help speed up reaction processes. They achieve this by stabilizing transition states, interacting with reactants or breaking/forming chemical bonds, ultimately lowering the activation energy of the reaction.[3] Catalysts that can be used in industrial processes can be separated into two categories, biocatalysts or inorganic catalysts.

Inorganic catalysts are catalysts that do not contain any carbon atoms in their structure, generally being composed of metal elements or metal complexes. Biocatalysts, also known as enzymes, are proteins produced by living organisms that play a fundamental roles in important

functions within cells, generally increasing the rates of a wide range of reactions to enable timely physiological responses. Due to their importance and variable uses, enzymes have a variety of different structures and functions (Table 1.1).[3-7]

Table 2.1: Different Classes of Enzymes and their Function, with Examples of Some Subclasses.[8]

Class	Reactions	Enzymes
Oxidoreductases (EC 1)	Transfer of Hydrogen, Oxygen or an electron between molecules	Dehydrogenases, oxidases, oxygenases, peroxidases
Transferases (EC 2)	Transfer of groups of atoms from one molecule to another	Fructosyltransferases, transketolases, acyltransferases, transaminases
Hydrolases (EC 3)	Hydrolytic cleavage of bonds	Proteases, amylases, acylases, lipases, phosphatases, cutinases
Lyases (EC 4)	Non-hydrolytic by elimination or addition reactions	Pectate lyases, hydratases, dehydratases, decarboxylases, fumarase,
Isomerases (EC 5)	Transfer of group from one position to another within a molecule	Isomerases, epimerases, racemases
Ligases (EC 6)	Covalent joining of two molecules	Synthetases, ligases

When considering the differences between these two types of catalysts (inorganic versus bio-based) for application to industrial processes, many pros and cons are identified (Table 1.2) such as environmental effects, specificity and profitability. Overall, this analysis suggests that enzymes should be the more favoured in industrial processes, being able to very specifically and selectively bind to their desired substrate and catalyze a specific reaction on it, while also being able to perform in mild reaction conditions.[9-13] Enzymes are also renewable and can reduce

environmental impact when compared to inorganic catalysts. Inorganic catalysts generally have less substrate specificity, react with multiple substrates and produce multiple products (which can be of value), often need harsh chemical or physical conditions to function and require extra energy for the reaction to initiate, and are sometimes environmentally unfriendly being non-renewable and create or are pollutants.

Table 1.3: Pros and Cons of Inorganic and Bio-base Catalysts in Industry Usage [13]

Type	Pros	Cons
Biocatalyst	• Increased Efficiency • High specificity & Selectivity • Mild Reaction Conditions • Reduced Environment Impact • Compatibility with Biological Systems • Versatility & Diverse • Potential for Renewable Resources	• Production costs • Sensitivity to Environmental Conditions • Limited stability and shelf life • Complexity and challenges in understanding catalytic mechanisms
Inorganic Catalyst	• Robustness and durability under harsh conditions • Longevity and reusability • Scalability for large-scale industrial production • Economic viability with abundant and low-cost raw materials	• Limited selectivity • May require higher temperatures and pressures for certain reactions. • May generate environmentally harmful by-products

There have been many advances in the usage of biocatalysts in industry over the last 20 years such as Carbios, that use enzymes to recycle plastic or EginZyme that develops sustainable biomanufacturing processes using enzyme-based technology. More and more companies are

realizing the advantages biocatalysts have over their inorganic counterparts.[14, 15] This increase in demand for biocatalysts can be linked to the often-conflicting list of increasing demands for consumer goods, reduction of natural resources depletion, and environmental safety.

Despite all their benefits, challenges still arise that can make it hard to incorporate microbial and/or enzymatic breakdown into many industrial processes.[15] Some existing problems include insufficient activity or stability for industrial performance and environmental sensitivity causing them to denature.[16] These problems are exacerbated by the time and cost it takes to research and optimize biocatalysts for industrial applications.

1.3 Protein Engineering

The evolution of biocatalysts in industry can be visualized as occurring in three waves where each wave can be described as an impact on industrial processes.[7,10,17] The first wave was the development of methods that enabled application of naturally occurring enzymes in their native structure, with some modern examples being proteases in laundry detergents, glucose isomerases to convert glucose to fructose and utilizing penicillin G acylases to make antibiotics. The second wave of biocatalysis development involved the applications of protein engineering techniques looking to increase enzyme performance of naturally occurring enzymes via re-design to elicit improved stability, activity, selectivity or substrate specificity. Examples of such protein engineering techniques include site directed mutagenesis and directed evolution. The third and final wave of biocatalysis development builds on the second wave incorporating synthetic biology and the creation/improvement of novel enzymes.

Expanding on this, one must consider techniques like rational design, semi rational design and/or de novo design to improve a variety of different criteria such as activity, substrate specificity, stability.[18-20] Rational design uses previous knowledge of the protein's function to

guide the design process and is best used if there is a enough information on the structure-function relationship of the protein of interest. Semi-rational is like rational design but also includes experimental approaches creating libraries of variants and is best used when the structure-function relationship is not fully understood. *De novo* design is the creation of a new protein, new scaffolds hosting new or exciting chemistries.[19, 21-23]

Also notable, the application of computational prediction and design software. One example of these techniques include Protein Repair One Stop Shop (PROSS), a software that combines bioinformatics with the predictive computational algorithm, RosettaFOLD to predict possible stabilizing mutations and/or by introducing new types of chemistries that do not occur naturally within proteins (or even within nature) by incorporating non-naturally occurring amino acids using genetic code expansion. Perhaps the culminating concept in the third wave is the idea of de novo generation of novel proteins with desired functionalities.[20,24]

Detailed literature reviews pertaining to the specific protein engineering methodologies used in this study can be found in the introductory sections to chapter 2 (genetic code expansion) and chapter 3 (computational stabilization prediction and computation library screening).

1.4 Function and Application of Tannins

With respect to the substrates targeted by the biocatalyst *Lp*Tan, tannins are large water-soluble polyphenols in plant leaves, bark, fruit wood and roots, with molecular masses ranging from 3000 to 20,000 Da.[25,26] They are the fourth most abundant plant constituent after cellulose, hemicellulose and lignin. Tannins can be classified into 3 groups: condensed, hydrolysable and phlorotannins (Figure 1.1). Phlorotannins are a unique group of tannins that occur solely in marine algae and are less complex in structure than the other two groups of tannins. Condensed tannins are the most common type of tannins, being found in forage legumes, trees and most

plant stems. Condensed tannins have a more complex structure and higher molecular weight with their structure consisting of flavonoid units linked by carbon-carbon bonds. Hydrolysable tannins consist of a sugar core (often D-glucose) that are esterified with gallates molecules (hydroxylated benzene rings) and are found at lower concentrations in plants when compared to condensed tannins.[27]

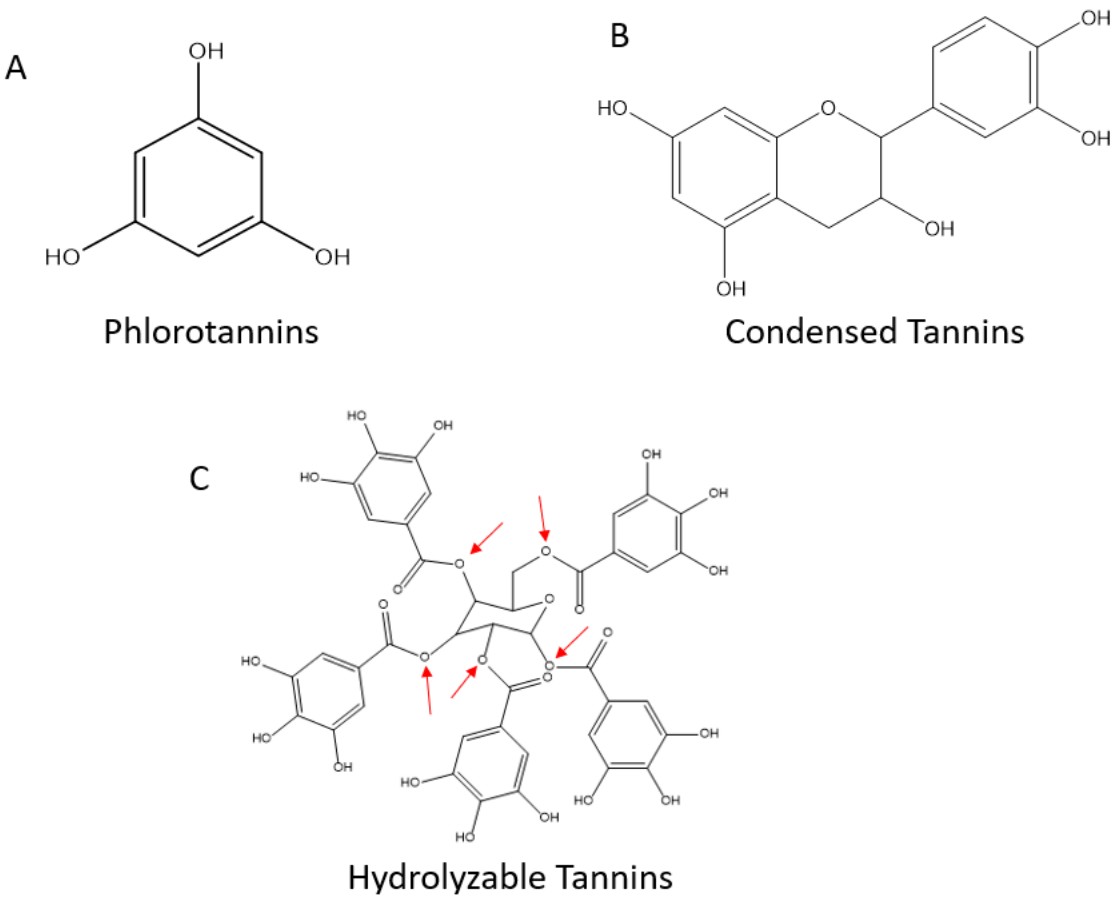


Figure 1.1 : Different Structures of Tannins Found in Nature. A) shows the typical chemical structure of phlorotannins, B) shows the typical structure of condensed tannins and C) shows the typical chemical structure of hydrolyzable tannins the tannin of interest. (red arrows indicate hydrolyzable sites)

Within plants, tannins are used as a storage form of secondary metabolites, serving as part of the plants chemical defense system against invasion by pathogens or attack.[25,28,29] Tannins, themselves can inhibit enzymes, interfere with substrates required for growth, affect

metabolism through the inhibition of oxidative phosphorylation, deprive systems of metal ions and are able to form complexes with the cell membranes increasing membrane permeability. These attributes are caused by the multiple phenolics groups contained within tannins, which bind to proteins and metal ions.

Interestingly, tannins have also found applications across a variety of different industries, with the global market for them increasing rapidly over the years, with a predicted compound annual growth rate of 5.8 % from 2016-2025.[25,30] In the food industry, tannins can be used as packing materials, food enhancers and food preservatives, while in the wood industry tannins are used to improve bio-durability. Tannins are also used in the pharmaceutical industry, based on their various molecular inhibitory properties described above.[31, 32]

1.5 Applications of Tannase in Industry

Tannase is an enzyme that has been used in a wide variety of different industrial applications, such as food, brewing and pharmaceutical due to its ability to cleave hydrolysable tannins producing glucose, gallic acids and gallic acid derivatives. An example of tannase in industry is the production of gallic acid. Gallic acid is mainly used in the pharmaceutical industry for the production of trimethoprim, an antibacterial agent.[33,34] Further investigations into gallic acid also show antimicrobial activity against human and animal pathogenic bacteria such as *Escherichia coli*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*.[35] Gallic acid has also been shown to exert anti-cancer activities via biological pathways, including migration, apoptosis, cell cycle arrest, angiogenesis, and oncogene expression. [33]

Another industry that uses tannases is the food industry. Due to its ability to break down catechins, tannases see extensive usage in the instant tea business by increasing the extractability and cold-water solubility of catechins found in teas.[36-38] This is true also for fruit juices as it

eliminates bitter tasting tannins while also releasing the glucose core, as well as the beer industry where it reduces haze formation by hydrolyzing wort phenolics, which complex with other chemicals in the beer.[38] Tannase is also used in the tanning industries, acting as a hydrolyzing agent in cleaning up highly polluting tannins from the effluent.

Currently methods for gallic acid production from gallnuts are performed using chemical hydrolysis, which in turn produces large amounts of industrial wastewater.[39] Therefore, there is a need to create a more industrial suited enzyme.

1.6 The Fundamentals of *Lp*Tan

The industrial enzyme of focus in this study is tannase (E.C 3.1.1.20) or tannin acyl hydrolase from *Lactobacillus plantarum* (*Lp*Tan). It is a serine esterase that catalyzes the hydrolysis of the galloyl ester bond playing a key role in breaking down tannins.[40, 41] More broadly, tannases are present in plants, animals, and many different microorganisms such as bacteria, yeast and fungi.

1.6.1 Serine Hydrolase Family

Tannase is a member of the larger family of serine hydrolases that are hydrolytic enzymes with broad range of different substrate specificities and regiospecificities. These are important enzymes being used in biological processes like digestion, blood clotting, immune responses and various metabolic pathways.[42] Belonging to the α/β- hydrolase superfamily, most serine hydrolases have two distinct domains: the catalytic domain and the cap domain.[43-45] The catalytic domain or the α/β- hydrolase domain (ABHD) is generally characterized by eight parallel β-sheets surrounded by six α-helices and contains the core catalytic machinery – catalytic dyad/triad and the oxyanion hole. The cap domain contains substrate binding sites, and the more conserved catalytic domain containing the catalytic triad (or in some cases a dyad)

formed by a group of key molecules (usually amino acid side chains) that include a highly nucleophilic serine, that initiates catalysis of a reaction.[46-48] In addition to the nucleophilic serine, a base and an acid are also required. The base (usually a histidine) is used to deprotonate the nucleophile (serine) and start the nucleophilic attack on the substrate, the acid is used to help with stabilization during transition states.[48] The oxyanion hole is a "pocket" in the active site and is used to help with stabilizing the negative charge on an alkoxide formed in the transition states, utilizing the backbone of the amino acids involved (Figure 1.2). The catalytic triads/dyads differ in different serine hydrolases with most presenting a 'classical' serine-histidine-aspartic acid catalytic triad, but also with instances of serine-histidine-glutamic acid triad and even histidine-cysteine catalytic dyads having been characterized.

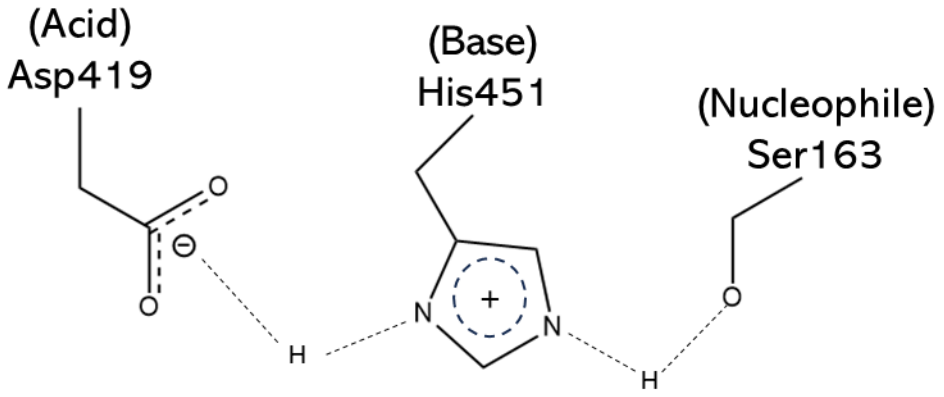


Figure 1.2: Key Residues in the Serine Hydrolase Enzyme *Lp*Tan's Catalytic Triad. D419 acts as an acid by donating its proton to help stabilize H451, while H451 acts as a base taking the proton of S163 turning it into an alkoxide. This in turn makes the S163 more nucleophilic inciting the reaction.

Unlike the catalytic domain, which is conserved among serine hydrolases, the cap domain varies in location, sequence and structure (Figure 1.3).[49] The cap domain contains mostly helical secondary structures, situated between the beta structures of the catalytic domain. It is usually at

the N-terminus of the protein and generally consists of residues that help bind the substrate in the catalytic site.

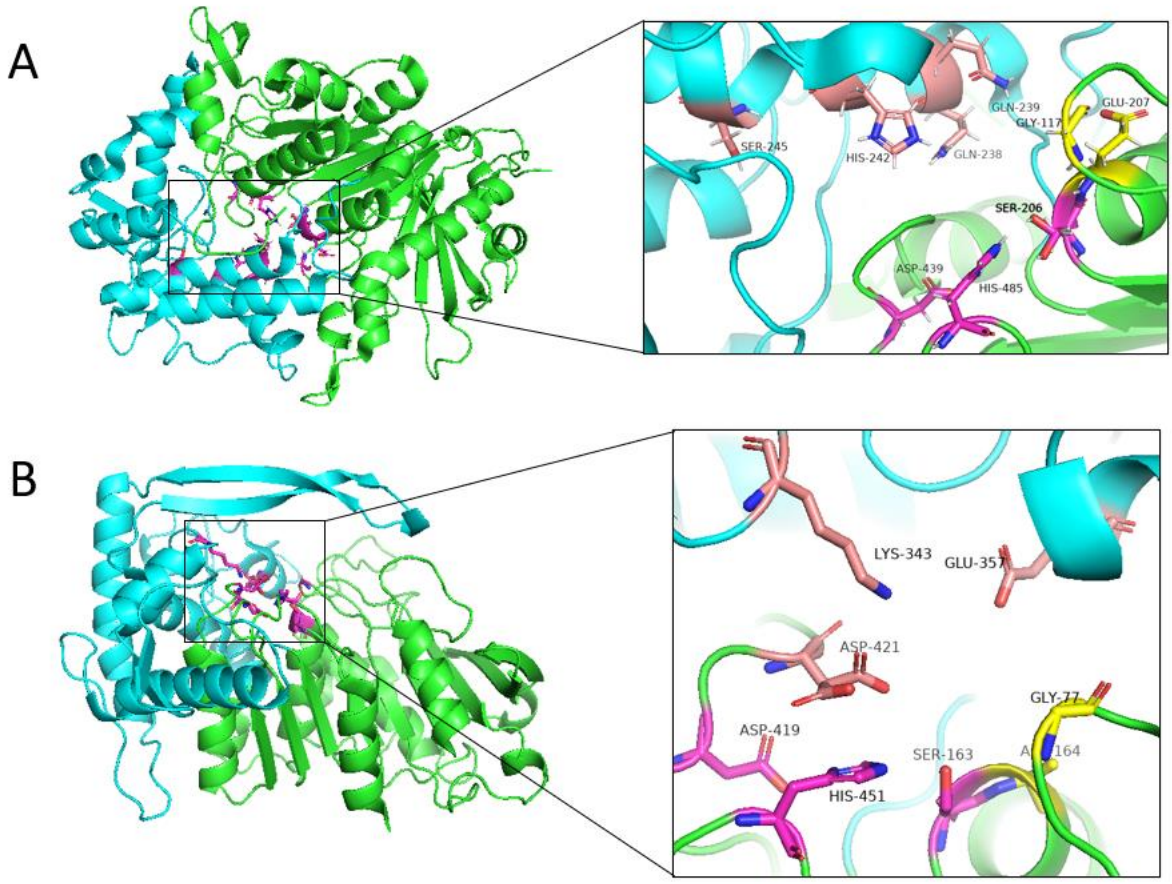


Figure 1.3: Comparison of Domains and Active Sites of *Aspergillus niger* and *Lactobacillus plantarum* Serine Hydrolases (Tannases). (A) tannase from *Aspergillus niger*,[50] and (B) *Lactobacillus plantarum*.[51] The green represents the conserved α/β hydrolase fold, while in cyan is depicted the cap region. Active site is shown in magenta in protein model, while for active site zoom window, substrate binding residues are in peach, the catalytic triad are in magenta and the oxyanion hole residues are in yellow. This image depicts pdbid:7K4O(A) and 3WA6(B) with the images generated using PyMOL.[52]

1.6.2 *Lp*Tan Structure and Mechanism

Like most bacterial tannases, *Lp*Tan exists as a monomer in solution and contains a conserved tripeptide comprised of C204-P205-I206, that is a signature used to identify tannases.[53, 54] There exists little sequence homology between bacterial tannases and tannases

from other different organisms. When compared to other bacterial tannase homologues, *Lp*Tan has 27-88 % sequence identity, with substate binding and catalytic triad residues being conserved. Although, the Asp group used as an acid in the catalytic triad is sometimes substituted Glu in other bacterial tannases.

*Lp*Tan follows the typical structure of serine hydrolases, containing an ABHD and catalytic triad located in the core of the enzyme (Figure 1.4). *Lp*Tan has a total of 18 α-helices and 13 β-strands, of these β-strands 9 (1-7,12 & 13) form a twisted left-handed β-sheet, with strands 1 and 3 running antiparallel to the other strands.[51, 53, 54] Most of the helices are positioned opposite to helices 3 and 18, which are located on the concave side of the major β-sheet. Helix 10 runs perpendicular to sheet 13 on the C-terminal side of the major β-sheet, combining with helices 11, 12 and a 2 stranded antiparallel sheet of strands 10 and 11. *Lp*Tan also has a "flap" formed from the antiparallel sheets 8 and 9 (Figure 1.4). The loop region between sheets 8 and 9 forms bonds with β4 and α3 forming of a tunnel that is approximately 12 Å which is seen on the protein's surface.

