Improving the unnatural and promiscuous nitroaldolase activity of modern and ancestral HNLs

A Thesis

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Huey Yee Lim

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Advisor: Romas J. Kazlauskas

May 2016

Acknowledgements

I would like thank everyone from Kazlauskas group especially those who were also involved in the ancestral enzyme project. I thank Bryan Jones for providing very useful feedbacks and suggestions for my project. I thank Titu Devamani for helping me in synthesizing chemicals used in my experiments. I thank Alissa Rauwerdink and Joanna Mooney for giving me ideas on this project.

I specially thank Professor Kazlauskas for giving me opportunity to be his graduate student and work on this project. Our weekly meeting has helped tremendously in the progress in this project. I also thank Dr Seelig and Professor Wackett for willing to serve in my committee and providing good feedbacks on this project. I also thank Professor Travisano for the help he provided on statistical analysis.

I thank my husband, Soon Li Teh, for supporting me throughout the graduate study. I would also like to thank Donna and Waihon Liew for being such good friends during my time in Minnesota.

Dedication

This thesis is dedicated to everyone who is interested in biocatalysis, protein engineering or green chemistry.

Abstract

Enzymes in superfamilies such as the α/β-hydrolase fold superfamily have similar structures and active sites because they evolved from a common ancestor. Since enzymes within a superfamily catalyze different reactions, the common ancestor may have been a promiscuous catalyst. Our hypothesis is that promiscuous ancestral enzymes may be better starting points to obtain new catalytic activities compared to modern enzymes. To test this hypothesis, I tried to improve the promiscuous and unnatural nitroaldol activity in both a modern and an ancestral hydroxynitrile lyase using semi rational and random approaches. My results showed that the ancestral enzyme was not a better starting point compared to the modern enzyme after the modern enzyme was stabilized by a single mutation.

Table of Contents

Chapter 4. Experimental section

Chapter 5. Results and Discussion

List of Tables

List of Figures

Chapter 1. Introduction to Enzyme and α/β hydrolases

1.1 Definition of enzyme and its application

While some enzymes could be catalytic RNAs, most enzymes are proteins. Also known as biocatalysts, they act as catalysts to speed up chemical reactions in organisms. Enzymes are divided in six classes: oxidoreductases (catalyze oxidation and reduction involving electron transfer), transferases (catalyze transfer of functional groups), hydrolases (catalyze breakage of C-O, C-N and C-S with addition of H_2O), lyases (catalyze breakage or formation of C-C, C-N, C-O and C-S), isomerase (catalyze rearrangement of atoms in a molecule) and ligases (catalyze linkages of two molecules).[1] Enzymes are not only essential for the survival of organisms; they have also proven to be useful in industrial applications. They are used extensively in food, beverages, textile, detergent and other industries. Just to name two examples, proteases are used in cheese production while phytases are used in animal feed to liberate phosphorus to improve uptake by farm animals.[2,6]

In the recent attempt to venture into green chemistry, enzymes have huge potential in replacing traditional synthetic chemistry, which uses harsh, toxic chemicals and extreme reaction conditions, for their high efficiency, specificity, reusability and non-toxicity and biodegradability. Enzymes also operate in neutral pH, ambient temperature and pressure.[1] Reactions carried out by enzymes are not only more environmental friendly but are also energy efficient. Enzyme technology or whole cell biocatalysis has received 16 awards in the Presidential Green Chemistry Award Challenge since 2000 from the US Environmental Protection Agency that awards 5 prizes each year.[1]