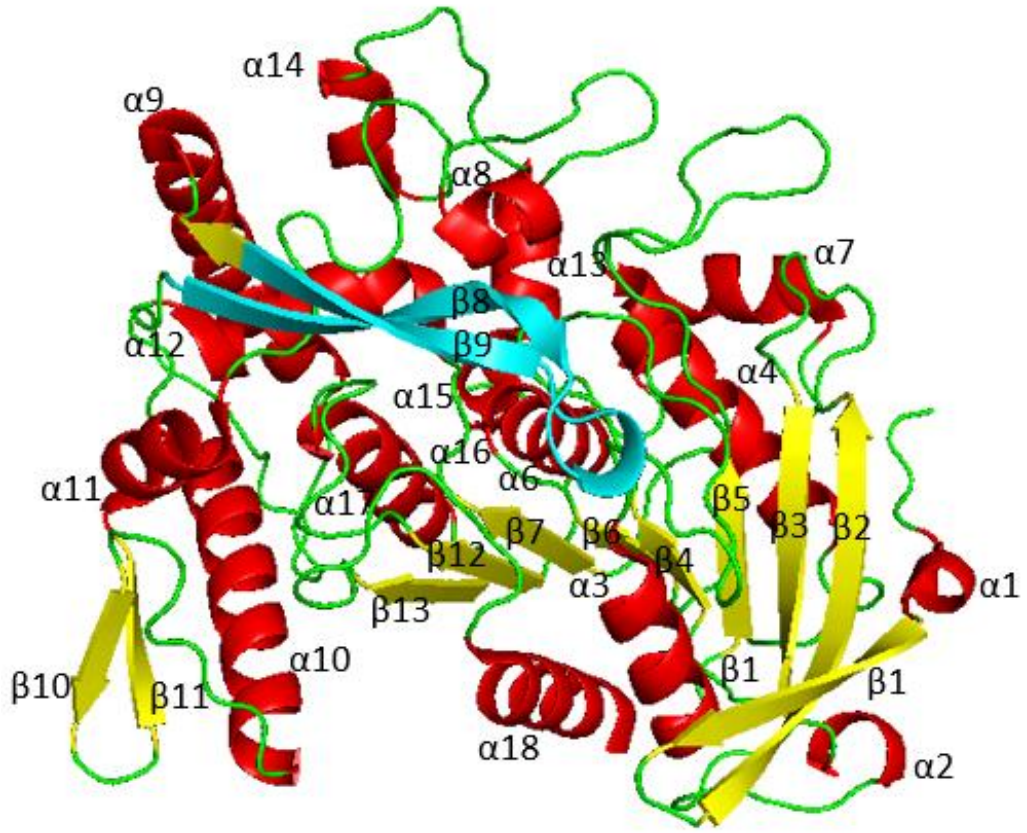


Figure 1.4: Secondary Structures of *Lp*Tan. The α-helices are shown in red, β-sheets in yellow, loops are in green and the large flap region in cyan. This image depicts pdbid:3WA6 and was generated using PyMOL.[51, 52]

When looking at the catalytic domain of *Lp*Tan, this domain contains the classical ABHD found in all serine hydrolase with 8 helices and 4 strands.[53, 54] The cap is located between sheet 7 and helix 16 and is positioned at the C-terminus of the structure. *Lp*Tan uses S163, H451 and D419 to form its catalytic triad (Figure 1.3), with the serine being positioned at the tip of the nucleophile elbow of β-strand 6 and α-helix 6 and under the flap region residing in the groove, while both H451 and D419 are located towards the end of the protein. The oxyanion hole is formed from the nitrogen backbone of G77 and A164, helping to stabilize the transition states of the substrate by forming H-bonds to oxygen atom of the carboxyl group. Side chains of D421, K343 & E357 form H-bonding networks with the hydroxyl groups of the substrate. Unlike the

other two binding residues, D421 is in the catalytic domain and has two conformations when in the apo structure.

The catalytic mechanism of tannase (Figure 1.5) is based on the hydroxyl group of S163 being depronated by H451 and becoming an alkoxide group.[54] This alkoxide group starts the nucleophilic attack on the carbonyl carbon creating a tetrahedral intermediate, which is further stabilized by the oxyanion hole H-bonding of G77 & A164. H451 then acts as an acid and collapses the tetrahedral intermediate, producing an alcohol and the acyl-enzyme intermediate. The lone pair electron on H_2O then proceeds to attack the carbonyl carbon of the acyl-enzyme intermediate, causing the collapse of acyl-enzyme intermediate, creating gallic acid (and derivatives). Throughout this process H451 interacts with D419 and C204, with D419 acting as an acid and donating a proton to H451, while C204 is hypothesized to play a role in proton transfer.[54, 55]

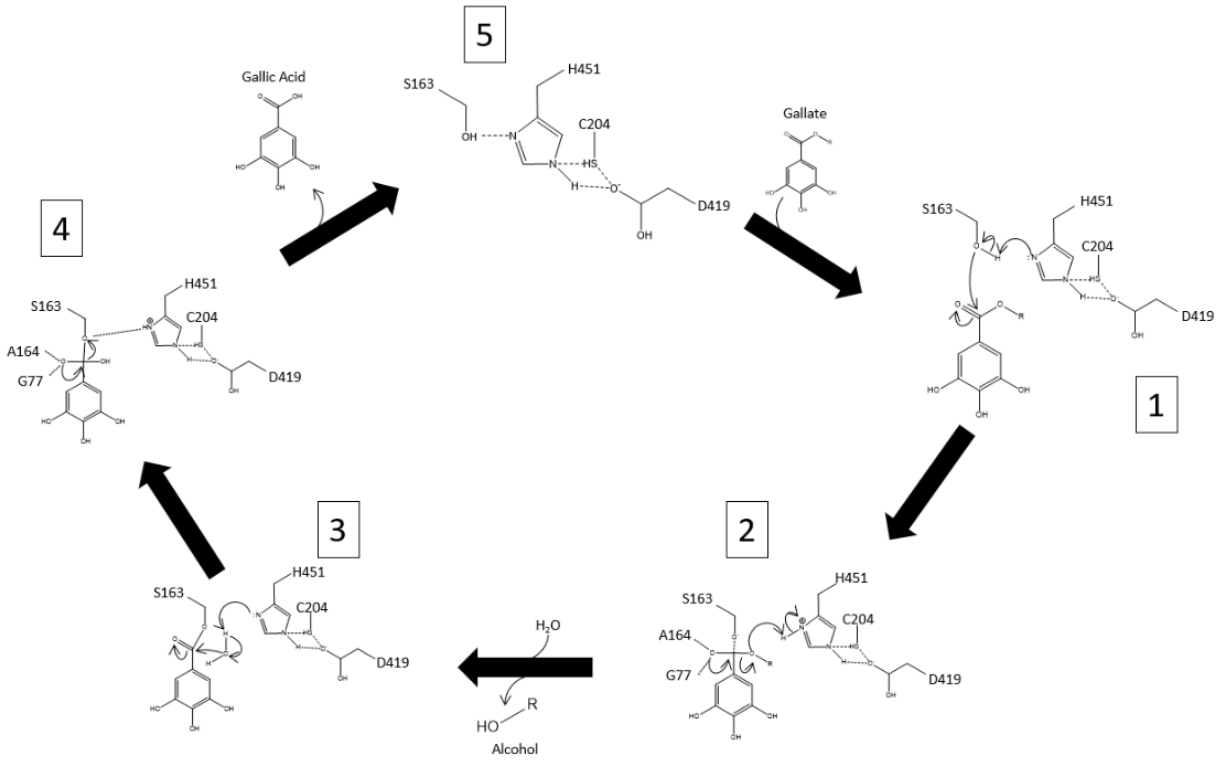


Figure 1.5: Catalytic Mechanism of *Lp*Tan. Movement of electrons as reaction occurs is denoted are shown using arrows.[41]

1.7 Aim of Study

In this study, I looked to increase the effectiveness of *Lp*Tan's ability to breakdown tannins, toward improving its performance in industrial application. Of note, this project is collaborative with feed producers in Western Canada, who were interested in reducing (bitter tasting and un-palatable) tannin content in feedstocks for improved animal feeding. It is important to keep feed production costs as low as possible, and thus any enzymes being applied must be optimized for efficiency to overcome cost barriers. Tannase from *Lactobacillus plantarum* was selected for this study as there were previous studies involving sequence/expression information and a basic characterization of the enzyme.[51,54-56] The aim of this study was to demonstrate the viability of optimizing the functionality of tannase. In particular imparting an increase in the total catalytic turnover of *Lp*Tan for industrial applications

in the animal feed sector. This was attempted using two different methods: Method one (Chapter 2) targeted the active site by using genetic code expansion to convert a catalytic triad to a dyad by mutating the Histidine at site 451 to its noncanonical counterpart 3-methyl-histidine, while method two (Chapter 3) looked to increase the stability, by decreasing the rate of inactivation or increasing the melting temperature using computational prediction software, (PROSS) to identify possible sites for mutagenesis for screening by combinatorial library.

Chapter 2: Testing Genetic Code Expansion for Increased Catalytic Efficiency in *Lp*Tan

2.1 Abstract

Tannin acyl hydrolase or tannase functions as an esterase that catalyzes the hydrolysis of ester bonds in gallo-tannins producing gallic acid and its derivates. Tannase is used in a variety of different industries such as food, chemical, agriculture, and pharmaceutical and thus represents an important target for biocatalytic development. Recent studies have emphasized the potential of genetic code expansion (GCE) to modify proteins and increase catalytic activity of enzymes.[57-59] GCE allows for the incorporation of non-canonical amino acids (ncAA). In this study the impact of replacing the Histidine 451 in the esterase catalytic triad with its non-canonical counterpart 3-methyl-histidine (NmH) was evaluated. A decreased reliance on the Asp of the catalytic triad, while shortening the distance between the His and the Ser was proposed to create a more efficient catalytic dyad. A pyrrolysl tRNA synthetase (PyIRS)/ tRNApyl pair were used in combination with a tannase mutant encoding an amber stop codon at position 451. The activity of the derived NmH451 tannase was compared to wild-type and unexpectedly showed a decrease in activity and substrate affinity. Mass spectrometry was applied to validate incorporation but proved to be inconclusive.

2.2 Introduction

Classical site directed mutagenesis is a technique used to delete one or more amino acids or change one or more amino acids in a protein sequence to one of the other 19 naturally occurring amino acids to induce a predicted functional change that may test a mechanistic hypothesis or improve activity, substrate affinity or specificity under given environmental conditions (temperature, pH and solvent). While mutagenesis continues to be a popular method for protein engineering based on its simplicity and ease of application, the limited number of

possible amino acids that can be used as substitutes can be limiting when trying to improve or create new interactions that haven't been studied. GCE is an emerging technique that helps to introduce new interaction between amino acids in protein, adding non natural amino acids, referred to as non-canonical amino acids (ncAA).[60,61] By expanding the list of possible amino acid side chains that can be incorporated, the hope is to create new and unique interactions and reactions within proteins.

GCE works by first substituting a selected three-base codon encoding a particular amino acid in each protein, with a stop codon.[62, 63] The stop codon will usually encode an amber stop codon (UAG) due being used the least amount in *Escherichia coli*.[62] During translation of the protein, when the mRNA is read, an engineered orthogonal translation system (OTS) comes into play.[63] Essentially, an engineered aminoacyl-tRNA synthetase attaches the ncAA to an engineered tRNA that recognizes the amber codon and will then incorporate the ncAA into the protein at the site (Figure 2.1). When looking to GCE, orthogonality is an important consideration, as the aminoacyl tRNA synthetase (aaRS) must be selective for its cognate tRNA. To achieve this, OTSs are derived from aaRS/tRNA pairs sourced from phylogenetically distant organisms. In the case of OTSs made for bacterial expression systems archaeal and eukaryotic sources are mined and optimized. For this study we leveraged an OTS from *Methanosarcina Mazei*, including a mutated $tRNA_{CUA}^{Pyl}$ and cognate tRNA for application to an E.coli recombinant system.[64]

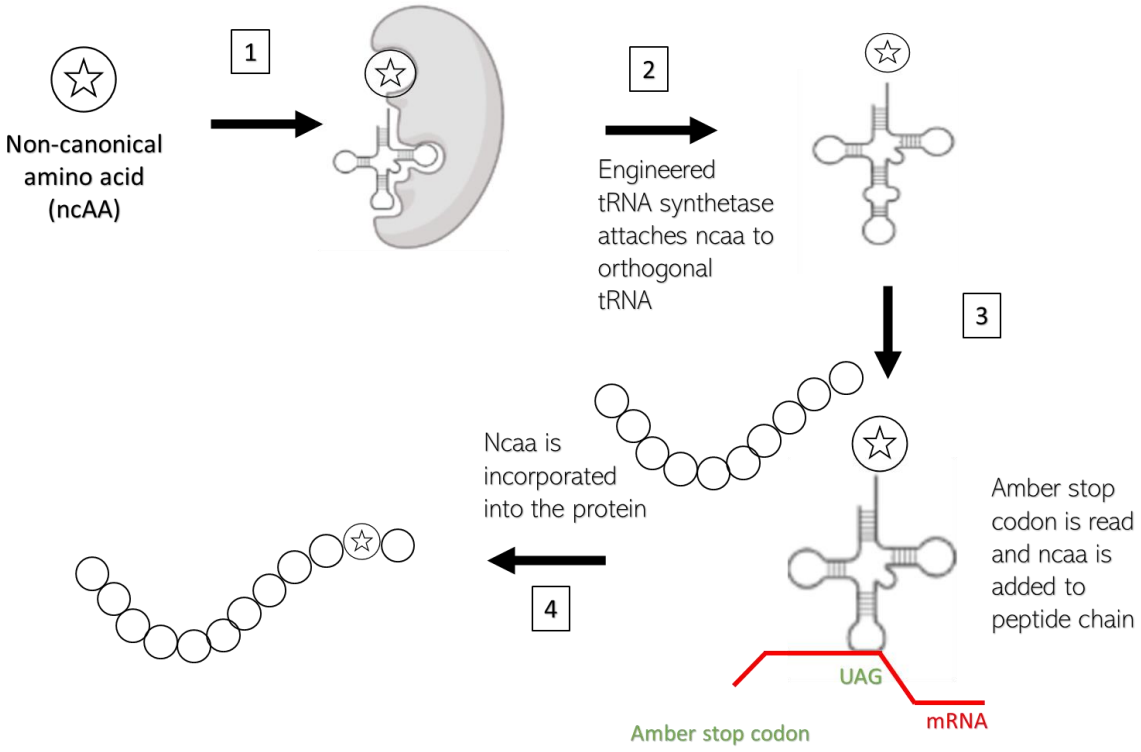


Figure 2.1: Flowchart Showing NcAA Incorporation into a Growing Peptide Chain. 1 & 2) Associated engineered tRNA synthase attaches ncAA to engineered orthogonal tRNA. 3) When the amber stop codon (UAG) is read in the mRNA sequence during translation. 4) NcAA is incorporated into the peptide chain at that site.

GCE is proving to be a useful technique with a variety of different demonstrated applications to date, including new chemical functions, fluorescence, ligand binding, cross-linking or photocaging.[65, 66] In a previous study, GCE was used to significantly increase the catalytic activity of TEM-1-β-lactamase, where V216 was substituted with the ncAA p-acrylamido-phenylalanine, yielding approximately a 10-fold increase in activity.[57] In this same paper they also substituted V216 with the other 19 canonical amino acids where the most active mutant (V216I) showed only a two-fold increase in activity. In another studies, GCE using the ncAA NmH was used to increase the catalytic function of myoglobin by about 4-fold.[58] At the same time, in a designed hydrolytic enzyme, an NmH substitution was found to improve catalytic function, by preventing the formation of unreactive acyl-enzyme intermediates derived from the

canonical histidine.[59] Through the introduction of NmH, it was hypothesized that by creating a catalytic dyad between NmH451 and S163 would result in a faster catalytic rate, due to decreased reliance on D419 for catalytic function resulting in a faster turnover rate. Catalytic dyads work in the same way as catalytic triads, but they do not utilize an acid stabilizer, consisting solely of a nucleophile and a base.

In this study, GCE was used to explore the possibility of improving the catalytic activity of *Lp*Tan. In this chapter, the catalytic triad histidine at site 451 in *Lp*Tan is substituted for its noncanonical NmH counterpart (Figure 2.2). This will test the hypothesis that introduction of NmH can enable a faster catalytic rate by increasing the turnover number in hydrolases more broadly, putatively by converting the catalytic triad into a catalytic dyad and thereby decreasing reliance on the catalytic aspartic acid.

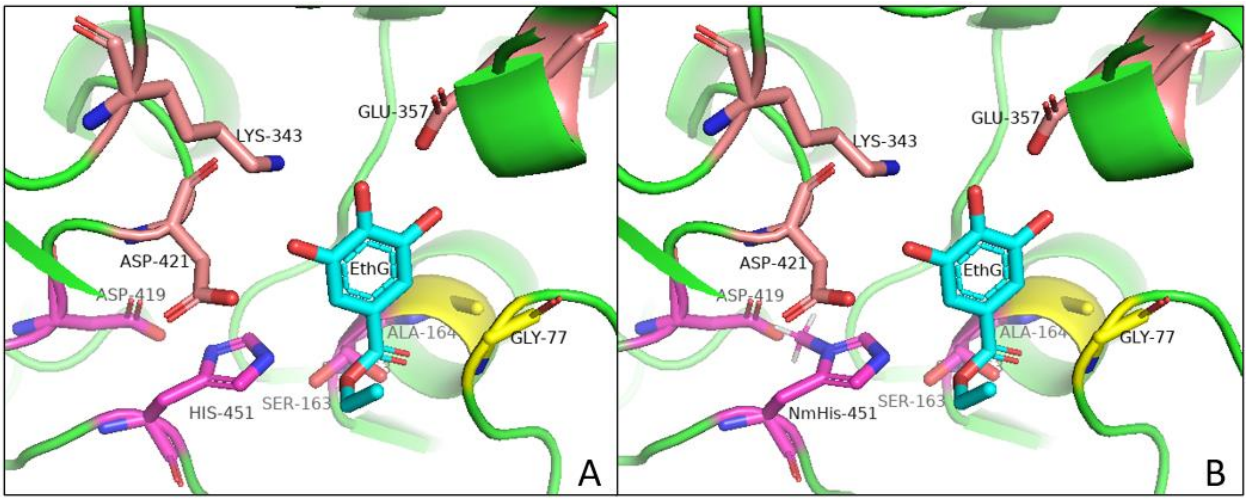


Figure 2.2: Active Site Comparison between *Lp*Tan and GCE mutant *Lp*Tan-His451NmH. A) Crystal structure of LpTan. B) Model of LpTan-His451NmH. The catalytic triad residues are shown in magenta, the oxyanion hole in yellow, active site stabilizers in peach and the substrate, ethyl gallate in cyan. Blue and red colouring on molecules indicate nitrogen (blue) and oxygen (red). (GCE mutant model (B) was not crystalized, but created using Pymol) The structures and model were generated from pdbid: 4J0K using PyMOL with energy minimization.[52]

2.3 Methods

2.3.1 Chemicals

All chemicals and reagents were purchased from Millipore-Sigma unless otherwise stated below.

2.3.2 Construction of the *Lp*Tan Expression Vector

Codon optimized coding sequence for the *Lp*Tan protein (Accession #: CP020564.1) was synthesized by Biobasics Inc and then cloned in the NdeI and XhoI restriction sites without a stop codon at the C-terminal end, in the multiple cloning regions of the pET31b+ vector, to encode the *Lp*Tan protein with a C-terminal His-tag.

2.3.3 Production and Purification of *Lp*Tan

A starter culture of 10 mL of Luria Broth (LB) media (10 g Tryptone, 10 g Sodium Chloride and 5 g Yeast Extract) with 100 µg/mL of ampicillin was inoculated with BL21 DE3 GOLD cells (Aligent Technologies) *Escherichia coli* containing the pET31b+-*Lp*Tan vector. The starter

culture was grown overnight in a MaxQ6000TM incubator at 37 °C at 250 rpm and then used to inoculate 100 mL of LB media supplemented with 100 µg/mL ampicillin and incubated at 37 °C at 250 rpm until OD600 reached 0.3-0.4. After OD600 reached 0.3-0.4, the temperature was reduced to 20.5 °C. The culture was grown to an OD600 of 0.6-0.7, after which a final concentration of 0.7 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression after which the culture was left to grow overnight at 20.5 °C at 150 rpm.

The overnight culture was centrifuged at 3750 x *g* for 30 min at 4 °C and the supernatant discarded, and pellets placed on ice. Lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 10 mM imidazole) with protease inhibitors (0.1 mM phenylmethylsulphonyl fluoride (PMSF), 3 U/mL benzonase, 1 mg/mL lysozyme, 5mM Dithiothreitol (DTT), 20 mM $MgCl_2$) was used to re-suspend the pellet. The cells were lysed on ice by sonication for 30 cycles (each cycle being 25 seconds at 30 % duty and 40 output power and 30 second cooldown) using a Sonics Vibra cell VC 505. Obtained lysates were centrifuged at 17000 x *g* for 30 minutes at 4 °C. Obtained lysate pellets were frozen at -20 °C, and lysate supernatants were applied to nickel nitriloacetic acid (Ni-NTA) affinity chromatography as follows.

QIAexpressionist™ Ni-NTA affinity resin (Qiagen) was used to purify the recombinant wild type tannase from the supernatant. Stacked Ni-NTA resin (1 mL) was equilibrated in 10 mL pH 8.0 lysis buffer. All collection tubes were placed on ice during elution. The cell lysate supernatant was loaded onto the column by gravity flow and the flowthrough was collected. The loaded column was washed three times with 1mL pH 8.0 wash buffer (10 mM Tris-HCl, 150 mM NaCl, 20 mM imidazole), each time collecting the flowthrough. The column was placed on ice and 1 mL of pH 8.0 Elution Buffer (10 mM Tris-HCl, 150 mM NaCl, 250 mM imidazole) added and incubated for 4 minutes before flowthrough was collected and the elution process was

repeating two more times. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to confirm successful Ni-NTA purification.

For size exclusion chromatography (SEC), Ni-NTA elution fractions were pooled and concentrated using an Amicon Ultra-4 mL 30k MWCO concentrator and centrifuged at 3250 x *g* for 10 minutes. The process was repeated until only 1-1.5 mL of eluate was remaining. The concentrated *Lp*Tan sample was subjected to SEC using a HiLoad 16/600 Superdex 200 column attached to an AKTA PURE 25L Fast Protein Liquid Chromatography (FPLC) instrument. A flowrate of 0.300 mL/min was used using SEC Buffer (10 mM Tris-HCl, 150 mM NaCl pH 8.0) with UV detection at 280 nm. After eluting from SEC-FPLC, the final concentration ogf pure *Lp*Tan protein was determined by bicinchoninic acid (BCA) assay.[67]

2.3.4 Production of GCE Encoded His451NmH *Lp*Tan

The codon representing site H451 (CAT) in the pET31b+-*Lp*Tan expression vector was changed to encode an amber stop codon (TAG) by site directed mutagenesis at Bio-Basics Inc yielding pET31b+- *Lp*Tan- His451Amb expression vector.

The evolved (N346A/C348A) *Methanosarcina mazei* PylRS scaffold (MmPyL(N346A/C348A) HRS) and corresponding tRNApyl gene ($tRNA_{CUA}^{Pyl}$) in the pEVOL-pylT vector was a gift from Dr. Dieter Soll at Yale University.[68] This was subsequently mutated to incorporate 5-point mutations (L270I, Y271F, L274G, C313F, Y349F) shown to increase selectivity for NmH in a *Methanosarcina barkeri* PylRS with the final 'evolved' MmPylRS containing vector (pEVOL-MmPyl-EVOL-HRS) kindly provided by previous lab members Hannah Parker and Tony St-Jacques.[64]

Starter cultures of *E. coli* BL21(DE3) GOLD (Aligent technologies) co-transformed with the pET31b+ -*Lp*Tan-His451Amb and the pEVOL-MmPyl-EVOL-HRS vectors were grown overnight in 10 mL of LB supplemented with 100 µg/mL ampicillin and 34 µg/mL of chloramphenicol. These were used to inoculate 100 mL of LB media supplemented with 100 µg/mL ampicillin and 34 µg/mL of chloramphenicol, which were incubated at 37 °C at 250 rpm until OD600 reached 0.3-0.4. After the OD600 reached 0.3-0.4, the temperature was reduced to 20.5 °C, and the cultures grown to an OD600 of 0.6. At this point a final concentration of 2 mM NmH and 20 % w/v arabinose were added to induce expression and activation of the OTS machinery. After a further hour of incubation, a final concentration of 0.7 mM IPTG was added to induce *Lp*TanHis451Amb expression, and the culture was left to grow overnight at 20.5 °C at 250 rpm. *Lp*Tan-His451NmH was then purified and assessed as described above for wild type *Lp*Tan. To ensure proper expression, and validate incorporation, a variety of controls were conducted in which the reagents, arabinose, NmH and IPTG were added or not added.

2.3.5 Kinetic Analysis of *Lp*Tan and *Lp*Tan-His451NmH

Tannase activity was assessed based on chromogen formation between the tannase reaction product (gallic acid) and rhodanine.[69] The chromogen was detected at a wavelength of 520 nm using a SpectraMax M5 microplate reader read in 96 wells Greiner microplates. A standard curve using gallic acid concentrations from 0 – 0.3 mM was prepared by reaction with 0.067% w/v rhodanine in methanol for quantification of obtained chromogen. The substrate methyl gallate (MeG) (0-3 mM) was incubated with ~0.25 µg of SEC-purified enzyme in 50 mM Tris pH 8.0 for 5 minutes at 30 °C. 0.067% w/v rhodanine in methanol was then added and incubated at 30 °C for 10 minutes. Finally, 0.5 M sodium hydroxide was added and incubated for 10 minutes at 30 °C before chromogen formation is quantified. Final reaction volume was 300 µL.

The tannase activity (U/µL) was defined as the amount of enzyme required to produce 1 µmol of gallic acid / sec under assay conditions. All tests were performed in triplicate and analyzed with Excel solver.

2.3.6 Chymotrypsin Digestion and Mass Spectrometry

*Lp*Tan-His451NmH sample was diluted with Mass spectroscopy buffer (50 mM Triethylammonium bicarbonate and 4.5% SDS at pH 8.0) using a total of 122.5 µg. A ratio of 1:40 (chymotrypsin: *Lp*Tan-His451NmH) was used for chymotrypsin digestion. S-Trap™ mini spin column (ProtiFi, NY, USA) was used for protein digestion and was carried out at 25 °C for 2 hours. Digested peptides were then quantified using Pierce™ quantitative colorimetric peptide assay kit (Thermo Scientific™, catalog number 23275).

Digested peptide was then injected into Vanquish Ultra High-performance Liquid Chromatography (50 µL) instrument coupled to a Thermo Scientific™ Q Exactive™ mass spectroscopy. The reverse phase C18 column Acclaim™ RSLC120 (2.1 x 250 mm, 2.2 µm, Thermo Scientific™) was used in Ultra High-performance Liquid Chromatography at 60 °C during separation. Gradient method consisted of mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in 80% acetonitrile with water) and were used as follows: 4-40% of mobile phase B for 83 min, 40-65% of mobile phase B for 5 min, 65-100% of mobile phase B for 2 min, and held at 100% mobile phase B for the next two min at a flow rate of 50 µL/min. Eluted peptide samples were injected into Q Exactive™ mass spectrometer, where ionization was achieved by heated electro ionization ion source.

Chymotrypsin digestion and shotgun mass spectrometry were performed offsite by Prem Das of the National Research Council Canada in Saskatoon.

2.4 Results

2.4.1 Contributors

Mass spectrometry was performed by Prem Das, at the National Research Council of Canada at Saskatoon.

2.4.2 Recombinant Production of *Lp*Tan-His451NmH

Recombinant production of *Lp*Tan-His451NmH was conducted along with several control conditions. Five different conditions were tested where one or all the inducers including either of IPTG, arabinose and NmH were excluded, compared to the complete induction of *Lp*Tan production. The goal being to demonstrate that without the addition of all three reagents production of the *Lp*Tan-His451NmH variant protein does not occur. IPTG is used to induce expression of the lac operon, allowing the production tannase, arabinose works to express the needed orthogonal engineered tRNA synthetase and NmH is the ncAA that gets incorporated by the orthogonal translation system, when the amber stop codon is read in the sequence. Treatment 1 was used as a negative control containing none of the reagents, treatment 2 used only NmH and arabinose, treatment 3 used only arabinose and IPTG, treatment 4 used only NmH and IPTG and treatment 5 was the test control using all the reagents.

Production of *Lp*Tan-His451NmH under various conditions is shown in lysated supernatants (Figure 2.3), with the expected molecular weight of full-length enzyme being 52 kDa. Looking at the negative control treatments 3 and 4, feint bands around 52 kDa are seen. These bands could result from some leaky production of wild type enzyme or represent truncated GCE mutant as both conditions lack one of the two components essential for full NmH incorporation (NmH or OTS), such that the amber stop codon served as a stop codon. The expected molecular weight of the truncated enzyme is 49 kDa (around 3 kDa less then full

length), although this remains to be validated. Treatment 5, when compared to treatments 3 and 4 shows a stronger band at a similar molecular weight. Thus, this strong band likely corresponds to the full-length GCE mutant protein of interest since condition five contains all the reagents needed for *Lp*Tan-His451NmH production. Notably, further analysis of the pellet from the cell lysate shows that there is a large amount of insoluble protein in the *Lp*Tan-His451NmH variant compared to wild type (Figure 2.4). This insoluble protein might be the truncated protein as based on the large amount of protein seen in Figure 2.4C appearing around the same molecular weight as eluted untruncated protein.

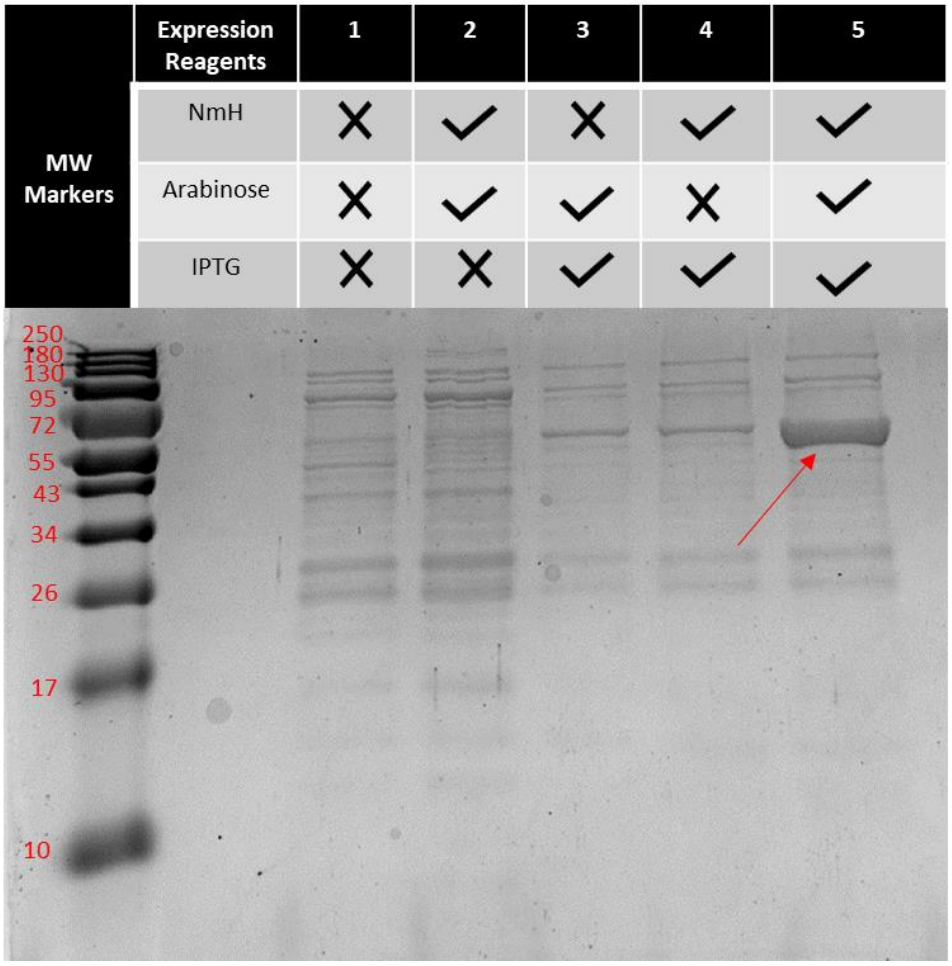


Figure 2.3: SDS-PAGE of the Soluble Fractions of Cell Lysates Arising from the Recombinant Production of *Lp*Tan-His451NmH. SDS-PAGE gel stained with Coomassie Brilliant Blue showing reagents added to difference cultures. Lane 1 is the negative control culture to which none of the inducers were added. Lane 2 is uninduced for LpTan expression, lacking IPTG, but induced for tRNA synthetase expression and includes NmH. Lane 3 is induced for both *Lp*Tan and tRNA synthetase expression but does not contain NmH. Lane 4 is induced for *Lp*Tan but uninduced for tRNA synthetase (lacking arabinose) and contains NmH. Lane 5 is the positive protein production test containing all the inducers.

2.4.3 Purification of *Lp*Tan-His451NmH

Based on the outcomes of the initial control experiments, purification of both wild-type *Lp*Tan and the NmH incorporated *Lp*Tan-His451NmH was attempted. The combination of an initial affinity chromatography step using Ni-NTA followed by a SEC separation yielded a purer sample. (Figure 2.4). Comparing the Ni-NTA purification between *Lp*Tan and *Lp*Tan-His451NmH in Figure 2.4A & 2.4C respectively, we see substantial enrichment of both protein species at the expected molecular weight of 52 kDa for each. The relative enrichment of wild type appears greater than the NmH incorporated variant. Looking at the subsequent SEC elutions, Figure 2.4B shows cleaner fraction lanes, while Figure 2.4D sows some residual contaminating proteins in the *Lp*Tan-His451NmH sample.

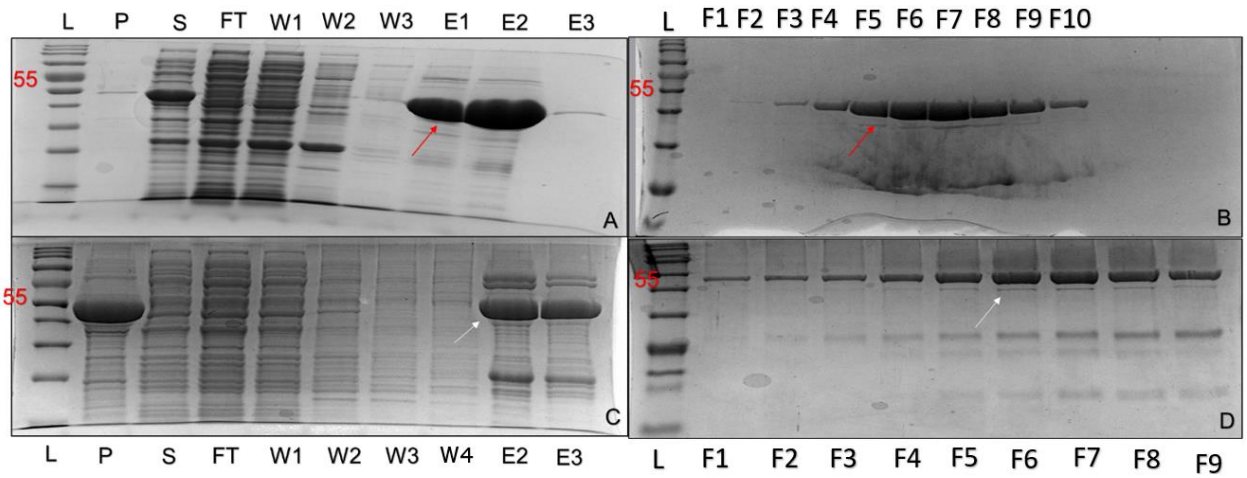


Figure 2.4: SDS-PAGE of Fractions arising from Ni-NTA and SEC purification of *Lp*Tan and *Lp*Tan-His451NmH. A) Ni-NTA purification and B) SE-FPLC of wild type *Lp*Tan. C) Ni-NTA purification and D) SE-FPLC of *Lp*Tan-His451NmH. For Ni- NTA, from left to right in figures A and C samples loaded into wells are pellet (P), supernatant (S), flow through (FT), wash 1 (W1), wash 2 (W2), wash 3 (W3), elution 1 (E1), elution 2 (E2), and elution 3 (E3). For SE-FPLC lanes indicate fractions collected. NEB P7718S colour protein standard ladder was used (L). The molecular weight of *Lp*Tan is 52 kDa with red arrows indicating where LpTan is located and white arrows corresponding to *Lp*Tan-His451NmH.

2.4.4 Functional Characterization of WT and *Lp*Tan-His451NmH

Kinetic testing of the purified proteins revealed binding affinity and activity of the wild-type vs the *Lp*Tan-His451NmH. The experimental results of the kinetic testing yielded for *Lp*Tan a K_m of 0.43 mM and k_{cat} of 250.5 sec^{-1} vs the *Lp*Tan-His451NmH variant which showed 1.69 mM and 0.92 sec^{-1} respectively. The results from the kinetic assays as shown, reveals that *Lp*Tan-His451NmH had substantially decreased catalysis (lower k_{cat}) as well as decreased productive binding (higher K_m) to the substrate MeG when compared to wild type tannase.

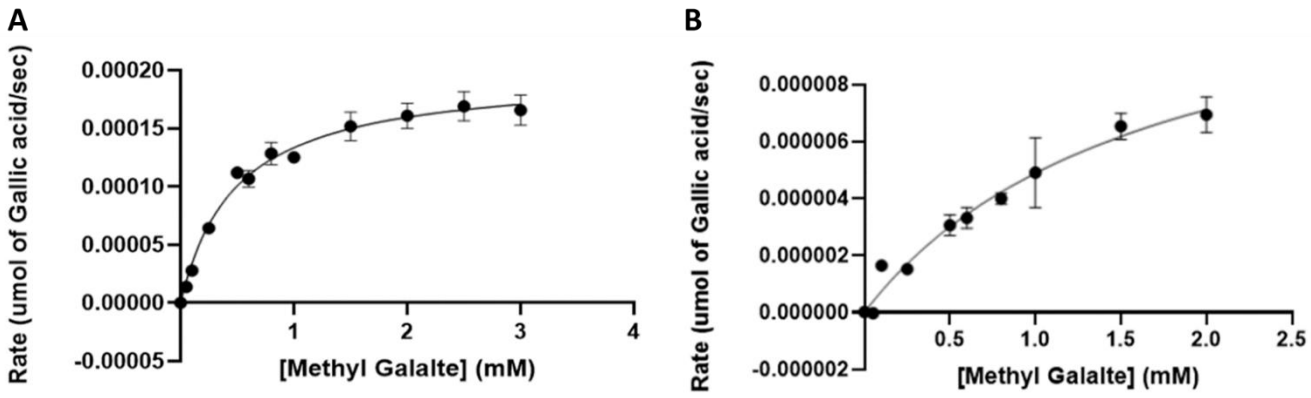


Figure 2.5: Comparison of Michaelis-Menten Plots for Recombinant *Lp*Tan and *Lp*Tan-His451NmH Proteins. *Lp*Tan (A) and *Lp*Tan-His451NmH(B). The substrate MeG was used at a concentration range of 0-3mM for the *Lp*Tan and 0-2mM for the LpTan-His451NmH.

2.4.5 Mass Spectroscopic Analysis of Recombinant *Lp*Tan and *Lp*Tan-His451NmH

Samples of the purified recombinant NmH incoporated *Lp*Tan enzyme as well as wild type were subjected to chymotryptic digestion and liquid chromatography with tandem mass spectroscopy analysis. Coverage for wild types *Lp*Tan duplicates was 50 and 57% of the amino acid sequence. The elution profile showed the expected peptide fragment (containg the targeted His451 residue) DIPHSGDY eluting at~17 min, with a mass of~903 Da when analyzed by mass spectrometry. In contrast, mass spectrometry data of the digested *Lp*Tan-His451NmH variant did not yield evidence of the presence of the His451NmH-containing peptide. The expected mass of the equivalent NmH incorporated peptide is~917 Da. At the same time, there was no evidence of the presence of the peptide containing H451 either (903 Da). Thus, the outcome is inconclusive as the expected NmH-relevant peptide could not be located in the digested *Lp*Tan-His451NmH sample.

2.5 Discussion

In this study we evaluated the effect of incorporating NmH in place of His451 in the catalytic site of *Lp*Tan. The results showed a protein approximate to the correct molecular weight and was expressed conditionally in the presence of NmH, with IPTG and arabinose inducers. Control experiments highlighted that individual omission of either of the two inducers or the NmH, reduced accumulation of this apparent *Lp*Tan-His451NmH protein. That, this accumulation is largely dependent on arabinose induction of the PYLHRS and the presence of NmH indirectly supports successful incorporation of the NmH into the protein. The observed weaker bands at 52 kDa under the conditions that included IPTG induction but lacked either arabinose (and thus AARS induction) or NmH, is consistent with *Lp*Tan being expressed in a truncated form where transcription is arrested at the amber stop codon at position 451. A truncation at this position is only 26 amino acids from the C-terminus, thus the expected difference in molecular weight between the full length NmH incorporated version and the truncated version would be approximately 3 kDa. However, the possibility that it is some residual full length wild type *Lp*Tan cannot be excluded at this time, possibly arising from a leaky promoter during the initial growth of the cell culture in complete media, or some read-through the stop codon, or even incorrect loading of the amber tRNA. It should also be noted that the evolved pEVOL-MmPyl-EVOL-HRS vector used in this study, although based on a literature construct [54], was not previously tested in our lab. Thus, the possibility that this PYLHRS was not functioning correctly cannot be strictly excluded.

Toward validation of ncAA incorporation, other studies have used mass spectrometry.[70] From the mass spectrometry data in Figure 2.6, we can see the desired peak from the fragment containing the H451 in the wild type *Lp*Tan sample, but in the *Lp*Tan-His451NmH

variant we are unable to identify this peak as seen in Figure 2.7. This means it remains unknown whether there was any actual incorporation of NmH into *Lp*Tan at position 451. Possible reasons for the disappearance of the desired peptide fragment might be the introduction of the methyl group to H451. This may have increased the hydrophobicity of the peptide causing it to aggregate out of solution or elute in a different fraction during liquid chromatography. Another paper that also incorporated NmH into their protein of interest analyzed the protein using X-ray crystallography, which could provide a definitive answer to whether incorporation was successful.[71] However obtaining a crystal structure is a non-trivial endeavor that would require increased yields and increased levels of purity.