1.2 How enzymes work

Enzymes provide a favorable environment to speed up chemical reactions by accepting substrate in its binding pocket, which is also known as an active site.[5] The surface of an active site is made up of amino acids with substituent groups that facilitate the chemical reaction and binding[5]. Most of the time, the substrate is 'enclosed' in the active site and is protected from solution. Binding results in an enzyme-substrate (ES) complex. Next, catalysis occurs to form product.[5]

$$E + S \rightleftharpoons ES \longrightarrow E + P$$

Enzymes, like other catalysts, speed up the reaction rates without affecting reaction equilibria since they are not consumed in reaction. They do so by lowering the activation energies.[5] Emil Fischer was one of the first who proposed that enzymes are structurally complementary to the substrate in a way that they fit like a lock and key.[5] Haldane and Pauling further developed the idea and proposed that an enzyme is complementary to its transition state to lower activation energy.[5] Contrary to the belief that enzymes are fixed structures, Koshland (1985) came out with the term "induced fit" to describe how enzymes could be 'induced' to undergo conformational changes once they are bound by the substrate and therefore, are flexible.[3] Mobile character of enzymes was studied and confirmed by nuclear magnetic resonance (NMR). Transition between conformational states determines enzymatic activity. Enzymes undergo conformational changes to bind or orient substrates and release products, as well as stabilizing intermediates.[3]

Enzyme proficiency can also be explained by non-covalent factors as well as covalent factors. Noncovalent factors that have been proposed are induced fit, binding to near attack

conformations, transition-state electrostatic stabilization and ground state desolvation and destabilization.[4] All these effects contribute to better complementarity of enzyme for the transition state. On the other hand, Zhang and Houk (2005) proposed that covalent factor also contributes to enzyme proficiency. Covalent bond that results from acid-base catalysis or hydrogen bonds between enzyme and substrate could also lead to reaction.[4]

1.3 Enzyme kinetics

Enzymes are usually characterized by their reaction rate and response to different experimental parameters.[5] A kinetics experiment could be carried out by measuring initial velocity, V_0 by using substrate concentration, [S], much higher than the enzyme concentration, [E]. When enzymes are first mixed with high amount of substrates, the concentration of [ES] increases in microseconds and it is usually not observable. This is called pre-steady state. After that, the concentration of ES complex stops changing over time because the rate of ES formation and the rate of its breakdown are the same. The reaction is in a steady state. By varying the [S], its effect on V_0 could be examined (Fig 1.1). At lower concentrations of substrates, V_0 increases in a linear fashion. V_0 eventually increases less and less (shown as plateau) as substrate concentration is increasing. The plateau region is called maximum velocity, V_{max} (Fig. 1.1). This enzyme kinetics is also known as steady-state kinetics.[5]

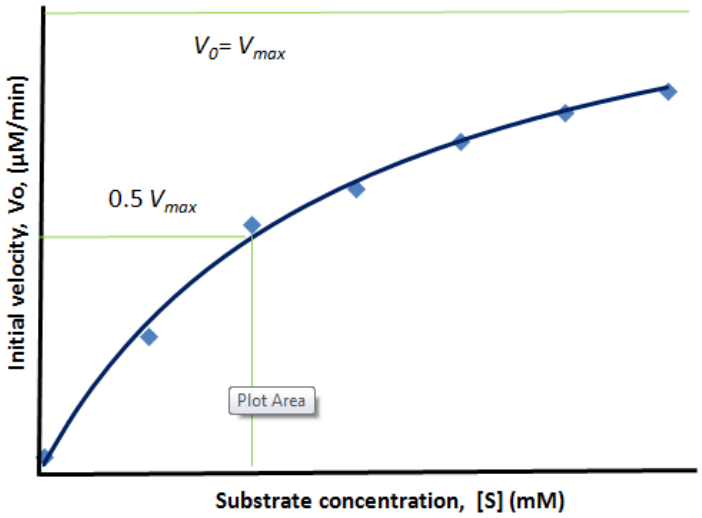


Figure 1.1. Michealis Menten curve. Initial velocity of enzyme, V_0 (µM/min) in response to different concentrations of substrate (mM) are shown.