Following purification, kinetic characteristics of the obtained putative *Lp*Tan-His451NmH were compared to *Lp*Tan. Assuming substitution was successful, the substantial decrease in activity could have occurred due to possible clashing between the methyl group of NmH451 and Asp419. Such a clash might disorient or translate the position of the His451 side chain such that it can no longer form the hydrogen bond to Ser163, and thus preclude activity.

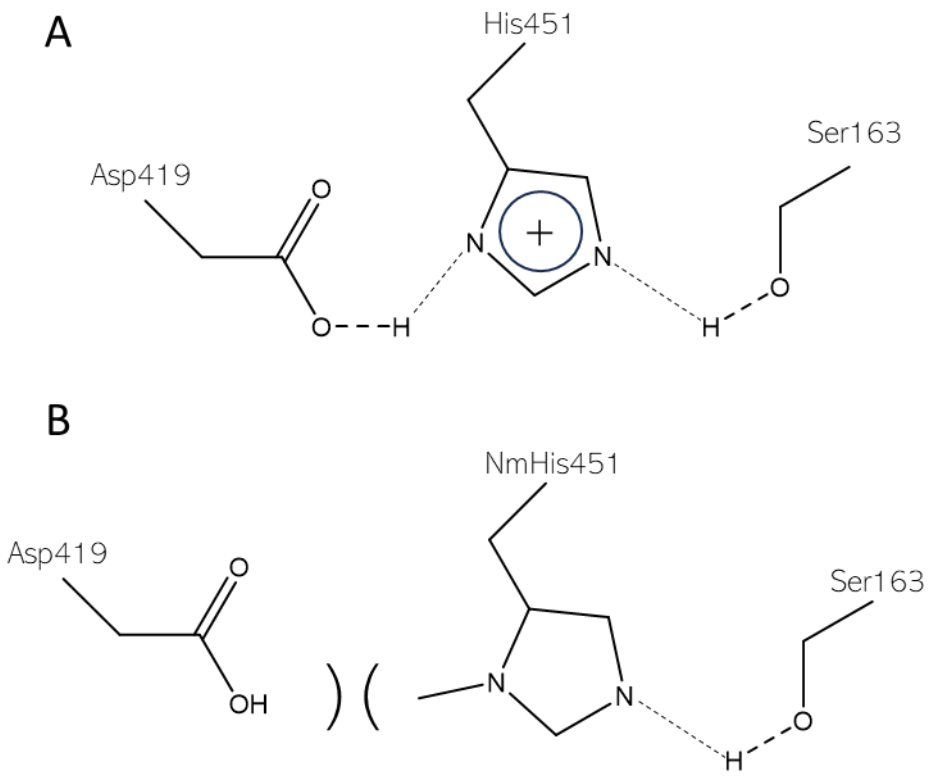


Figure 2.6: Possible Clashing Effects in the Active site of *Lp*Tan-His451NmH Between Asp419 and the Methyl Group of NmH451. *Lp*Tan (A) shows the interaction of the nitrogen lone pair and Asp419 in stabilizing His451, while in *Lp*Tan-His451NmH (B), this interaction is blocked due to interference of the introduced methyl group.

One possible way to try and address this proposed clash in *Lp*Tan-His451NmH is to create a mutation of D419A in the *Lp*Tan-His451NmH. This may eliminate any possible clashing that may have resulted from the introduction of the ncAA, due to the decrease in size of alanine. This mutation could also contribute to a hydrophobic interaction with the methyl group of the NmH451 that places the NmH451 in the correct position to interact with Ser163. Another mutation that could be explored in combination with D419A is S163C. This substitution of the nucleophile group could help to improve activity. It is possible that the addition of the methyl group causes NmH to become a weaker base, decreasing its ability to deprotonate the serine.[72-74] By making the S163C substitution, it might be better to present a nucleophile that is able to be deprotonated more easily. Looking at the pK_a values of Cys vs Ser, Cys has a pK_a value of ~ 8.3

while Ser has a pK_a value of ~13, meaning deprotonating the Cys is easier than Ser. That said, the alkoxide anion is a stronger nucleophile than the thiolate, so there are competing effects here.

That said, the possibility that the protein product detected by SDS-PAGE was only the truncated version has not been strictly excluded at this time. The truncated version lacking the H451 from the catalytic site would be completely inactive as seen in a previous study where the H451A mutation killed all catalytic activity.[39] Whether the protein is truncated or not should be easily determined by mass spectroscopy of the undigested purified species. The 3 KDa difference between full-length and the truncated species should be easily detectable by mass spectrometry.

2.6 Conclusion

In conclusion, the incorporation of NmH into the *Lp*Tan catalytic site was shown to have a significant decrease in both catalytic function and productive binding for the tested substrate MeG, compared to wild type. However, actual production of *Lp*Tan-His451NmH remains to be confirmed due to results from the mass spectrometry being inconclusive. The overall goal of creating a faster turnover rate through the conversion of a catalytic triad to a dyad through site specifically substituting NmH for His451 in the active site was not achieved.

Chapter 3: Testing a Combinatorial Multi-site Directed Mutagenesis Strategy for Improved Stability of *Lp*Tan

3.1 Abstract

Protein engineering has seen some success in increasing the efficiency of enzymes industrially, with a common strategy being to increase their stability by looking at their folded and unfolded states, and designing improvement mutation predictions that will stabilize the folded state. This design strategy cannot be applied universally and does not account for conformational changes during catalysis. Using the 'flapless (FLT)' version of tannase from *Lactobacillus plantarum* (*Lp*Tan) this study was initiated with Protein Repair One Stop Shop (PROSS),which used sequence and structural data to generate stabilizing mutation predictions for 4 different *Lp*Tan states. The obtained list of 62 mutations was reduced to 8, including only those predicted in all 4 states. These included sites Q63T, A65D, A184Y, A257D, V276Y, T321G, G421D and G439D. Combinatorial screening of the impact of these site changes on functionality has, to date, yielded 4 variants that were characterized for catalytic efficiency (k_{cat}/K_m), melting temperature (T_m), inactivation constant (k_{inact}) and total turnover numbers (TTN; k_{cat}/k_{inact}). Variant P6H7, contained 7 of the 8 site variations (all variations except V276Y) showed the highest significant k_{cat} value, (17% higher than FLT and 30% higher than *Lp*Tan) and a 4.5 °C higher T_m value, but intriguingly, showed a significant 3-fold increase in k_{inact}, reducing potential usefulness. A second variant, P8E5, with 6 of 8 site variations (all variations except A257D and G439D) yielded a 6.5 °C increase in T_m compared to FLT, but no improvements in kinetic values or TTN. The remaining variants also all had T_m greater than FLT and wild-type, but the increases were more moderate. The mechanisms underlying the stabilization imparted by these mutations, along

with the potential value of these engineered designs as biocatalysts better suited for industrial applications is discussed.

3.2 Introduction

Proteins are used in everyday functions within our bodies, performing molecular duties such as transportation, imparting structure, and catalyzing biochemical reactions. Industries are interested in possible applications of enzymes based on their ability to catalyze specific reactions, often quite efficiently with out having to input excessive energy into the reaction.[75] A common downside of using enzymes in industrial processes is that enzymes function optimally only under their biological-like conditions, decreasing their applicability under harsh industrial conditions (i.e, high or low temperatures, high or low pH). Protein engineering is an important technique when looking to increase the industrial capabilities of enzymes, modifying their designs to impart them with higher thermostability. These modifications can include optimizing secondary structure, adding salt bridges, disulfide bonds and maximizing side chain-side chain hydrogen bonding or even deleting or shortening of loop regions.[76, 77]

Previous protein engineering of *Lp*Tan, has shown that deletion of the flap domain covering the active site of the enzyme (deletion of amino acids: 225-247; Figure 3.1) imparts the enzymes with a higher k_{cat} and K_m.[78] However the thermal stability of this flapless tannase variant (FLT) was unknown, as the prior study did not evaluate this.[76,79, 80] Thus, it was hypothesized that the FLT variant might have improved stability, when compared to the naturally occurring version, as previous papers talked about the flexibility of the flap structure of *Lp*Tan, making the FLT variant a good starting point for the stabilization project.[54]

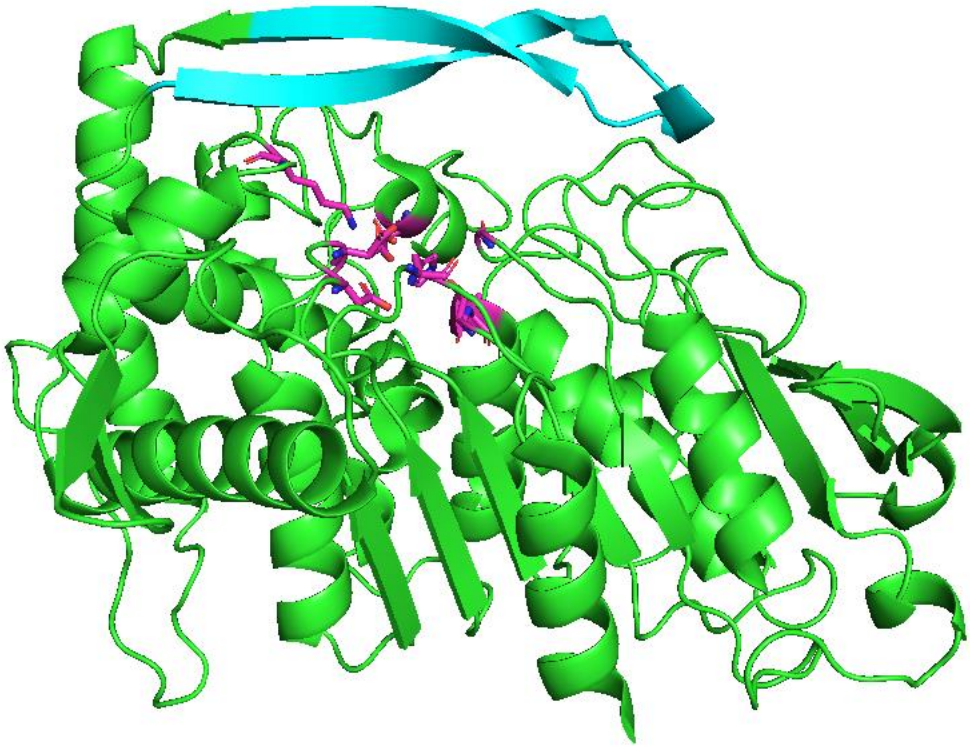


Figure 3.1: Crystal Structure of *Lp*Tan. The blue region highlights the flap that was deleted in the 'FLT' variant of this enzymes, while magenta sticks show amino acids that comprise the active site. This image depicts pdbid:3WA6 and the image was generated using PyMOL.[52]

Most protein engineering strategies have followed a single state design (SSD) concept, looking only at the functional folded state, while not taking the non-functional unfolded state into consideration. This means computational predictions based on a single fixed protein template. While there are many studies reporting proven success using SSD, it should be noted that enzymes are rarely in a "singular state".[81] States like the substrate, intermediate and product binding forms need also to be considered. Thus a 3-state design concept has been proposed that considers flexibility and the different conformations that are adopted during catalysis. By taking these forms and apo structures of the protein into consideration, we hope to create better engineered designs, that more readily and reliably improve enzyme catalytic function and/or thermostability should arise. On this basis, the software protein repair one stop shop (PROSS) was used to generate possible stabilizing site designs for the apo (pdb: 3WA6) substrate binding

(pdb: 4JUI) and product binding (pdb: 4J0H) *L. plantarum* Tannase structures. PROSS has been used successfully to predict designs that increase the thermal stability of enzymes at least once previously, with respect to *Bacillus subtilis* Lip A.

PROSS works by applying 3 steps over the course of which amino acids are selected to generate a final draft of possible stabilizing mutations.[24,82,83] The first step looks at scanning the natural sequence diversity surrounding the target amino acid sequence in genomic databases and generating a sequence alignment. This alignment allows for the elimination of undesirable mutations by identifying and excluding amino acids that are rarely or never seen in the natural diversity at any given site. The second step relies on Rosetta computational mutation scanning and independently models the mutations screened in after Step 1. Using the energies calculated for each variant's overall protein fold, in Rosetta energy units (R.e.u), selected single point changes that produce R.e.u greater than -0.45 are also eliminated. The R.e.u is used to represent how stable a mutation is with a lower R.e.u representing a more stable mutation. In the final step, PROSS uses Rosetta combinatorial sequence design to find an optimal combination of the remaining mutations using the available amino acids screened in from Step 2, again based on screening R.e.u of generated variants folds.

When looking at ways to measure whether an enzyme is more applicable to an industrial setting, three different criteria can be assessed. How the enzyme handles rise in temperature (thermostability), how fast a protein inactivates (inactivation rate) and how much product an enzyme can generate (efficiency). The thermostability of an enzyme can be judged with respect to the melting temperature (T_m), a value that represents the temperature at which 50 % of an enzyme is denatured.[84]The total turnover number (TTN) is a measurement that considers both the rate at which an enzyme inactivates or inactivation rate (k_{inact}) and the enzyme's catalytic

efficiency (k_{cat}). TTN is itself dimensionless being comprised of the expected number of reaction the enzyme can catalyze in its lifetime.[85]

Using the optimized 'flapless (FLT)' version of *Lp*Tan as a starting point, this study tests the ability of PROSS (Protein Repair One Stop Shop) to generate stabilizing mutation predictions. A set of selection criteria were used to narrow down the list of PROSS predicted site variations, which were then screened combinatorically. Detailed characterization of the most active variants from the screen, yielded variants with increased T_m and TTN comparable to wild type. The mechanisms underlying the stabilization imparted by these variants along with the potential value of these engineered designs as biocatalyst better suited for industrial applications is discussed.

3.3 Methods

3.3.1 Chemicals

All chemicals were obtained from SigmaMillipore unless otherwise stated below.

3.3.2 Construction of FLT Recombinant Expression Plasmid

The FLT variant replaces the amino acids in the flap domain of *Lp*Tan (amino acids 225-247) with a Gly-Gly-Ser linker, as was performed previously.[78]

The FLT variant was created following the protocol from New England BioLabs (NEB) for bridging double stranded DNA (dsDNA) oligos. The full length optimized *Lp*Tan coding sequence was previously cloned into the pET31b+ expression vector at the Nde1 and Xho1 restriction sites in frame with the C-terminal His-tag encoding sequence and placing the gene under the control of the T7 promoter (see Chapter 2). Linearized dsDNA of this vector was created using overlap extension PCR on a T100 thermo cycler (BioRad), using primer FFLT and

RFLT (Table 3.1), and with 5 µL 5X Q5 Reaction Buffer, 0.5 µL of 10 µM dNTPs, 1.25 µL each of 10 µM primers FFLT and RFLT, 0.38 µL of 261.9 ng/µL pET31B+-*Lp*Tan template DNA, 0.25 µL of Q5 High-Fidelity DNA polymerase and the remaining volume was made up with nuclease free water to create a final volume of 25 µL. The PCR reaction conditions included, an initial heating to 98 ℃ for denaturation of dsDNA for 30 seconds, after which the sample was subjected to 35 cycles of 98 ℃ for 10 s, 65 ℃ for 30 s and 72℃ for 3 min. After the cycles were completed the reaction was held at 72 ℃ for an additional 2 min for the final extension.

The final FLT expression construct was created by combining 1pmol of ssDNA oligo encoding the Gly-Gly-Ser substitution, 0.005 pmol of linearized dsDNA vector obtained by PCR as described above, 10 µL of NEBuilder HiFi DNA Assembly Master Mix 2X and topped up with nuclease-free water to a final volume of 20 µL. The reaction was incubated in a T100 thermo cycler (BioRad) at 50 ℃ for 1 h.

Table 3.4: Primers used in FLT Creation

FLT Forward primer (FFLT)	5’ TCTGGCCAGCTGACTGTG 3’
FLT Reverse primer (RFLT)	5’ GTTGATACCGTTGAACTGCC 3’
ssDNA oligo	5’ gcgtacgaatggcagttcaacggtatcaacGGAGGATCCtctggccagctgactgtggaagaacaggcg 3’

The FLT expression construct was transformed into competent NEBuilder *E. coli* cells to take advantage of their increased transformation efficiency compared to BL21(DE3) Gold *E. coli* cells (Agilent). Competent NEBuilder cells were thawed on ice for 10 min, after which 100 ng of FLT expression construct were added and placed on ice for more than 30 min (a negative control with NEBuilder cells-only was included). Cell samples were heat shocked at 42 ℃ for 10 secs

then placed back on ice for 5 more min, after which 450 µL of super optimal broth (5 g yeast extract, 20 g tryptone, 0.5 g NaCl, 0.186 g KCl, 2.4 g anhydrous $MgSO_4$ 1 L H_2O) was added to the cells which were then incubated at 37 °C for 1 hour at 250 rpm. Finally, 50 µL of each transformed cell sample was plated on LB agar plates (7.5 g agar, 5 g tryptone, 2.5 g yeast extract and 5 g NaCl for 500 mL H_2O), supplemented with 100 ug/mL of ampicillin and incubated overnight at 37 °C. The next day, 10 mL of LB liquid media supplemented with 100 ug/mL of ampicillin was inoculated with a single colony from the agar plate and the culture grown overnight in a MaxQ6000TM incubator at 37 °C, shaking at 250 rpm. From this 10 mL culture, pET31b+-FLT plasmid was extracted using plasmid DNA minipreps kit (BIOBasic) and then the obtained plasmid transformed again following the same protocol as stated earlier into competent BL21(DE3) Gold cells for recombinant expression.

3.3.3 Recombinant Production and Purification of LpTan and FLT

LB media (10 mL) supplemented with 100 ug/mL of ampicillin was inoculated with BL21(DE3) Gold cells transformed with either the pET31b+-FLT or the pET31b+-*Lp*Tan expression constructs and grown overnight in a MaxQ6000TM incubator at 37 °C with shaking at 250 rpm. This starter cultures were used to inoculated 100 mL of LB media supplemented with 100 ug/mL ampicillin and incubated at 37 °C at 150 rpm until OD600nm reached 0.3-0.4, at which time the temperature was reduced to 20.5 °C. The cultures were then grown to an OD600nm of 0.6-0.7, at which time a final concentration of 0.7 mM IPTG was added to induce protein expression. The culture was then left to grow overnight at 20.5°C shaking at 150 rpm.

The overnight culture was centrifuged at 3250 x *g* for 30 min at 4 °C. The obtained supernatant was discarded, and the pellets were placed on ice. Extraction buffer (4 mL) with protease inhibitors (0.1mM PMSF, 3U/mL benzonase, 1mg/mL lysozyme, 5mM DTT, 20mM

$MgCl_2$, 10 mL lysis buffer) was used to re-suspend the pellet. The cells were lysed on ice by sonication for 30 cycles, for 25 seconds at 30% duty and 40 output power and 30 second cooldown per cycle using a Sonics Vibra cell VC 505 sonicator. Lysates were then centrifuged at 17000 x g for 30 min at 4°C to remove cell debris and any un-lysed cells.

QIAexpressionist™ Ni-NTA affinity resin from Qiagen was used to purify the expressed FLT and *Lp*Tan proteins. Qiagen Ni-NTA resin (1 mL) was equilibrated in 10 mL pH 8.0 lysis buffer (10mM Tris-HCl, 150mM NaCl, 10mM imidazole) in a column. All collection tubes were placed on ice during and following elution. The cell lysate supernatant was loaded onto the column and flowthrough was collected. The loaded column was washed three times with 1mL of pH 8.0 wash buffer (10mM Tris-HCl, 150mM NaCl, 20mM imidazole), each time collecting the flowthrough into separate tubes. The bound protein was then eluted with the addition of 1 mL of pH 8.0 Elution Buffer (10mM Tris-HCl, 150mM NaCl, 250mM imidazole) with incubation on ice for 4 min before the flowthrough was collected and the elution process repeated two more times. SDS-PAGE was used to confirm successful Ni-NTA purification.

Elutions were pooled and concentrated using an Amicon µLtra-4mL 30k MWCO concentrator and centrifuged at 3250 x *g* for 10 min. This process was repeated until only 1-1.5 mL of eluate was remaining. The obtained purified *Lp*Tan and FLT samples were then subjected to size exclusion FPLC and had their concentrations determined by BCA assay all as described in Chapter 2.

3.3.4 Michaelis-Menten Enzyme Kinetic Assays

Activity of all enzymes described below was assayed using the same technique as described for *Lp*Tan in Chapter 2.

3.3.5 PROSSS-Directed Mutation Generation

Predicted mutations were generated by submitting 3 crystal structures of *Lp*Tan and a Phyre2 derived model of FLT into the PROSS algorithm.[86] The crystal structures included PDB IDs: 3WA6 (apo), 4J0K (substrate: Ethyl gallate) & 4J0H (product: Gallic acid). A multiple-sequence alignment was generated by PROSS using the National Center for Biotechnology Information database with a BlastP E value of 0.0001, a minimal sequence identity of 35 %, 75 % coverage and a maximum of 4000 targets. The energy function ref2015 was selected and amino acids corresponding to the active site (G77, A164, S163, H451, D419, D421, E357& K343) were constrained.

3.3.6 Construction of an FLT-based Combinatorial Library

Generation of the combinatorial library vectors covering the 8 single site variants (Q63T, A65T, A184Y, A257D, V276Y, T321G, G421D and G439D) that were ultimately screened in, using pET31b+-FLT as the template was performed by GenScript. Competent BL21-(DE3) GOLD cells were thawed on ice for 10 min, after which 100 ng of combinational library plasmid DNA was added and transformations conducted as described above. To ensure full coverage of library 768 colonies was picked and screened where 768 is the result of the 256 possible mutant combinations arising from the eight possible sites, multiplied by 3.[87]

Multiple 96 well Nunc plates (mother plates) were autoclaved and filled with 300 µL of LB supplemented with 100 µg/mL ampicillin. Each well was inoculated with 1 cell colony from the plated combinatorial library as well as a colony of the original BL21-DE3-Gold expression pET31b+-FLT. The Nunc plates were incubated overnight at 37 °C and shaking at 250 rpm in a MaxQ6000 incubator from Thermo Scientific. Overnight Instant Express TB Media from Novagen was prepared and supplemented with 100 µg/mL ampicillin. New Nunc plates

(daughter plates) were filled with 300 µL of Overnight express TB media. A sterile replicator was used to inoculate the daughter plate from each mother plate. Daughter plates were then covered with the breathable rayon membrane (VWR) and then placed into the MaxQ6000 incubator from Thermo Scientific at 37 °C and 250 rpm overnight as well. Mother plates were supplemented with 50 µL of 50 % glycerol and 1 % NaCl, covered with a foil layer (VWR Product Number 89049-034 in North America) and stored at -80 °C.

Daughter plates were then centrifuged at 3000 x *g* at 4 °C for 30 min in an Allegra X-12 R Centrifuge from Beckman Coulter. The supernatant was then removed, and cells were then washed twice with 200 µL of Phosphate Buffer Saline (PBS). Each wash involved completely resuspending the cells in PBS using an Integra VIAFLO96 96 well pipette with the mix setting set at 100 µL, and centrifuging the cells down using the same centrifuge conditions as above. The pellets were then resuspended in 100 µL of Lysis Buffer (100 mM reaction buffer with 10X Bug buster (Novagen), 5 U/mL Benzonase nuclease and 1 mg/mL lysozyme). This was done using the 96-pipette mix setting at 50 µL. The plate was then centrifuged again with the same conditions. All 100 µL of the cell lysate supernatants were then collected and transferred to new Nunc plates.

3.3.7 High-throughput Stability Screening Assay

High-throughput assays were performed using the gallate-rhodanine assay as described in Chapter 2, using 1 mM substrate and 10 µL of 100x dilution of the mutant enzyme extracts from cell lysates in 50 mM Tris Buffer pH 8.0. This assay was used to quantify the activity and stability of mutants compared to *Lp*Tan. Using 2 separate 96 well 2 mL plates, 2 different reactions were performed at 30°C. One plate (plate A) containing enzyme, was not incubated before the addition of substrate and followed the normal gallate rhodanine procedure. The other

plate (plate B) containing enzyme, was incubated for 2 hours at 30 °C before adding substrate and then rhodanine. The activity deference between these two plates were compared yielding a ratio (amount of activity remaining after 2-hour preincubation/initial activity before incubation) to obtain the decrease in activity of the enzyme upon being held at 30 °C for 2 hours.

3.3.8 Enzyme *In*activation Kinetic Assay and Calculation of TTN

Selected variants were recombinantly produced and purified using the same protocol as described for *Lp*Tan and FLT. TTN was determined by taking the ratio of the k_{cat} over k_{inact}. The gallic acid - rhodanine assay as described in chapter 2 was applied to each selected enzyme variant at 5 different enzyme pre-incubation times (over 0-85 min) at 30 °C to allow denaturation using ~0.025µg of enzyme and MeG concentration of 1 mM. The ln of the detected specific activity was plotted against incubation time with the slope representing the rate of inactivation (k_{inact}). The half-life ($t_{1/2}$) was determined using the first order kinetics equation from the k_{inact} (equation 2).[85] k_{cat} was determined from previous kinetic assay ran at the same temperature.

3.3.9 Enzyme Melting Temperature (T_m) Assay

Selected variants were recombinantly produced and purified using the same protocol as described for *Lp*Tan and FLT.[88] Enzymes were diluted to 1 mg/mL and 10 µL aliquots were transferred into a 96 well PCR plate from ThermoFisher. Adding in 2.5 µL of a 50X stock solution of SYPRO Orange (generated by diluting 2.5µL of a 5000X stock into 250µL of MQ water) and 12.5 µL of size exclusion buffer. Negative controls included one sample that excluded protein and another that had no dye. Samples were placed in an iQ5 qPCR machine (BioRad). Melt curve function started with the temperature at 20 °C and went up in intervals of 0.5°C, ending at 95°C, with 15 seconds spent at each temperature interval. Software was analyzed using excel.

3.4 Results

3.4.1 Contributors

Combinatorial library of plasmids for FLT variants was generated by Genscript

3.4.2 Characterization of Wild-Type *Lp*Tan and FLT

FLT and *Lp*Tan wild type proteins were produced recombinantly and purified by Ni-NTA affinity and SEC. They were characterized in terms of kinetic abilities, inactivation rate and T_m to determine if FLT was more thermostable and longer lived and thus more suited for industrial applications than the wild-type. A kinetic Michaelis-Menten analysis (Figure 3.2A) revealed that both of the *Lp*Tan and FLT k_{cat} values are substantially larger (~ 2 times) than literature values (Table 3.2), which could be due to subtle differences in recombinant production, details, purification methods and kinetic assay techniques.

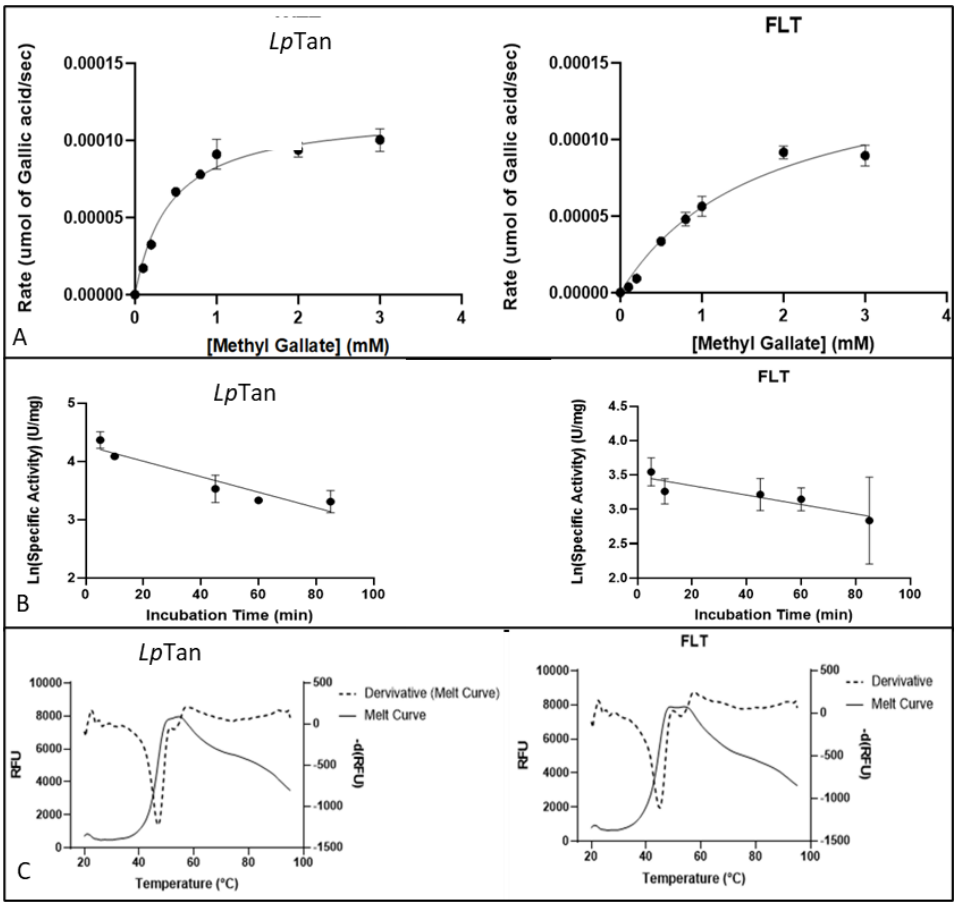


Figure 3.2: Comparison of Kinetic and Stability Plots Between *Lp*Tan (Left) and FLT (Right). A) Michaelis-Menten plots, B) k_{inact} plot C) Melting Temperature (T_m) analysis. Plots were analyzed using GraphPad Prism.

Table 3.5: Comparison of Characterization Parameters Between *Lp*Tan and FLT. Error range shows upper and lower standard deviation (n=3). Statistical significance was assessed by student T-test comparing empirical FLT to empirical *Lp*Tan and are represented by an * where $p<0.05$.

	FLT		*Lp*Tan	
k_{cat} (sec^{-1})	Empirical	Literature[66]	Empirical	Literature[66]
	306.6 ± 18.4*	115.5 ± 16.6	250.5 ± 18.1	99.4 ± 11.4
K_m (mM)	1.7±0.1*	1.1±0.4	0.43 ± 0.02	0.67 ± 0.15
k_{cat}/K_m ($sec^{-1}mM^{-1}$)	180.4 ± 12.2	95.3	582.6 ± 52.0	158.4
k_{inact} (min^{-1})	0.0069 ± 0.0052	N/A	0.0133 ± 0.0023	N/A
Half-Life (min)	101.2	N/A	52.1	N/A
TTN	2.69e06	N/A	1.13e06	N/A
Tm (°C)	45.0 ± 0.0*	N/A	46.8 ± 0.3	N/A

When comparing the empirical values obtained for FLT and *Lp*Tan to each other, it is shown that the FLT variant has a significantly larger k_{cat} than that of the *Lp*Tan, but a decrease in productive binding for the substrate MeG. It should be noted though, that the average rate of inactivation (k_{inact}) for FLT (Figure 3.2B and Table 3.2) was found to be reduced, suggesting that the rate at which the enzyme inactivates may be slower. However, a statistical analysis highlighted that the difference was not significant, thus, being that TTN was determined from k_{inact} values TTN results are undermined. As such although the obtained values suggest that FLT might have a larger TTN value being approximately twice as large as *Lp*Tan, this result can not be trusted. At the same time, analysis of T_m shows a significant reduction in FLT when compared to the wild-type *Lp*Tan, suggesting that wild-type is more stable.

3.4.3 PROSS Generated Mutations

PROSS generated 62 possible mutations across all three binding forms of the wild-type *Lp*Tan enzyme and the phyre 2 model FLT variant. By only focusing on mutations shared between all 3 binding forms, the 62 total mutations sites were narrowed down to 12 sites. This list of 12 mutations was further narrowed down to 8 after comparing the results to the FLT Phyre 2 model, and only including mutations shared between the FLT model and the pre-existing 12 mutations, with 4 of the 12 sites not being predicted to stabilize the structure of the FLT model. These 8 mutations are at positions 63, 65, 184, 257, 276, 321, 421, 439 in the FLT model. These sites are scattered throughout the FLT mutant with 3 being in the conserved ABHD, while the other 5 are in the cap region (Figure 3.3). Most mutations consistently vary residues from polar to non-polar and vice versa as follows Q63T, A65D, A184Y, A257D, V276Y, T321G, A421D, A439D (Figure 3.4). It should be added that a Phyre 2 model was used to create FLT model for PROSS. Phyre 2 uses homology detection methods to build 3D structures, as such limitations of would

include if homology can not be detected which would make modelling impossible or its inability to predict the structural effects of a point mutation.[89]

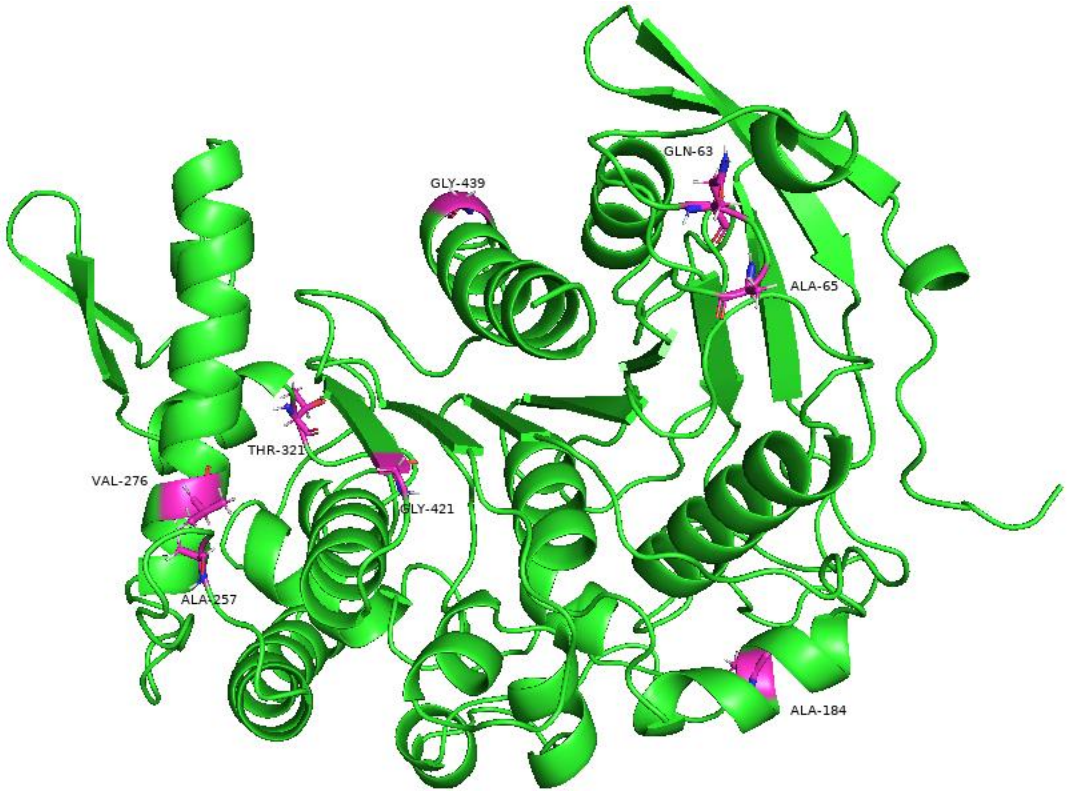


Figure 3.3: Generated FLT Structure (Green) with Predicted PROSS Mutation Sites (magenta). FLT structure was created using AlphaFold and modelled using PyMol. PROSS was used for mutation prediction.[52, 86]

3.4.4 Combinatorial Screening of the PROSS Predicted Mutations

The combinatorial library covering all 256 mutants showed some possible hits (Figure 3.5) that were both more active (> 1.0 Activity) and showed a decrease in rate of inactivation (k_{inact}) (>1.0 Residual Activity). A selection of these potential hits were purified, characterized and compared against FLT.

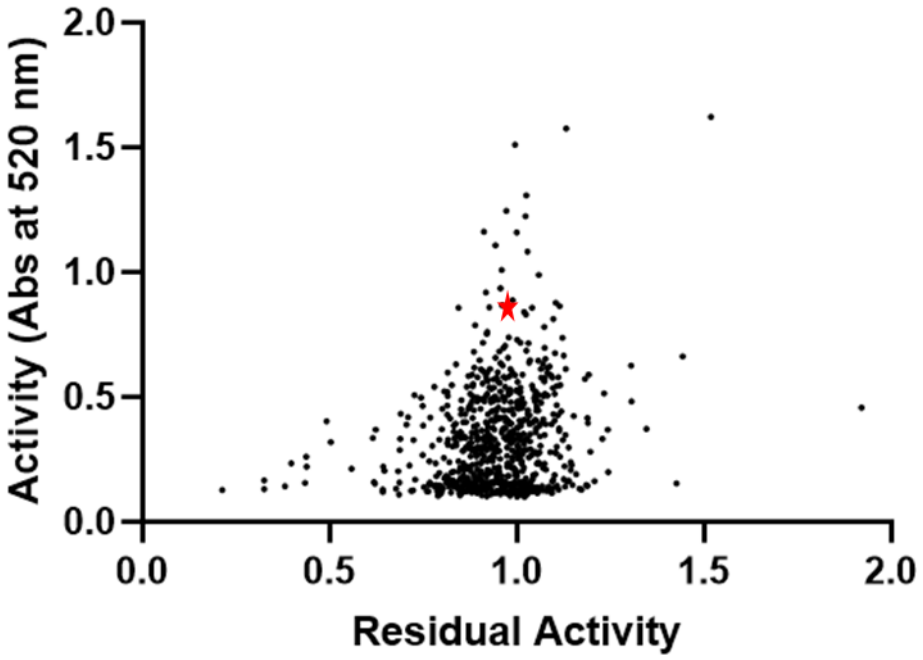


Figure 3.4: Comparison of Activity and Residual Activity (stability) of the 768 Combinatorial Library Mutants against FLT. Combinatorial variants are shown in black, while FLT is shown as a red star as a reference. Residual activity was measured by taking the ratio between the activity detected with no pre-incubation of the enzymes and the activity detected after 2-hours pre-incubation of the enzyme at 30 °C. Activity was the measured absorbance at 520 nm upon no preincubation of the enzymes. Plot was created using GraphPad Prism.

With respect to the objective of obtaining a more stable and more efficient variant of FLT a collection of 4 variants with activity and/or residual activity equal to, or great than, FLT were considered further (Figure 3.6). DNA sequencing of these showed that they all shared the mutations Q63T, A65D, A184Y, with each combined differently with other mutations as highlighted in Table 3.3 and Figure 3.7. None of the 4 selected variants contained all 8 mutations.

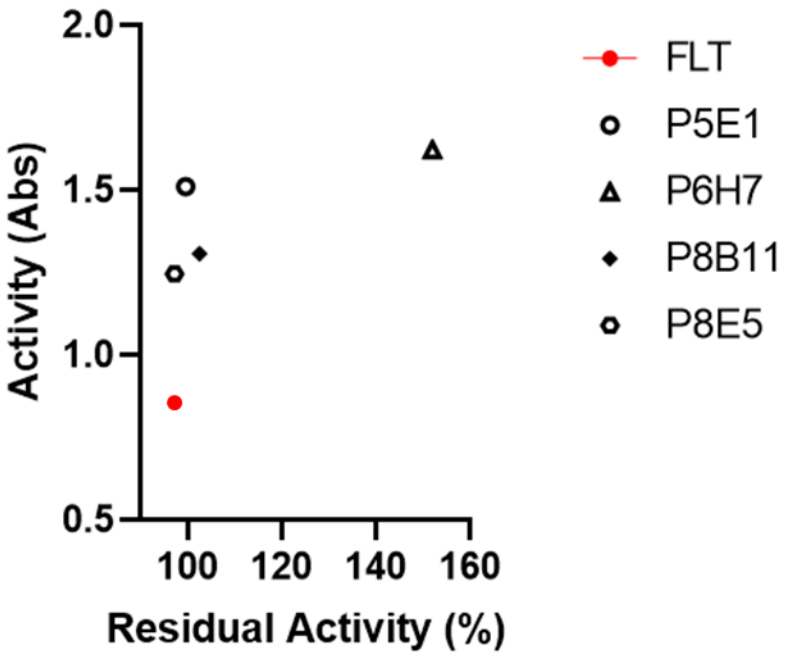


Figure 3.5: Comparison of Activity and Residual Activity (stability) of the 4 Selected Mutants against FLT. Mutants are shown in black, while FLT in red was used as a reference. Residual activity was measured by taking the ratio between the activity detected with no pre-incubation of the enzymes and the activity detected after 2-hours pre-incubation of the enzyme at 30 °C. Activity was the measured absorbance at 520 nm upon no preincubation of the enzymes. Plot was created using GraphPad Prism.