The curve expressing the enzyme kinetics relationship between substrate concentration and initial velocity can also be expressed by Michaelis Menten equation where K_M, the Michaelis constant is the substrate concentration when $V_0 = \frac{1}{2}V_{max}$. K_M is the ratio of ES breakdown rate to ES formation rate and is used to describe enzyme affinity. Another useful constant, k_{cat}, which is also called the turnover number, can be used to describe rate-limiting step. Turnover number can be obtained dividing the total enzyme concentration from V_{max} and is used to describe number of molecules of substrates being converted into products in a given unit of time catalyzed by one molecule of enzyme saturated with its substrates. To compare different enzymes, the ratio of $k_{cat}$to K_M (specificity constant) is generally used. These parameters allow us to evaluate the kinetic efficiency of enzymes. The upper limit of k_{cat}/K_M (diffusion rate of enzyme and substrate in water) is between 10^8 and 10^9 M^{-1} S^{-1}. It is said that enzymes have obtained 'catalytic perfection' when it is close to this upper limit.[5]

$$V_o = \frac{V_{max}[S]}{K_m + [S]}$$

$$k_{cat} = \frac{V_{max}}{[E_t]}$$

1.4 α/β Hydrolase Fold Enzymes

The α/β hydrolase fold is a subclass of α/β domain, which is frequently found in nature, composing of central, parallel or mixed β sheets enclosed/bordered by α helices.[8] This fold is being shared by many enzymes with very diverse catalytic functions. Members belong to this family include hydroxynitrile lyase, lipase, esterase, haloalkane dehalogenase, acetylcholinesterase, epoxide hydrolase, dienelactone hydrolase.[7] These members do not share high sequence similarity or use the same nucleophiles or substrates but they are similar in structure and have a conserved arrangement of catalytic residues.[8] This suggests that these enzymes might have evolved from a common ancestor.[8] Modification to the fold such as insertion is well-tolerated by this protein fold without affecting the conserved catalytic machinery, demonstrating the plasticity of the fold and its remarkable adaptability.[8]

The primary features of α/β hydrolase fold consist of α helices surrounding eight stranded β sheets including the second, anti-parallel β sheet. β sheets display a left-handed superhelical twist and its degree of twisting can vary substantially, especially in between β5 and β6. Some helices could be absent except helix αC, which is very well-conserved as it plays important role of positioning nucleophile residue in the active site.[8] Many members of this family have lids or caps, which are formed due to loop insertion, play important role in the regulation of accessibility of substrate (Fig. 1.2).[8,9]

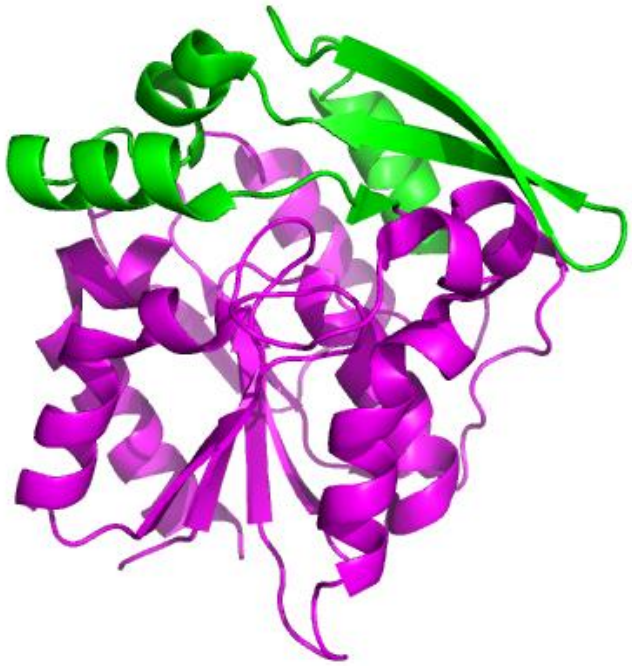


Figure 1.2. HbHNL as an example of α/β hydrolase. Purple region is the core domain while the green region is the cap domain of HbHNL(PDB:1YB6)

All α/β hydrolases have a highly conserved Nucleophile-His-Acid catalytic triad consisting of a nucleophile (serine, cysteine or aspartic acid) at the end of β5, an absolutely conserved histidine at the end of β strand, and an acidic residue at the end of β7 (Fig 1.3).[8] The nucleophile is situated in the "nucleophile elbow", defined by a common sequence S_m-X-Nuc-X-S_m (S_m=small residue, X=any residue, Nuc=nucleophile) that constitutes a sharp turn.[7,8] It is easily approached by the substrate and hydrolytic water molecule. A pocket called "oxyanion hole" is also formed to stabilize the negative charge of transition state (from an alkoxide) during hydrolysis in some α/β hydrolases. Oxyanion hole normally consists of two backbone nitrogen atoms: one is on a residue in between helix αA and strand β3 and the other one is on a residue right after the nucleophile. [8]