Table 3.6: Mutations made in the 4 Selected Mutants from the Combinatorial Library

	Q63T	A65D	A184Y	A257D	V276Y	T321G	G421D	G439D
P5E1	Yes	Yes	Yes	Yes	Yes	No	Yes	No
P6H7	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
P8B11	Yes	Yes	Yes	No	No	Yes	No	No
P8E5	Yes	Yes	Yes	No	Yes	Yes	Yes	No

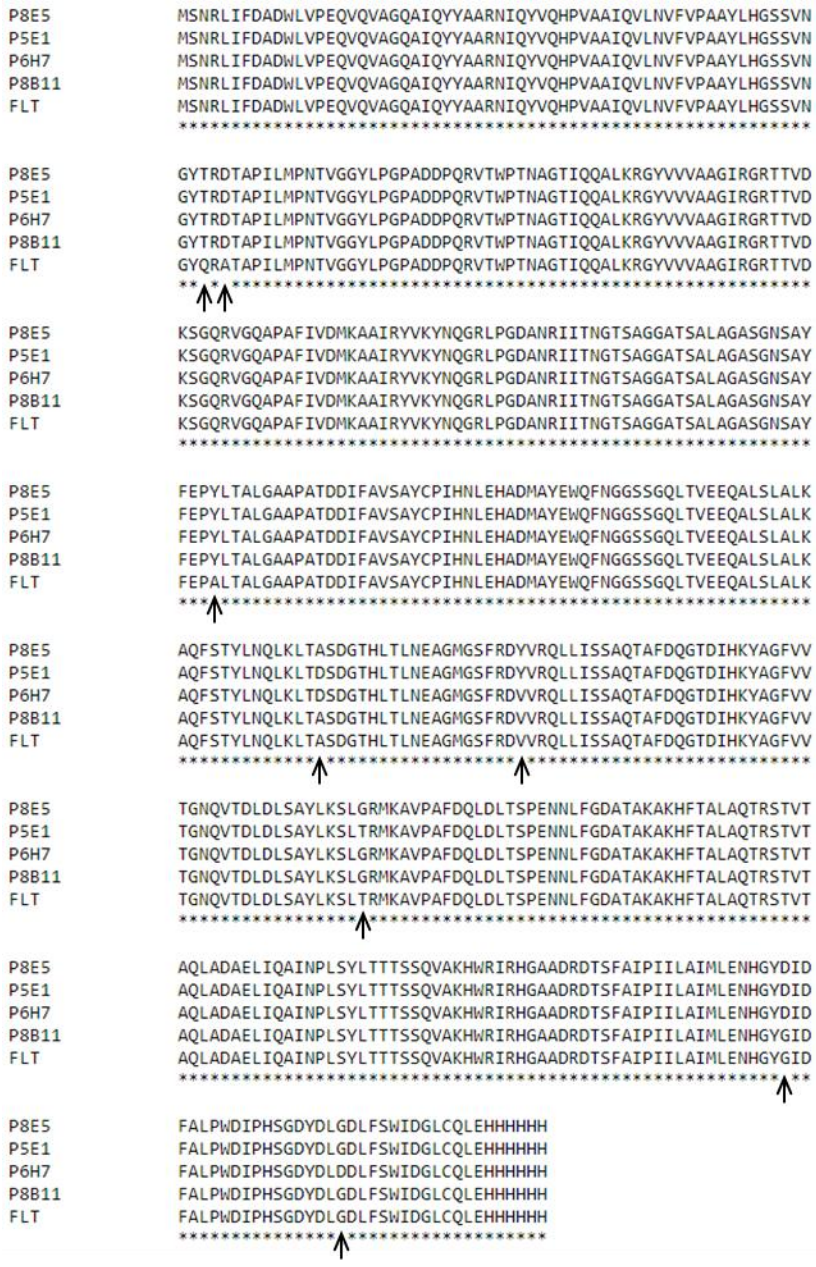


Figure 3.6: Multiple Sequence Alignment of the 5 selected Mutants against FLT. Black arrows indicate mutagenesis sites 63, 65, 184, 257, 276, 321, 421, 439. MSA was performed using CLUSTALW.[90]

3.4.5 Characterization of the Selected 4 Top Combinatorial Variant Hits

The 5 selected variants and FLT were recombinantly produced and purified by Ni-NTA affinity and SEC. Initial Michaelis-Menten kinetic analyses (Figure 3.8; Table 3.4) showed that P5E1,

P6H7 and P8B11displayed larger k_{cat} than FLT. Interestingly, the K_m values for P5E1, P6H7, and P8E5 variants were comparable to FLT, although only P6H7 was shown to yield a statistically significant increase. Interestingly the K_m values for P5E1, P6H7 and P8E5 were all comparable to FLT, with P8B11being an exception, showing a significantly higher K_m. Overall, the variants catalytic efficiencies were variable, with only the P5E1 showing any hint of improvement and its' supporting data being significantly different from FLT. Notably P6H7 showed an efficiency comparable to FLT, despite its significantly larger k_{cat}. The other two had overall efficiencies less than FLT, consistent with the significant increase in K_m at least in the case of P8B11.

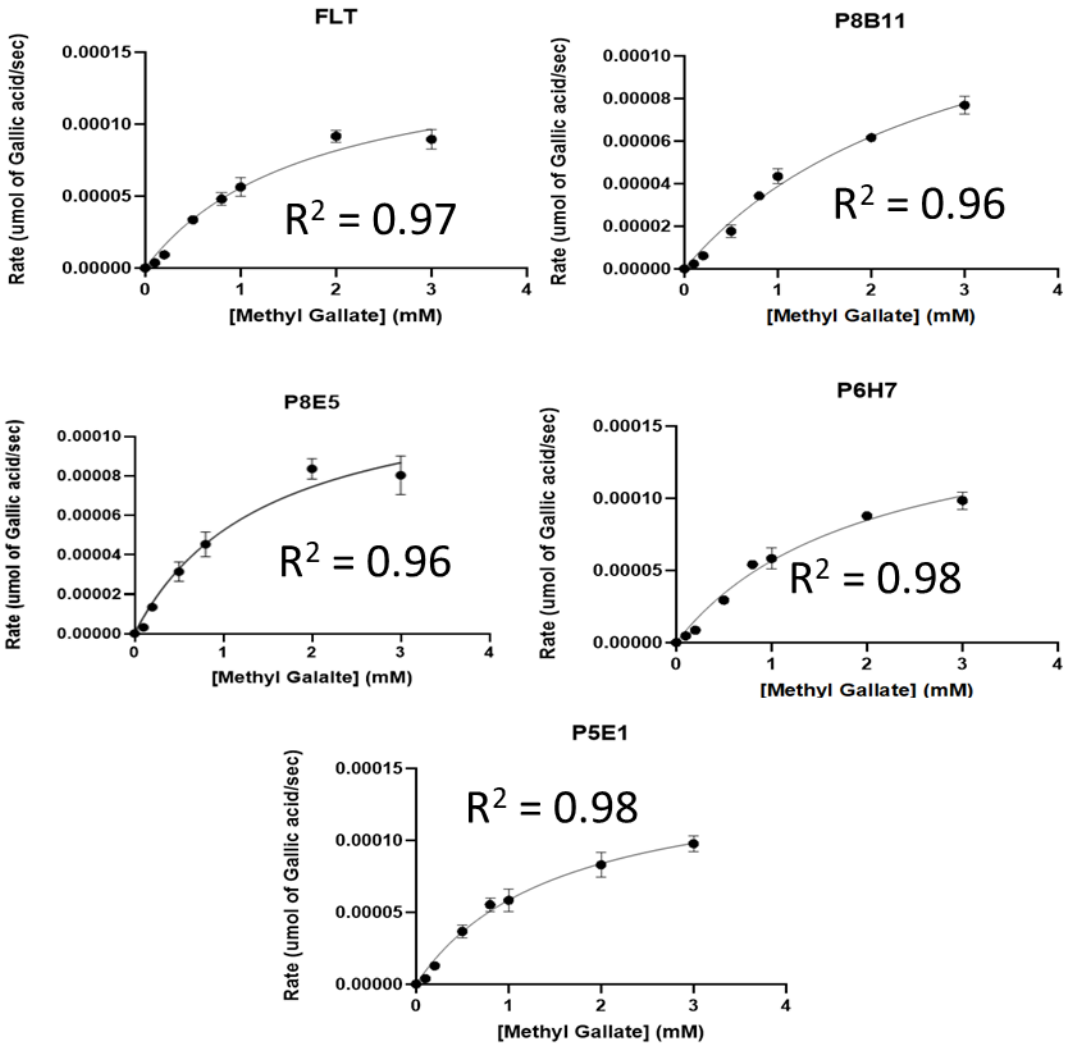


Figure 3.7: Comparison of Michaelis-Menten Plots of FLT and Selected Variants. Curves were generated using GraphPad Prism software and performed in triplicate. The substrate MeG was used in a concentration curve of 0-3 mM and the formation of a rhodanine chromagen formed by reaction with the gallic acid product monitored.

Table 3.7: Comparison Table of Kinetic and Stability Parameters and Number of Mutations of FLT against the 5 Selected Variants. Error bars show the standard deviation (n=3). Statistical significances were assessed by student T-test compared against FLT and are represented by an * while a ˣ shows statistical significance for FLT variants against wild-type, where $p<0.05$.

	FLT	P5E1	P6H7	P8B11	P8E5
k_{cat} (sec^{-1})	306.6 ± 18.4	345.0 ± 19.2ˣ	365.2 ± 37.2*ˣ	327.4 ± 18.6ˣ	266.1 ± 39.2
K_m (mM)	1.7±0.1	1.5±0.2ˣ	2.0±0.3ˣ	3.0±0.3*ˣ	1.5±0.4ˣ
k_{cat}/K_m (sec^{-1}•mM^{-1})	180.4 ± 12.2	230 ± 35.6	182.5 ± 33.8	109.1 ± 12.8	139.1 ± 48.9
K_{inact}(min^{-1})	0.0069 ± 0.0052	0.0136 ± 0.0052	0.0205 ± 0.0011*ˣ	0.0157 ± 0.0030	0.0139 ± 0.0050
Half-Life (min)	101.2	51.1	33.8	44.0	49.9
TTN	2.69e06	1.53e06	1.07e06	1.25e06	1.32e06
T_m (°C)	45.0±0.0	48.5±0.3*ˣ	49.6±0.3*ˣ	47.3±0.3*	51.5±0.0*ˣ
# Mutations	0	6	7	4	6

Next the inactivation rate k_{inact} of the selected mutants was considered (Figure 3.9; Table 3.4). In contrast to expectations, all variants showed at least a slight increase in inactivation rate, meaning catalysis slowed down faster over time for the variants compared to the FLT, Notably, only that of P6H7 was deemed a significant decrease. This increase in k_{inact} is reflected in the TTN having a substantially lower value obtained than FLT (Table 3.4). This is surprising in that the P6H7 variant that showed a 160 % increase in residual activity in the high-throughput assay (Figure 3.5). However, the bell-shaped curve centered on 1 in x-axis of the screening experiment (Figure 3.4) suggests there may have been significant error in residual activity measurements (potentially arising from variation in expression levels). At the same time there is small lack of concordance between the different variant inactivation initial activities (Figure 3.8), which may also reflect other systemic sources of error.

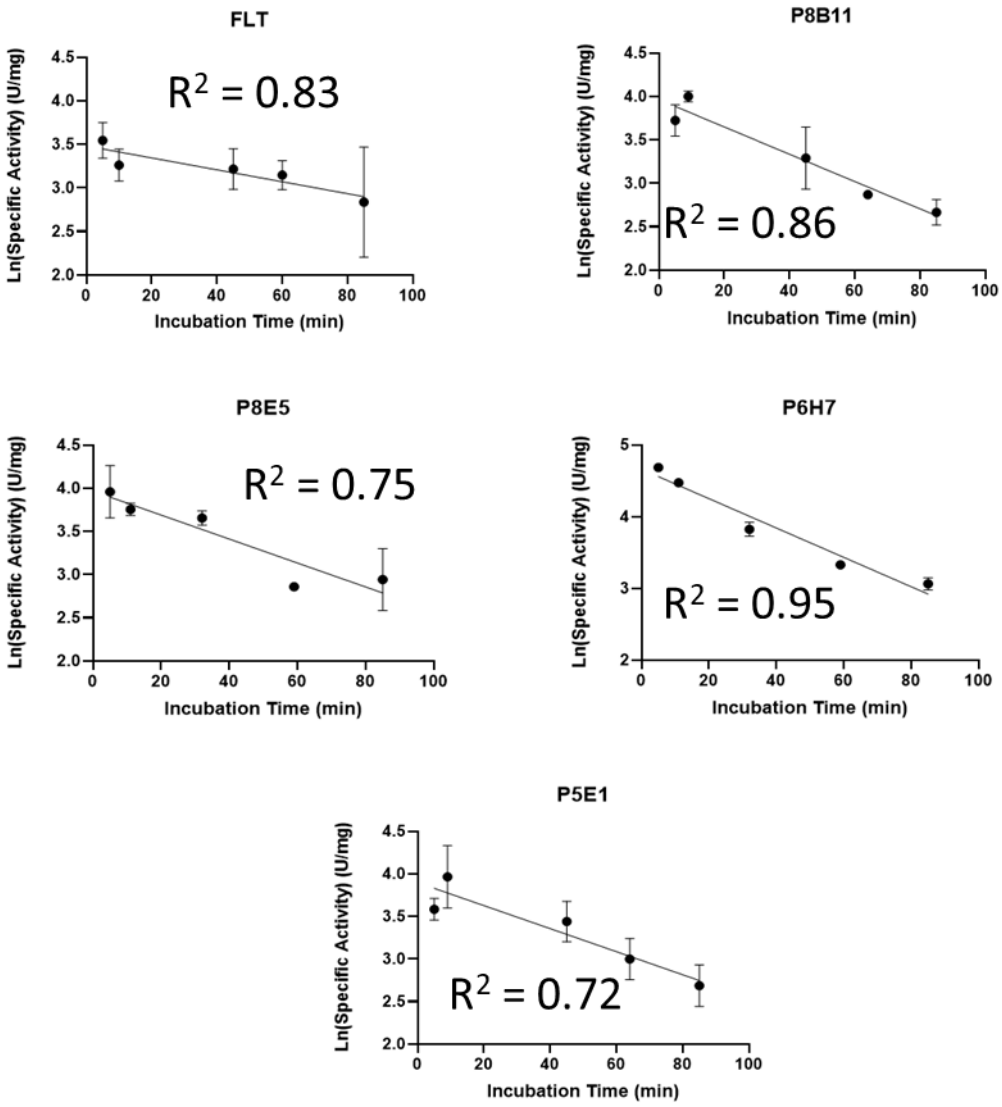


Figure 3.8: Comparison of Inactivation Rates Plots of FLT and Selected Variants. Plots were generated using GraphPad Prism software and performed in triplicate. Reactions were carried out at 30 °C.

Toward trying to glean more insight into the increased inactivation rates, Tm values were subsequently obtained (Figure 3.9; Table 3.4). Interestingly, and in contrast to the inactivation results, all FLT mutants displayed T_m values significantly higher than FLT. Notably this included P8E5 having a substantial and significant increase of 6.5 °C. P6H7 also showed significant increases in Tm of 4.6 °C. Even P5E1 and P8B11 showed significant increases of 3.5 and 2.7 °C respectively.

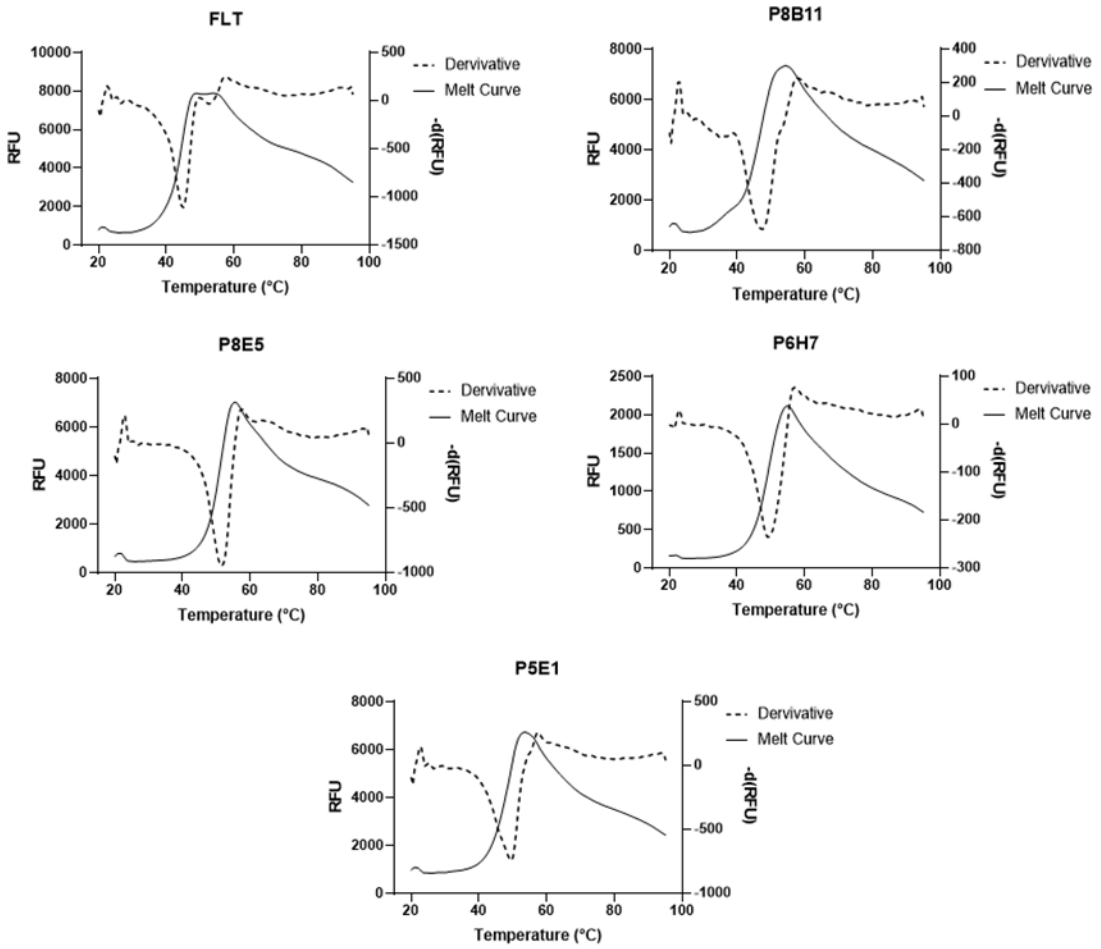


Figure 3.9: Melting Temperature Plots for FLT and Selected Variants. Experiments were performed in triplicate, with original absorbance being represented by the solid line (left y-axis) and its first derivative ($\frac{dT}{dF}$) represented by dashed lines (right y-axis). Data was plotted using GraphPad Prism.

3.5 Discussion

In terms of the underlying mechanisms that are leading to the observed changes in kinetics and inactivation, the "flap" is said to function as a gate for substrate (Figure 3.10). We see a significantly higher K_m value for FLT compared to wild-type, which is consistent with the idea that the flap is used to mediate/coordinate substrate binding in the active site of the enzyme. Removal of the flap eliminates residues that are predicted to make direct interactions with the substrate in the binding pocket, such that the overall affinity of the active site for the substrate is

diminished and competing with nearby non-productive binding events. From this, we might expect to see an increase in FLT k_{cat} (consistent with observations and literature (Table 3.2)) as there is nothing to "regulate" the flow of substrate to the active site, possibly increasing the total amount of substrate to be catalyzed by the FLT variant. Nonetheless the overall efficiency (k_{cat}/K_m) is significantly lower for FLT. The fact that FLT might have a smaller k_{inact} suggesting a higher stability than wild type overtime, is undermined by lack of significant differences, as well as the observed significant decrease in the FLT T_m value. Overall, the results obtained to date do not support FLT being an improved variant generally and further testing is recommended to get more accurate and reliable results. If the inverse correlation between k_{inact} and T_m were validated in the future, the mechanism underlying this discrepancy between inactivation and T_m could be further investigated by studying unfolding intermediates and aggregation. As well, with the idea that deletion of the loop was not constructive, alternatively, the effects of strengthening interactions between the catalytic region and the flap might be worth investigating. Nonetheless, FLT was selected as the template to use as the starting point for development of a more stable enzyme based on its significantly increased k_{cat}.

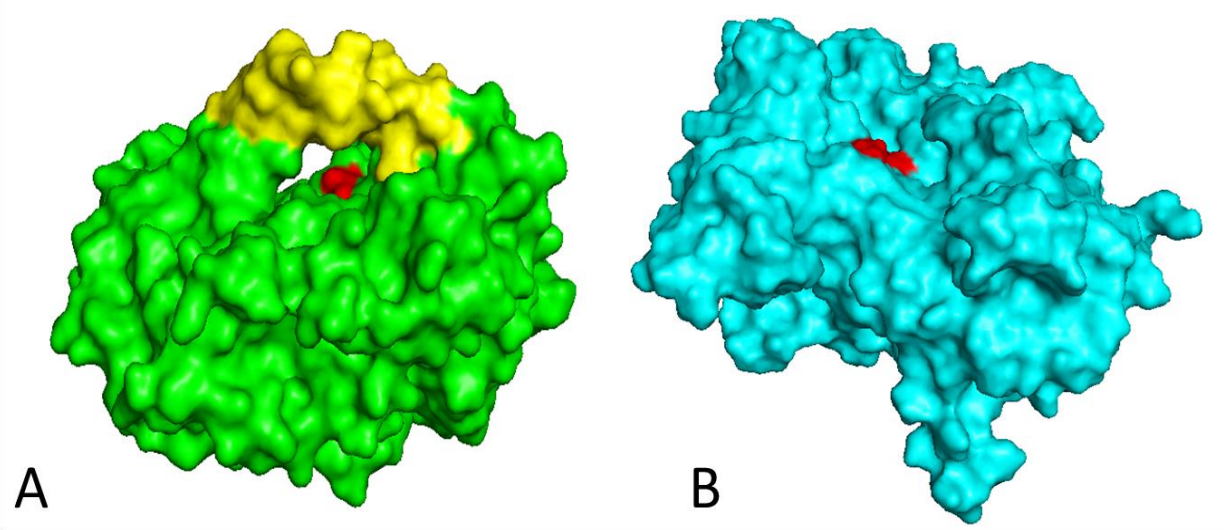


Figure 3.10: Structure Comparison of LpTan and FLT. A) shows the *Lp*Tan with the "flap" region highlighted in yellow, while B) shows FLT variant lacking the flap. Both figure A and B show location of active site in red. PyMol was used to model proteins with pdbid: 3WA6 being used in figure A. FLT was generated using AlphaFold. The Pymol Molecular Graphics System, Version 2.0 Schrodinger, LLC.[52, 91, 92]

A high throughput assay was utilized to detect key mutants from the combinatorial library screen, from which 4 variants were selected for full characterization. It should be noted that in the high throughput assay there were mutants that displayed residual activities above 1 (100%). This perceived error may be due to the differences in protein expression levels between each of the wells (variant-dependents).

Interestingly, k_{inact} and T_m were both shown to be increased in all 4 variants when compared to FLT. Variants displayed an opposite inverse correlation between T_m and k_{inact} compared to FLT is somewhat perplexing, although only the k_{inact} of P6H7 was deemed significantly different. With a focus on P6H7, which contained 7 of the 8 sites variations (all except V276Y), it is notable that it showed the (significantly) highest k_{cat},17 % higher than FLT and 30 % higher than wild-type and a significant T_m value of 4.5 ˚C higher than FLT (the latter

not significantly different than wild type). However, it would seem then that while this variant is more catalytically active and thermally stable than the FLT, this comes with the cost of faster inactivation over time (increased k_{inact}). As mentioned above with respect to FLT, it is difficult to explain this inverse correlation between k_{inact} and T_m logically, and thus additional experiments are recommended to evaluate unfolding intermediates and aggregation to glean insight into the underlying mechanism. Nonetheless, the increased k_{inact} values are still comparable to the wild-type *Lp*Tan, and thus this catalytically (k_{cat}) more active variant may still maintain industrial relevance compared to wild type for short fast reactions. Another variant, P8E5, with 6 of the 8 mutations (all variations except A257D and G439D) yielded a 6.5 °C increase in T_m compared to FLT (5.0 °C increase compared to *Lp*Tan), and no other significant changes to its characteristics, and thus could be useful industrially for processes at high temperatures.. Overall, with observed standard deviations being so high across the collected data, it is recommended that all experiments be repeated with more replicates to clarify and verify findings.

Another interesting observation is the number of mutations in each variant, where apart from P6H7, the number of mutations follows a general trend, with the more mutations a variant had (Table 3.3) the more stable (both kinetically and thermally) that mutant was (Table 3.4). Results of k_{inact} and T_m show that P8B11, with only 4/8 mutations yielded the lowest stability measurements, while mutants P5E1 and P8E5 with 6/8 mutations yielded the highest stabilities. Correlating this back to PROSS, these results suggest there is value in the application of this software to tannase engineering. It makes one wonder what the outcome might have been if all 62 originally predicted mutations were incorporated, or if all mutations suggested for each one of the binding states or the FLT model were tested respectively as 4 different sets. This latter experiment would test the value of the multi-state prediction concept.

Finally, when looking at the mutations in each variant, the effects of each mutation can be interpreted to some degree. The mutations Q63T, A65D, and A184Y are shared between all 4 variants. As such there is a high likelihood that they are at least partially responsible for the increases in the T_m values that were observed in all of the tested mutants, with variations in values attributable to the remaining mutations which differ between variants. These 3 variations all incorporate oxygen-hydrogen bonds that were not previously present in wild type. They also face towards the surface of the protein. Together these compiled sets of mutations increase the overall number of hydrogen bonds/polar and surface area, which in turn could increase the solubility and thus thermostability of the protein.[84, 93]

Mutation V276Y, which is in α-helix 10 (Figure 1.3), and is facing towards the surface of the protein may contribute to decreasing the K_m values as in the case of variants P5E1 and P8E5 (albeit not significant decreases). This possible increase in productive binding could be due to increase stabilization of the helix, which may help the protein bind to the selected substrate MeG.[94, 95] This effect may be caused by an increase in stabilization with the introduction of the larger and polar side chain, which may increase stabilizing contacts and allow for better affinity with MeG. However, visual inspection of the structure shows that α-helix 10 does is not interacting with any other secondary structure within *Lp*Tan, and with the V276Y mutation appearing distant from the active site of the enzyme. Thus, further computational tests are needed to help determine any contribution of this this mutation to the variants kinetics/stability.

Future directions, in addition to repeating all the existing experiments with increased repetitions for clarification and validation, would involve characterizing additional promising variants from the original high-throughput screening. Since some mutants displayed a higher T_m, another test measuring optimal temperature could be added, with another set of kinetic testing

performed at optimal temperature. Also, testing the effects of the 8 mutations on wild type, rather than FLT, should be studied. A final test of mutations not shared by the 3 binding models and FLT structures selected by PROSS could be done as negative controls to test if mutations shared by the models have a greater effect on thermostability and TTN.

3.6 Conclusion

Ultimately, FLT was not deemed to be a significant improvement over *Lp*Tan. This most likely goes to the flap being critical for stabilization, notable in the significant decrease in T_m observed upon deletion of the flap. With respect to the variants, the PROSS generated mutant combinatorial library was successfully created, with the overall goal of creating a more thermostable enzyme also being successful, The observed increased Tm values proved to be the most significant and reliable outcomes, compared to both FLT and *Lp*Tan., The observed discrepancies between k_{inact} and T_m may or may not be relevant due to the large standard deviations observed in the k_{inact} data; although if validated additional experiments probing folding, unfolding and aggregation states are recommended. Looking forward most of the experiments should be repeated for validation, with increased repetitions for reduction of standard deviations across the data to clarify and substantiate outcomes.

Chapter 4: General Discussion and Future Directions

4.1 Genetic Code Expansion

4.1.1 Summary of Incorporating an ncAA into the Active Site of *Lp*Tan

The experiment of creating the NmH451H mutation into *Lp*Tan, resulted in a significant decrease in activity (k_{cat}) and an increase in K_m. Future directions involve mutating D419A to increase hydrophobicity with the methyl group of NmH or mutating S163C, with the expectation that the thiol group would act as a better nucleophile than its hydroxyl counterpart. Finally, since the mass spectroscopy analysis was inconclusive about the incorporation of the NmH, X-ray crystallography was proposed for validation of incorporation, with the caveat that it is understood that x-ray crystallography would require increased yields and purity to achieve, and thus may not be practical.

4.1.2 Practicality of GCE for Industrial Usage

Although not a new idea, the utilization of GCE to site-specifically incorporate ncAA to increase activity is still under explored. Problems including low ncAA incorporation efficiency, price of creating incorporated enzymes, difficultly with confirmation of incorporation are immediately notable. Despite studies that investigate improving the efficiency of GCE,[96] due to the limited research done on them, these concerns make GCE not viable for industries in its current state, with further research needed.

4.2 Flapless *Lp*Tan

4.2.1 Summary of Loop Deletion in Increasing Stability in *Lp*Tan

When compared to the *Lp*Tan, the FLT variant showed an increase in the TTN, coincident with a decrease in the catalytic efficiency, likely due to having a decreased affinity with the reaction substrate, MeG. When looking at the possibilities relating to protein modification, deletion of the

flap region of tannase did create a more stable enzyme having a lower k_{inact}, but did not create a more thermostable one lowering the T_m. Further testing could look at kinetics with larger substrates that more closely mimic tannins.

4.2.2 Modifications of Secondary Structure in Improving Biocatalytic Performance

The secondary structure of a protein plays a large part in stability, increasing hydrogen bonding networks based on the regular repeating nature of the structures within the overall structure of the protein. This helps maintain the protein structure by providing rigidity. When comparing the secondary structures to one another, alpha helices are shown to help increase conformational stability of proteins with the stabilization of alpha helices being previously used to help in designing industrially relevant enzymes.[97] While alpha helices seem to help with increasing the stability of an enzyme, loops offer the protein increased flexibility and are used in the linking between beta sheets and alpha helices. In a previous study it was shown that changing the length of the loop of acylphosphatase affected its stability and unfolding rate, while showing that the reduction of loop length resulted in a gradual thermal and chemical stabilization of the protein.[98] In this current study on tannase the "flap" region of *Lp*Tan was removed. Alternately, it would be interesting to see the effects of stabilizing the flap by looking to see what residues between the flap and main body of the protein interact with each other and making mutations that would strengthen these interactions.[99] For example if the H-bond between N238 and W91 were replaced with a salt bridge or disulfide bond (Figure 4.1). This modification, while perhaps not exactly mimicking the backbone-backbone hydrogen bond conformation, could provide stronger stabilizing forces that help in keeping the activity that is seen in the *Lp*Tan while increasing stability. Of course, if flexibility of the loop is important for substrate access, the impact of pinning the loop in place could, alternatively, be quite detrimental.

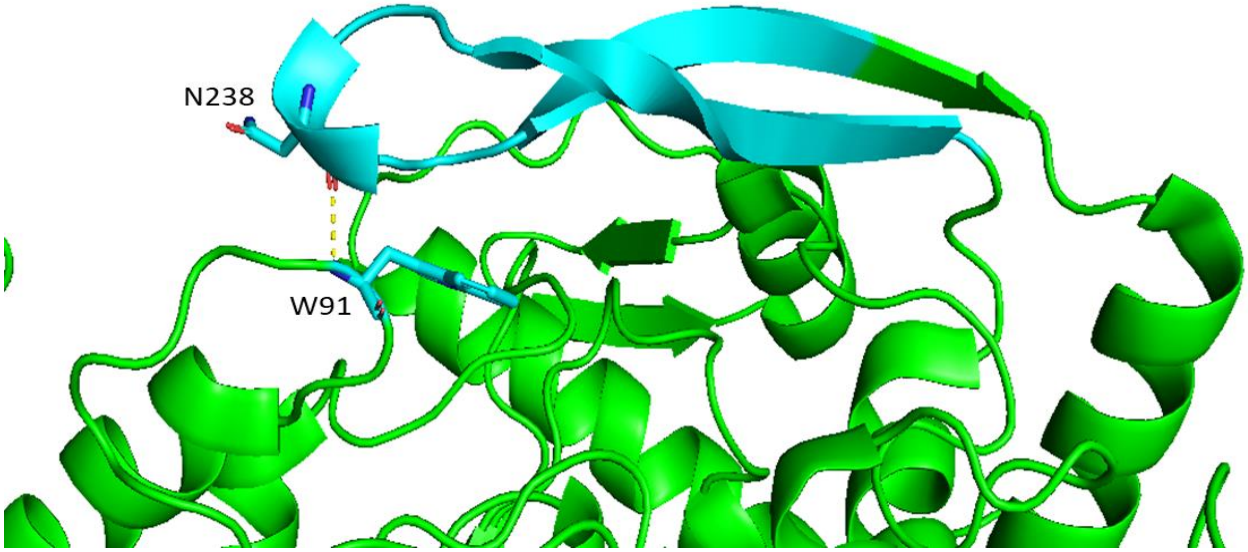


Figure 4.1: Structure of *Lp*Tan Showcasing "flap" Interaction. In Green is *Lp*Tan while in cyan shows the flap region, The yellow dotted line represents the H-bond between residues N238 and W91. pdbid: 3WA6.[52]

Modification of secondary structure does show relevance in not only changing the protein's stability but also activity with future directions for this project relating to how well the FLT mutant would interact with the preferred substrate, tannins.

4.3 PROSS Generated Combinatorial Library

4.3.1 Summary of PROSS Generated Mutations Effects

The 3-state design method of using the apo, substrate binding and product binding forms of *Lp*Tan is anticipated to consider any flexibility in the active site as it transitions through catalysis, while simultaneously looking at mutations that help to increase the stability of the enzyme. Further testing of these results would involve sequencing additional hits identified via the high throughput screening assay and characterizing their stability and activities. As noted in chapter 3, an experiment in which all mutations suggested for each of the binding forms or FLT model were tested respectively as 4 different sets would directly address the value of the 3-state design prediction method.

4.3.2 Improving the Combinatorial library Design.

The experiments performed have helped to identify problems with a general optimization procedure to help with creating industrially relevant enzymes. When comparing the combinatorial variants selected for characterization, and the FLT mutant that they were inspired by the lack of correlation between k_{inact} and T_m was notable. It was originally assumed that kinetic stability (k_{inact}) and thermal stability (T_m) would correlate, and that the results would show an increase in kinetic stability, as thermal stability was increased across most of the mutants.[98] However this was not the case. From this, further research into the relationship between kinetic and thermal stabilities needs to be explored. When looking at the differences between these two types of stabilities thermal stability, judged by an enzyme's melting temperature (T_m) is defined as the free energy of stabilization (difference in free energy between the folded and unfolded protein), while kinetic stability depends on the energy barrier to irreversible inactivation and is expressed as a rate of inactivation (k_{inact}). Future studies into this relationship need to be explored.

Things to consider for improving the library design strategy is that mutants should display an increase in either stability or activity, which both help to improve industrial relevance. When looking at reasons for a lack of stability in proteins, arise from evolutionary compromises made in order to facilitate binding and catalysis (activity-stability trade off), regulation of protein abundance in the organism, and adaptation to other environmental factors.[100, 101]

Focusing more on the activity-stability trade-off, there have been several advances in trying to counteract this. One method tries to differentiate between regions within a protein being classified as either a stability strength region or a stability weakness region, with their research showing oscillatory patterns of free energies centered around the catalytic region of selected

proteins.[102] These “shells” around the active site weakened in their effect on the stability by comparing the shells of between cold and heat adapted proteins. Hopefully this aspect can be incorporated into future designs which may help to contribute to creating more industrially relevant enzymes.

Industrial processes in general should look to optimize the cost of production with respect to ensuring profits. This study demonstrates that variants can be generated with unique characteristics, enabling application of the “best” engineered enzyme for a given industrial reaction condition.

Bibliography

1. Belur, P. D., & Mugeraya, G. (2011). Microbial production of tannase: State of the art. *Research Journal of Microbiology*, *6*(1), 25–40. https://doi.org/10.3923/jm.2011.25.40
2. Armor, J. N. (2010). A history of industrial catalysis. *ELSEVIER*, 3–9. https://doi.org/10.1016/j.cattod.2009.11.019
3. Roduner, E. (2014, October 14). *Understanding catalysis*. Chemical Society Reviews. https://pubs.rsc.org/en/content/articlehtml/2014/cs/c4cs00210e?casa_token=qwdUEbzqDXgAAAAA%3ALL0O_UCZQ0YaUJaRLDWWGymMqsBL8wzC2LIeRvDdelLKB2cmQEzUHMKlFWrjPa_nlGSlI285CVw_Hw
4. Harding, K. G., Dennis, J. S., von Blottnitz, H., & Harrison, S. T. L. (2008). A life-cycle comparison between inorganic and biological catalysis for the production of biodiesel. Journal of Cleaner Production, 16(13), 1368–1378. https://doi.org/10.1016/j.jclepro.2007.07.003
5. Kjaergaard, C. H., Rossmeisl, J., & Nørskov, J. K. (2010). Enzymatic versus inorganic oxygen reduction catalysts: Comparison of the energy levels in a free-energy scheme. Inorganic Chemistry, 49(8), 3567–3572. https://doi.org/10.1021/ic900798q
6. Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K., & Bornscheuer, U. T. (2020). Biocatalysis: Enzymatic synthesis for industrial applications. Angewandte Chemie International Edition, 60(1), 88–119. https://doi.org/10.1002/anie.202006648
7. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., & Robins, K. (2012). Engineering the third wave of biocatalysis. Nature, 485(7397), 185–194. https://doi.org/10.1038/nature11117
8. Porto de Souza Vandenberghe, L., Karp, S. G., Binder Pagnoncelli, M. G., von Linsingen Tavares, M., Libardi Junior, N., Valladares Diestra, K., Viesser, J. A., & Soccol, C. R. (2020). Classification of enzymes and catalytic properties. *Biomass, Biofuels, Biochemicals*, 11–30. https://doi.org/10.1016/b978-0-12-819820-9.00002-8
9. Dordick, J. S. (1991). An introduction to industrial biocatalysis. *Biocatalysts for Industry*, 3–19. https://doi.org/10.1007/978-1-4757-4597-9_1
10. Hauer, B. (2020). Embracing nature's catalysts: A viewpoint on the future of biocatalysis. *ACS Catalysis*, *10*(15), 8418–8427. https://doi.org/10.1021/acscatal.0c01708
11. Bommarius, A. S. (2015). Biocatalysis: A status report. *Annual Review of Chemical and Biomolecular Engineering*, *6*(1), 319–345. https://doi.org/10.1146/annurev-chembioeng-061114-123415
12. Sheldon, R. A., & Brady, D. (2019). Broadening the scope of biocatalysis in sustainable organic synthesis. *ChemSusChem*, *12*(13), 2859–2881. https://doi.org/10.1002/cssc.201900351
13. Gould, S. E. (2014, January 12). *Speeding up reactions: Biological vs. Chemical Catalysts*. Scientific American Blog Network. https://blogs.scientificamerican.com/lab-rat/speeding-up-reactions-biological-vs-chemical-catalysts/
14. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., & Witholt, B. (2001). Industrial biocatalysis today and Tomorrow. *Nature*, *409*(6817), 258–268. https://doi.org/10.1038/35051736
15. Singh, R., Kumar, M., Mittal, A., & Mehta, P. K. (2016). Microbial Enzymes: Industrial progress in 21st Century. *3 Biotech*, *6*(2). https://doi.org/10.1007/s13205-016-0485-8

16. Linares-Pasten, J., Andersson, M., & Karlsson, E. (2014). Thermostable glycoside hydrolases in Biorefinery Technologies. *Current Biotechnology*, *3*(1), 26–44. https://doi.org/10.2174/22115501113026660041
17. Poppe, L., & Vértessy, B. G. (2018). The fourth wave of biocatalysis emerges— the 13 th international symposium on Biocatalysis and Biotransformations. *ChemBioChem*, *19*(4), 284–287. https://doi.org/10.1002/cbic.201700687
18. Li, Y., & Cirino, P. C. (2014). Recent advances in engineering proteins for biocatalysis. *Biotechnology and Bioengineering*, *111*(7), 1273–1287. https://doi.org/10.1002/bit.25240
19. Wang, X., Xu, K., Tan, Y., Liu, S., & Zhou, J. (2023). Possibilities of using de novo design for generating diverse functional food enzymes. *International Journal of Molecular Sciences*, *24*(4), 3827. https://doi.org/10.3390/ijms24043827
20. Buller, R., Lutz, S., Kazlauskas, R. J., Snajdrova, R., Moore, J. C., & Bornscheuer, U. T. (2023). From nature to industry: Harnessing Enzymes for biocatalysis. *Science*, *382*(6673). https://doi.org/10.1126/science.adh8615
21. Korendovych, I. V. (2017). Rational and Semirational Protein Design. *Methods in Molecular Biology*, 15–23. https://doi.org/10.1007/978-1-4939-7366-8_2
22. Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and Design. *Current Opinion in Biotechnology*, *21*(6), 734–743. https://doi.org/10.1016/j.copbio.2010.08.011
23. Song, Z., Zhang, Q., Wu, W., Pu, Z., & Yu, H. (2023, January 16). *Rational design of enzyme activity and enantioselectivity*. Frontiers. https://www.frontiersin.org/articles/10.3389/fbioe.2023.1129149/full
24. Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., & Fleishman, S. J. (2016). Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Molecular Cell*, *63*(2), 337–346. https://doi.org/10.1016/j.molcel.2016.06.012
25. P. Singh, A., & Kumar, S. (2020). Applications of tannins in industry. *Tannins - Structural Properties, Biological Properties and Current Knowledge*.
26. Serrano, J., Puupponen-Pimiä, R., Dauer, A., Aura, A., & Saura-Calixto, F. (2009). Tannins: Current knowledge of food sources, intake, bioavailability and biological effects. *Molecular Nutrition & Food Research*, *53*(S2). https://doi.org/10.1002/mnfr.200900039
27. Harborne, J. B., Mabry, T. J., & Mabry, H. (1975). *The flavonoids*. Academic Press.
28. Dehghanian, Z., Habibi, K., Dehghanian, M., Aliyar, S., Asgari Lajayer, B., Astatkie, T., Minkina, T., & Keswani, C. (2022). Reinforcing the bulwark: Unravelling the efficient applications of plant phenolics and tannins against environmental stresses. *Heliyon*, *8*(3). https://doi.org/10.1016/j.heliyon.2022.e09094
29. Huang, Q., Liu, X., Zhao, G., Hu, T., & Wang, Y. (2018). Potential and challenges of tannins as an alternative to in-feed antibiotics for farm animal production. *Animal Nutrition*, *4*(2), 137–150. https://doi.org/10.1016/j.aninu.2017.09.004
30. *Tannin market size, share & growth analysis report, 2030*. Tannin Market Size, Share & Growth Analysis Report, 2030. (2017, April). https://www.grandviewresearch.com/industry-analysis/tannin-market
31. Hassanpour, S., Maheri-Sis, N., Eshratkhah, B., & mehmandar, F. baghbani. (2020, August 17). *Plants and secondary metabolites (tannins): A review*.