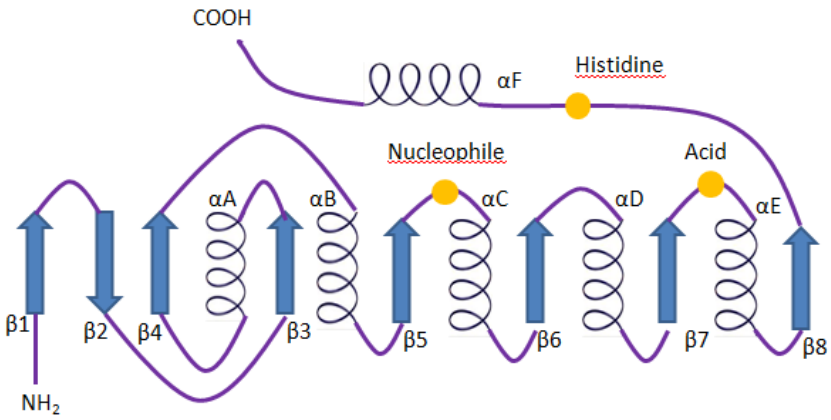


Figure 1.3. Two dimensional diagram of a typical α/β hydrolase. Catalytic domain is labeled in yellow dots.

Our group has worked on several enzymes from α/β hydrolase family, particularly on a modern esterase, salicylic acid binding protein 2 (SABP2) from the tobacco plant and hydroxynitrile lyase (HbHNL) from the rubber tree, *Hevea brasiliensis*. My project focuses on HbHNL as a model for modern enzyme from α/β hydrolase family. Several ancestral enzymes were reconstructed from these modern hydroxynitrile lyases and esterases and will be discussed in chapter 2.

Here, I am using SABP2 as an example mechanism of α/β hydrolase that contains an oxyanion hole. SABP2 is involved in plant defense system by catalyzing the hydrolysis of methyl salicylate (MeSA) to salicylic acid.[9] Salicylic acid plays important role in resistance signaling pathway by inducing expression of pathogenesis-related genes in plants.[9] The catalytic triad of SABP2 consists of Ser81, Asp210 and His 238. During catalysis, OH-group of Ser81 acts as a nucleophile that attacks the carboxyl carbon of MeSA. The negative charge of carbonyl oxygen that forms upon attack is stabilized by hydrogen bonding from oxyanion hole.[54] After the formation of a tetrahedral intermediate, the methanol group from salicyl serine intermediate

leaves.[54] His238 pulls a proton from a water molecule, initiating hydrolysis of carboxylic acid ester from Ser81.[54] The tetrahedral intermediate once again forms to stabilize the negative charge of the carbonyl group and the whole carboxylic acid leaves (Fig. 1.4).[54] Mechanism of HbHNL will be discussed in detail in the next section.

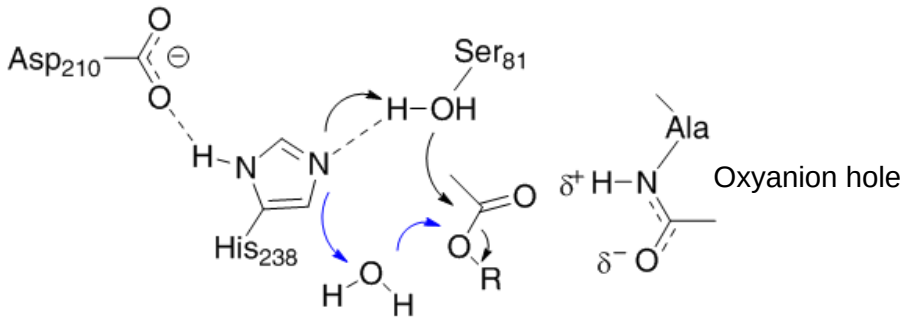


Figure 1.4. Mechanism of SABP2. Blue arrows shows reaction happens in later steps to form and release carboxylic acid. Adapted from Steiner and Schawab, 2012.[33]