216069956_Plants_and_secondary_metabolites_Tannins_A_Review. https://www.researchgate.net/profile/Shahin-Hassanpour/publication/216069956_Plants_and_secondary_metabolites_Tannins_A_Review/links/5f3a89f8a6fdcccc43cfdb1c/Plants-and-secondary-metabolites-Tannins-A-Review.pdf?origin=publication_detail

32. Pizzi, A. (2019). Tannins: Prospectives and actual industrial applications. *Biomolecules*, *9*(8), 344. https://doi.org/10.3390/ biom9080344
33. Jiang, Y., Pei, J., Zheng, Y., Miao, Y., Duan, B., & Huang, L. (2021). Gallic acid: A potential anti-cancer agent. *Chinese Journal of Integrative Medicine*, *28*(7), 661–671. https://doi.org/10.1007/s11655-021-3345-2
34. Choubey, S., Goyal, S., Varughese, L. R., Kumar, V., Sharma, A. K., & Beniwal, V. (2018). Probing gallic acid for its broad spectrum applications. *Mini-Reviews in Medicinal Chemistry*, *18*(15), 1283–1293. https://doi.org/10.2174/1389557518666180330114010
35. Keyvani-Ghamsari, S., Rahimi, M., & Khorsandi, K. (2023). An update on the potential mechanism of gallic acid as an antibacterial and anticancer agent. *Food Science & Nutrition*, *11*(10), 5856–5872. https://doi.org/10.1002/fsn3.3615
36. Aguilar, C. N., & Gutierrez-Sanchez, G. (2001). Review: Sources, properties, applications and potential uses of tannin acyl hydrolase. *Food Science and Technology International*, *7*(5), 373–382. https://doi.org/10.1106/69m3-b30k-cf7q-rj5g
37. Belmares, R., Contreras-Esquivel, J. C., Rodríguez-Herrera, R., Coronel, A. R., & Aguilar, C. N. (2004). Microbial production of tannase: An enzyme with potential use in food industry. *LWT - Food Science and Technology*, *37*(8), 857–864.
38. Beniwal, V., Kumar, A., Sharma, J., & Chhokar, V. (2013). Recent advances in industrial application of Tannases: A Review. *Recent Patents on Biotechnology*, *7*(3), 228–233. https://doi.org/10.2174/18722083113076660013
39. Pan, H., Zhan, J., Yang, H., Wang, C., Liu, H., Zhou, H., Zhou, H., Lu, X., Su, X., & Tian, Y. (2021). Improving the acid resistance of Tannase TANBLP (AB379685) from lactobacillus plantarum ATCC14917T by site-specific mutagenesis. *Indian Journal of Microbiology*, *62*(1), 96–102. https://doi.org/10.1007/s12088-021-00983-x
40. Curiel, J. A., Rodríguez, H., Acebrón, I., Mancheño, J. M., De Las Rivas, B., & Muñoz, R. (2009). Production and physicochemical properties of recombinant *lactobacillus plantarum* tannase. *Journal of Agricultural and Food Chemistry*, *57*(14), 6224–6230. https://doi.org/10.1021/jf901045s
41. Jiménez, N., Esteban-Torres, M., Mancheño, J. M., de las Rivas, B., & Muñoz, R. (2014). Tannin degradation by a novel Tannase enzyme present in some lactobacillus plantarum strains. *Applied and Environmental Microbiology*, *80*(10), 2991–2997.
42. Long, J. Z., & Cravatt, B. F. (2011). The metabolic serine hydrolases and their functions in mammalian physiology and disease. *Chemical Reviews*, *111*(10), 6022–6063. https://doi.org/10.1021/cr200075y
43. Holmquist, M. (2000). Alpha beta-hydrolase fold enzymes structures, functions and mechanisms. *Current Protein and Peptide Science*, *1*(2), 209–235. https://doi.org/10.2174/1389203003381405
44. Mindrebo, J. T., Nartey, C. M., Seto, Y., Burkart, M. D., & Noel, J. P. (2016). Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom.

Current Opinion in Structural Biology, *41*, 233–246. https://doi.org/10.1016/j.sbi.2016.08.005
45. Nardini, M., & Dijkstra, B. W. (1999). A/β hydrolase fold enzymes: The family keeps growing. *Current Opinion in Structural Biology*, *9*(6), 732–737.
46. Simon, G. M., & Cravatt, B. F. (2010). Activity-based proteomics of enzyme superfamilies: Serine hydrolases as a case study. *Journal of Biological Chemistry*, *285*(15), 11051–11055. https://doi.org/10.1074/jbc.r109.097600
47. Rauwerdink, A., & Kazlauskas, R. J. (2015). How the same core catalytic machinery catalyzes 17 different reactions: The serine-histidine-aspartate catalytic triad of α/β-hydrolase fold enzymes. *ACS Catalysis*, *5*(10), 6153–6176. https://doi.org/10.1021/acscatal.5b01539
48. Dodson, G. (1998). Catalytic triads and their relatives. *Trends in Biochemical Sciences*, *23*(9), 347–352. https://doi.org/10.1016/s0968-0004(98)01254-7
49. Distaso, M., Cea-Rama, I., Coscolín, C., Chernikova, T. N., Tran, H., Ferrer, M., Sanz-Aparicio, J., & Golyshin, P. N. (2022). *The Mobility of the Cap Domain Is Essential for the Substrate Promiscuity of a Family IV Esterase from Sorghum Rhizosphere Microbiome*. https://doi.org/10.1101/2022.09.10.507389
50. Dong, L., McKinstry, W. J., Pan, L., Newman, J., & Ren, B. (2021b, February 8). *Crystal structure of fungal tannase from Aspergillus niger*. Acta crystallographica. Section D, Structural biology. https://pubmed.ncbi.nlm.nih.gov/33559614/
51. Ren, B., Wu, M., Wang, Q., Peng, X., Wen, H., McKinstry, W. J., & Chen, Q. (2013). *Crystal Structure of Tannase from from Lactobacillus Plantarum*. https://doi.org/10.2210/pdb4jui/pdb
52. The Pymol Molecular Graphics System, Version 2.0 Schrodinger, LLC
53. Jiménez, N., Esteban-Torres, M., Mancheño, J. M., de las Rivas, B., & Muñoz, R. (2014a). Tannin degradation by a novel Tannase enzyme present in some lactobacillus plantarum strains. *Applied and Environmental Microbiology*, *80*(10), 2991–2997. https://doi.org/10.1128/aem.00324-14
54. Matoba, Y., Tanaka, N., & Sugiyama, M. (2013). *Crystal Structure of Tannase from Lactobacillus Plantarum in the Orthorhombic Crystal*. https://doi.org/10.2210/pdb3wa6/pdb
55. Ristinmaa, A. S., Coleman, T., Cesar, L., Langborg Weinmann, A., Mazurkewich, S., Brändén, G., Hasani, M., & Larsbrink, J. (2022). Structural diversity and substrate preferences of three tannase enzymes encoded by the anaerobic bacterium clostridium butyricum. *Journal of Biological Chemistry*, *298*(4), 101758.
56. Liu, L., Guo, J., Zhou, X.-F., Li, Z., Zhou, H.-X., & Song, W.-Q. (2022). Characterization and secretory expression of a thermostable tannase from Aureobasidium Melanogenum T9: Potential candidate for Food and Agricultural Industries. *Frontiers in Bioengineering and Biotechnology*, *9*. https://doi.org/10.3389/fbioe.2021.769816
57. Xiao, H., Nasertorabi, F., Choi, S., Han, G. W., Reed, S. A., Stevens, R. C., & Schultz, P. G. (2015). Exploring the potential impact of an expanded genetic code on protein function. *Proceedings of the National Academy of Sciences*, *112*(22), 6961–6966. https://doi.org/10.1073/pnas.1507741112
58. Pott, M., Hayashi, T., Mori, T., Mittl, P. R. E., Green, A. P., & Hilvert, D. (2018). A noncanonical proximal heme ligand affords an efficient peroxidase in a globin fold.

Journal of the American Chemical Society, *140*, 1535–1543. https://doi.org/10.1021/jacs.7b12621.s001
59. Burke, A. J., Lovelock, S. L., Frese, A., Crawshaw, R., Ortmayer, M., Dunstan, M., Levy, C., & Green, A. P. (2019). Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature*, *570*(7760), 219–223. https://doi.org/10.1038/s41586-019-1262-8
60. Nikić-Spiegel, I. (2020). Expanding the genetic code for neuronal studies. *ChemBioChem*, *21*(22), 3169–3179. https://doi.org/10.1002/cbic.202000300
61. Zhao, J., Burke, A. J., & Green, A. P. (2020). Enzymes with noncanonical amino acids. *Current Opinion in Chemical Biology*, *55*, 136–144. https://doi.org/10.1016/j.cbpa.2020.01.006
62. Drienovská, I., & Roelfes, G. (2020). Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nature Catalysis*, *3*(3), 193–202. https://doi.org/10.1038/s41929-019-0410-8
63. Arranz-Gibert, P., Patel, J. R., & Isaacs, F. J. (2019). The role of orthogonality in genetic code expansion. *Life*, *9*(3), 58. https://doi.org/10.3390/life9030058
64. Xiao, H., Peters, F. B., Yang, P.-Y., Reed, S., Chittuluru, J. R., & Schultz, P. G. (2014). Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-trna synthetase. *ACS Chemical Biology*, *9*(5), 1092–1096. https://doi.org/10.1021/cb500032c
65. Jackson, J. C., Duffy, S. P., Hess, K. R., & Mehl, R. A. (2006). Improving natures enzyme active site with genetically encoded unnatural amino acids. *Journal of the American Chemical Society*, *128*(38), 11124–11127. https://doi.org/10.1021/ja061099y.s001
66. Mattheisen, J. M., Wollowitz, J. S., Huber, T., & Sakmar, T. P. (2023). Genetic code expansion to enable site-specific bioorthogonal labeling of functional G protein-coupled receptors in live cells. *Protein Science*, *32*(2). https://doi.org/10.1002/pro.4550
67. Walker, J. M. (1994). The bicinchoninic acid (BCA) assay for protein quantitation. *Protein Protocols Handbook, The*, 11–14. https://doi.org/10.1385/1-59259-169-8:11
68. Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B., & Liu, W. R. (2012). A rationally designed pyrrolysyl-trna synthetase mutant with a broad substrate spectrum. *Journal of the American Chemical Society*, *134*(6), 2950–2953. https://doi.org/10.1021/ja211972x
69. Sharma, S., Bhat, T. K., & Dawra, R. K. (2000). A spectrophotometric method for assay of tannase using rhodanine. *Analytical Biochemistry*, *279*(1), 85–89. https://doi.org/10.1006/abio.1999.4405
70. Chung, C. Z., Amikura, K., & Söll, D. (2020). Using genetic code expansion for protein biochemical studies. *Frontiers in Bioengineering and Biotechnology*, *8*.
71. Shandell, M. A., Tan, Z., & Cornish, V. W. (2021). Genetic code expansion: A brief history and perspective. *Biochemistry*, *60*(46), 3455–3469. https://doi.org/10.1021/acs.biochem.1c00286
72. Ferreira, J. C., Fadl, S., Villanueva, A. J., & Rabeh, W. M. (2021). Catalytic dyad residues his41 and cys145 impact the catalytic activity and overall conformational fold of the main SARS-COV-2 protease 3-chymotrypsin-like protease. *Frontiers in Chemistry*, *9*. https://doi.org/10.3389/fchem.2021.692168
73. Gisdon, F. J., Bombarda, E., & Ullmann, G. M. (2022). Serine and cysteine peptidases: So similar, yet different. how the active-site electrostatics facilitates different reaction

mechanisms. *The Journal of Physical Chemistry B*, *126*(22), 4035–4048. https://doi.org/10.1021/acs.jpcb.2c01484
74. Carugo, O. (2022). *Interplay between Hydrogen and Chalcogen Bond in Cysteine*. https://doi.org/10.22541/au.165648965.56989348/v1
75. Illanes, A. (1999). Stability of biocatalysts. *Electronic Journal of Biotechnology*, *2*(1). https://doi.org/10.2225/vol2-issue1-fulltext-2
76. Kumar, S., Tsai, C.-J., & Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Engineering, Design and Selection*, *13*(3), 179–191. https://doi.org/10.1093/protein/13.3.179
77. Peccati, F., Alunno-Rufini, S., & Jiménez-Osés, G. (2023). Accurate prediction of enzyme thermostabilization with Rosetta using alphafold ensembles. *Journal of Chemical Information and Modeling*, *63*(3), 898–909. https://doi.org/10.1021/acs.jcim.2c01083
78. Wang, D., Liu, Y., Lv, D., Hu, X., Zhong, Q., Zhao, Y., & Wu, M. (2018). Retracted article: Substrates specificity of tannase from streptomyces sviceus and lactobacillus plantarum. *AMB Express*, *8*(1). https://doi.org/10.1186/s13568-018-0677-1
79. Kumwenda, B., Litthauer, D., Bishop, Ö. T., & Reva, O. (2013). Analysis of protein thermostability enhancing factors in industrially important*thermus*bacteria species. *Evolutionary Bioinformatics*, *9*. https://doi.org/10.4137/ebo.s12539
80. Thompson, M. J., & Eisenberg, D. (1999). Transproteomic evidence of loop-deletion mechanism for enhancing protein thermostability. *Journal of Molecular Biology*, *292*(4), 946. https://doi.org/10.1006/jmbi.1999.3149
81. Davey, J. A., & Chica, R. A. (2012). Multistate approaches in Computational protein design. *Protein Science*, *21*(9), 1241–1252. https://doi.org/10.1002/pro.2128 PMID: 27425410 PMCID: PMC4961223 PDF HTML
82. Dong, F., Zhang, M., Ma, R., Lu, C., & Xu, F. (2022). Insights of conformational dynamics on catalytic activity in the computational stability design of bacillus subtilis Lipa. *Archives of Biochemistry and Biophysics*, *722*, 109196. https://doi.org/10.1016/j.abb.2022.109196
83. Peleg, Y., Vincentelli, R., Collins, B. M., Chen, K.-E., Livingstone, E. K., Weeratunga, S., Leneva, N., Guo, Q., Remans, K., Perez, K., Bjerga, G. E. K., Larsen, Ø., Vaněk, O., Skořepa, O., Jacquemin, S., Poterszman, A., Kjær, S., Christodoulou, E., Albeck, S., … Fleishman, S. J. (2021). Community-wide experimental evaluation of the Pross Stability-Design Method. *Journal of Molecular Biology*, *433*(13), 166964. https://doi.org/10.1016/j.jmb.2021.166964
84. Jung, F., Frey, K., Zimmer, D., & Mühlhaus, T. (2023). DeepSTABp: A deep learning approach for the prediction of thermal protein stability. *International Journal of Molecular Sciences*, *24*(8), 7444. https://doi.org/10.3390/ijms24087444
85. Rogers, T. A., & Bommarius, A. S. (2010). Utilizing simple biochemical measurements to predict lifetime output of biocatalysts in continuous isothermal processes. *Chemical Engineering Science*, *65*(6), 2118–2124. https://doi.org/10.1016/j.ces.2009.12.005
86. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The PHYRE2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053
87. Walton, C. J. W., & Chica, R. A. (2013). A high-throughput assay for screening L- or D-amino acid specific aminotransferase mutant libraries. *Analytical Biochemistry*, *441*(2), 190–198. https://doi.org/10.1016/j.ab.2013.07.004

88. Tomasiak, T. M., Pedersen, B. P., Chaudhary, S., Rodriguez, A., Colmanares, Y. R., Roe-Zurz, Z., Thamminana, S., Tessema, M., & Stroud, R. M. (2014). General QPCR and plate reader methods for rapid optimization of membrane protein purification and crystallization using thermostability assays. *Current Protocols in Protein Science*, *77*(1). https://doi.org/10.1002/0471140864.ps2911s77
89. Varadi, M *et al*. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* (2021).
90. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673
91. Jumper, J *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021).
92. *A new bioinformatics analysis tools framework at EMBL-EBI (2010)* Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R Nucleic acids research 2010 JµL, 38 Suppl: W695-9 doi:10.1093/nar/gkq313
93. Vogt, G., & Argos, P. (1997). Protein thermal stability: Hydrogen bonds or internal packing? *Folding and Design*, *2*. https://doi.org/10.1016/s1359-0278(97)00062-x
94. Lyu, P. C., Sherman, J. C., Chen, A., & Kallenbach, N. R. (1991). Alpha-helix stabilization by natural and unnatural amino acids with alkyl side chains. *Proceedings of the National Academy of Sciences*, *88*(12), 5317–5320. https://doi.org/10.1073/pnas.88.12.5317
95. Cornish, V. W., Kaplan, M. I., Veenstra, D. L., Kollman, P. A., & Schultz, P. G. (1994). Stabilizing and destabilizing effects of placing .beta.-branched amino acids in protein .alpha.-helixes. *Biochemistry*, *33*(40), 12022–12031. https://doi.org/10.1021/bi00206a003
96. Fu, X., Huang, Y., & Shen, Y. (2022). Improving the efficiency and orthogonality of genetic code expansion. *BioDesign Research*, *2022*. https://doi.org/10.34133/2022/9896125
97. Yakimov, A. P., Afanaseva, A. S., Khodorkovskiy, M. A., & Petukhov, M. G. (2016). Design of stable α-helical peptides and thermostable proteins in biotechnology and biomedicine. *Acta Naturae*, *8*(4), 70–81. https://doi.org/10.32607/20758251-2016-8-4-70-81
98. Kumar, Shashi, & Deshpande, P. A. (2021). Structural and thermodynamic analysis of factors governing the stability and thermal folding/unfolding of sazca. *PLOS ONE*, *16*(4). https://doi.org/10.1371/journal.pone.0249866
99. Chan, C.-H., Yu, T.-H., & Wong, K.-B. (2011). Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0021624
100. Xie, Y., An, J., Yang, G., Wu, G., Zhang, Y., Cui, L., & Feng, Y. (2014). Enhanced enzyme kinetic stability by increasing rigidity within the active site. *Journal of Biological Chemistry*, *289*(11), 7994–8006. https://doi.org/10.1074/jbc.m113.536045
101. Goldenzweig, A., & Fleishman, S. J. (2018). Principles of protein stability and their application in computational design. *Annual Review of Biochemistry*, *87*(1), 105–129. https://doi.org/10.1146/annurev-biochem-062917-012102

102. Siddiqui, K. S. (2016). Defying the activity–stability trade-off in enzymes: Taking advantage of entropy to enhance activity and thermostability. *Critical Reviews in Biotechnology*, *37*(3), 309–322. https://doi.org/10.3109/07388551.2016.1144045

Appendixes

Equations

Since observed absorbance (A_{obs}) would be a combination of unreacted substrate (A_s) and product ($A_{p)}$, the concentration of product (c_p) was determined to be equal to the observed absorbance subtracted by initial absorbance of substrate at the tested concertation (A_s^0) then divided by the difference in the molar extinction coefficients of gallic acid (ε_p) and MeG (ε_s), obtained via standard curve plots seen in supplementary figure 1.

Equation 1: Concentration of Gallic Acid Product for Tannase Reaction

$$A_{obs} = A_p + A_s$$

$$A_{obs} = c_p l\varepsilon_p + c_s l\varepsilon_s$$

$$A_{obs} = c_p l\varepsilon_p - c_p l\varepsilon_s + c_s^{\circ} l\varepsilon_s$$

$$A_{obs} = c_p(l\varepsilon_p - l\varepsilon_s) + c_s^{\circ} l\varepsilon_s$$

$$A_{obs} - c_s^0 l\varepsilon_s = c_p(l\varepsilon_p - l\varepsilon_s)$$

$$c_p = \frac{(A_{obs} - A_s^0)}{(\varepsilon_p - \varepsilon_s)}$$

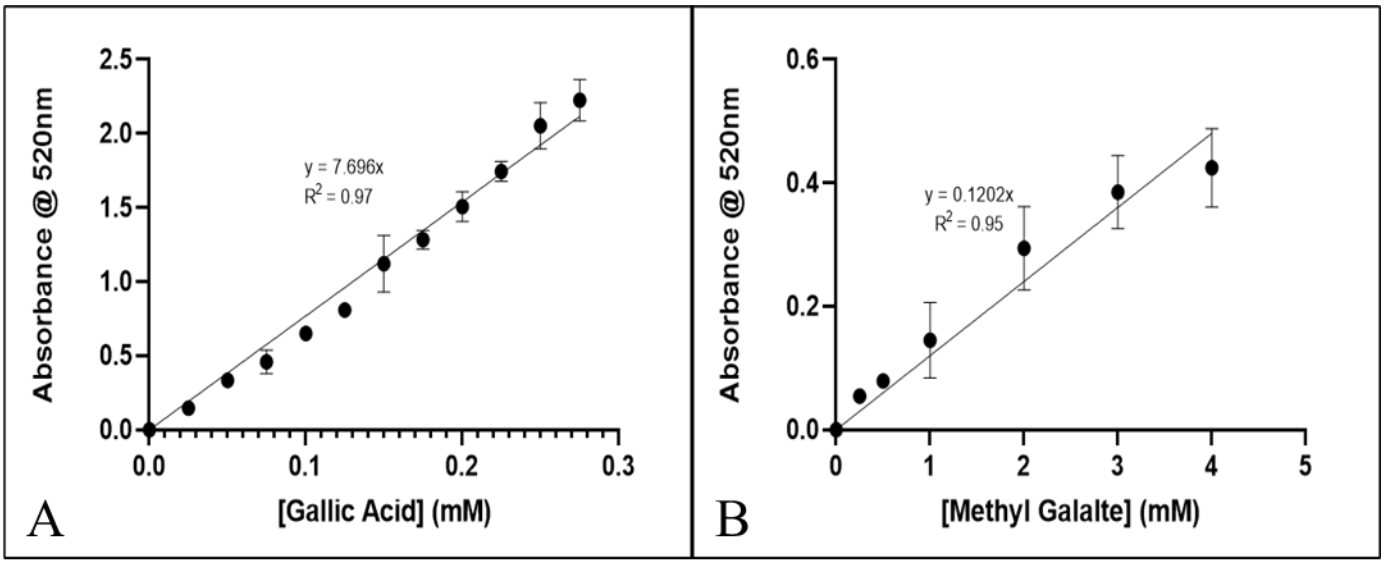
K_{inact} is obtained through inactivation plots.[74]

Equation 2: Determination of Half Life

$$\text{Half-life} = \frac{\ln 2}{k_{inact}}$$

Preliminary Tests

To accurately determine formation of gallic acid produced by tannase, two standard curves were created. The gallic acid standard curve was created to represent the theoretical amount of gallic acid produce from the tannase assay. The substrate MeG in high concentrations also produced an absorbance signal at 520 nm, so a standard curve of MeG was created and to account for unused MeG in reaction solution.



Supplementary Figure 1: Standard Curve Plots of Product (Gallic Acid, A) and Substrate (MeG, B). Standard curves were created using same rhodanine assay as prescribed for tannase apart from the addition of enzyme. Gallic acid plot used a concentration curve of 0-0.3 mM, while MeG plot concentration curve of 0-4mM. Plots were created using GraphPad Prism.

The high-throughput assay was performed with 768 colonies, theoretically representing the 256 possible combinations of the 8 site mutations generated from PROSS. Results were narrowed down, with selection of mutants for characterization focusing on mutants with a higher activity and residual activity than the FLT variant.