1.5 Hydroxynitrile lyase, its function in nature and industrial application

Hydoxynitrile lyases (HNLs) do not uniquely belong to α/β hydrolase family, but are also found in peptidase-S10 superfamily, medium-chain reductase/dehydrogenase superfamily as well as glucose-methanol-choline (GMC) oxidoreductase N and C superfamily. Hydroxynitrile lyases are catabolic enzymes that catalyze cleavage of cyanohydrins into hydrogen cyanide, HCN and corresponding carbonyl compounds (Fig. 1.6a).

Among the S-selective hydroxynitrle lyases are HNLs from *Manihot esculenta* (cassava), *Hevea brasiliensis* (rubber tree) and *Sorghum bicolour* (millet).[19] Both Manihot esculenta HNL (MeHNL) and *Hevea brasiliensis* (HbHNL) accept aliphatic or aromatic aldehydes and methyl ketones while *Sorghum bicolour* (SbHNL) accepts only aromatic aldehydes and methyl ketones.[19] The other two R-selective hydroxynitrile lyases are *Prunus amygdalus* (bitter almond), PaHNL

and *Arabidopsis thaliana* (mouse ear cress), AtHNL. Both of these HNLs accept both aliphatic and aromatic aldehydes and methyl ketones as substrates as well. The natural substrate is acetone cyanohydrins for MeHNL, HbHNL and mandelonitrile for PaHNL. No natural substrate was found for AtHNL. Only a few well characterized hydroxynitrile lyases from α/β hydrolase family have been identified, such as HbHNL, MeHNL and AtHNL.[19] HNLs from α/β hydrolase are closely related to many other esterases and lipases.

The release of HCN from plants tissues or other organisms, termed cyanogenesis, has been observed in more than 3000 plant species and is important in plant defense.[11] Plants generally have enzymes exhibiting monooxygenase activity such as P450 that catalyzes earlier steps in cyanogenesis from hydrophobic aliphatic amino acids such as phenylalanine, leucine, isoleucine, tyrosine and valine.[14] In rubber tree, two cyanogenic glucosides (linamarin and lotaustralin) are synthesized from valine (major) and isoleucine (minor) respectively. They are converted to their corresponding oximes by P450.[14] Next, hydroxynitriles form through nitrile formation by dehydration and hydroxylation of the α-carbon, also catalyzed by a P450 enzyme (Fig. 1.5). A UDP-glucose glucosyltransferase catalyzes the glycosylation of hydroxynitriles resulting in production of cyanogenic glucoside.[12-14] Both cyanogenic glucosides are stored in vacuoles while HNL and β-glucosidase are stored in apoplasts to avoid leakage of HCN from tissues. They are only cleaved by β-glucosidase and turn into acetone cyanohydrins when triggered by predator attack.[13] Acetone cyanohydrin are then transformed by hydroxynitrile lyase to release HCN and ketones. Cyanogenesis can also happen slowly and spontaneously at high pH without a hydroxynitrile lyase.

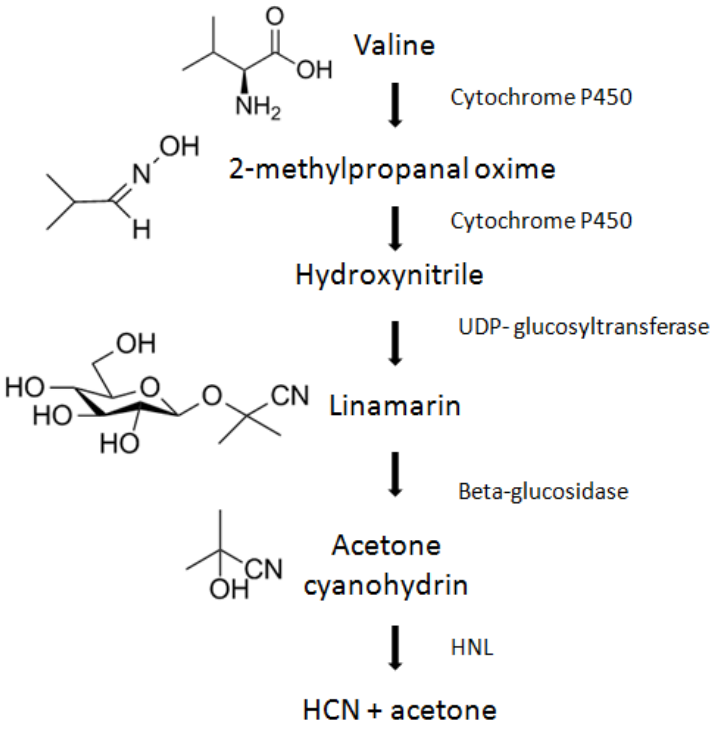


Figure 1.5. Cyanogenesis in rubber tree. Flow chart shows the role of hydroxynitrile lyase in the final step of cyanogenesis to produce HCN in rubber tree.

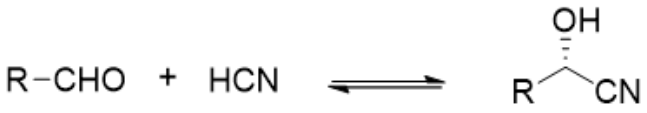
.

Although the original function of HbHNL is to produce HCN from acetone cyanohydrins for plant defense purpose, its reversible reaction to produce enantiopure cyanohydrins makes it an industrial relevant enzyme (Fig. 1.6a). Besides acetone, it is also able to take a bigger substrate such as benzaldehyde and turn it into mandelonitrile with optical purity up to 99%.[15] HbHNL is known to turn different aliphatic or aromatic aldehydes and methyl ketones into different cyanohydrins.[15] Cyanohydrins are useful precursors to agrochemicals and pharmaceutical products. For example, (R)-2-chlorobenzaldehyde cyanohydrin is the precursor of Clopidogel (Plavix), which is a blood clot inhibitor.[16] Pyrethroids, made from (S)-phenoxybenzaldehyde cyanohydrins, are an important synthesis route to insecticides.[16]

HbHNL catalyzes an unnatural reaction, nitroaldol reaction (Fig. 1.6b).[17,18] The nitroaldol reaction is an important industrial reaction as the enantiopure products could potentially be the precursors to useful intermediate such as 1,2-aminoalcohols and α-hydroxycarboxylic acids[17].

My overarching goals of this project are to: 1) improve the nitroaldol activity of HbHNL and HNL1, a reconstructed ancestral enzyme of HbHNL by using directed evolution; 2) determine which enzyme is more capable of producing better nitroaldolase by using suitable statistical analysis.

(a)



(b)

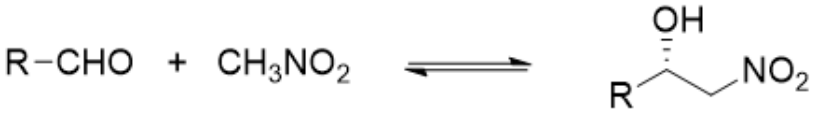


Figure 1.6. HNL (natural) and nitroaldol (unnatural) reaction catalyzed by hydroxynitrile lyase. (a) Natural HNL reaction (b) Unnatural nitroaldol reaction

1.6 Mechanism of modern hydroxynitrile lyases (HbHNL and MeHNL) and differences between HNL and esterase mechanisms

HbHNL and MeHNL share 77% sequence identity. HNL1 was reconstructed from these two enzymes along with other modern hydroxynitrile lyases. These three enzymes share the same catalytic triad consisting of Ser80, His235 and Asp207.[20,21] Gruber et al. 2004 concluded that there is no difference in their molecular mechanism between HbHNL and MeHNL. The root means square deviation (r.m.s.d) of superposition of crystal structures of acetone complexes is

0.42 for 253 common carbon atoms and r.m.s.d for atoms at active site is 0.23. The reaction mechanism of both HNLs starts with deprotonation of OH-group of Ser80 by His235 followed by deprotonation of the OH-group of the cyanohydrins. The O atom of the carbonyl compound is stabilized by hydrogen bonding to Thr11. Next, cyanohydrins cleavage happens with cyanide group of the substrate, held stably in the position by the positive charge of Lys236, acts as leaving group. At the last step, His235 reprotonates the cleaved cyanide and it is replaced by a water molecule.[20] Since HNL1 also has a similar active site, it is speculated that the mechanism is the same.

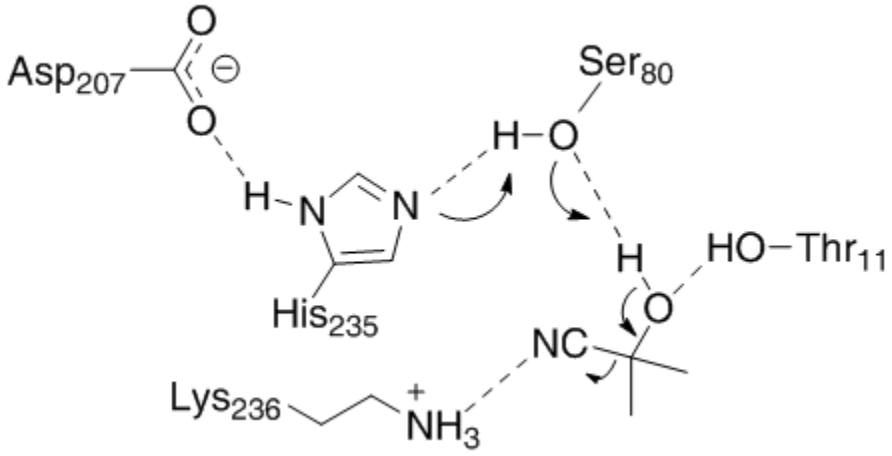


Figure 1.7. Catalytic mechanism of HbHNL, MeHNL and HNL1, the ancestral enzyme. Diagram was adapted from Gruber et al., 2004.

Since ancestral enzymes our group worked with were reconstructed from modern esterases and HNLs, it is interesting to point out the differences between esterases and HNLs. SABP2 could be turned into an HNL by only two amino acid substitution (G12T and M239K) at the active site by copying residues from HbHNL.[22] Conversely, an HNL could also be turned into an esterase by three amino acid substitutions (T11G, K239G & E79H) copied from SABP2.[23] Both mutants lost their original activities and gained new activities. They share 45% amino acid

sequence similarity, same catalytic triad and protein fold but catalyze very different reactions. By identifying the differences of the homologous enzymes focusing at the active sites, one enzyme could be turned into another enzyme by copying residues from the other enzyme, albeit with lower activity compared to the enzyme that was being copied. Mechanisms of both enzymes have been described so the differences in the mechanisms are being summarized in Table 1 below.

Table 1.1. Differences between HbHNL mechanism and SABP2 mechanism[22,23]

HbHNL	SABP2
A single transition state without involving acyl enzyme intermediate	Transition states involves short-lived acyl enzyme intermediate formation and hydrolysis
Oxyanion hole is blocked by threonine residue	Presence of oxyanion hole for tetrahedral intermediate stabilization
Serine reacts with the O atom of the cyanohydrin	Serine acts as a nucleophile that attacks the carbonyl carbon
Lysine is present to stabilize the negative charge of CN group	Lysine is not present

1.7 Ancestral enzyme and enzyme promiscuity

SABP2 and HbHNL are good examples of α/β hydrolases that share similar fold but catalyze very different reactions. A sequence identity of about 45% between the two also suggests that both share a common ancestral enzyme.[22,23] It is probable that this common ancestral enzyme is a promiscuous enzyme that is able to catalyze both reactions. There are two types of promiscuity such as substrate promiscuity and catalytic promiscuity. An enzyme is said to be substrate promiscuous when it is able to accept various substrates and carry out the same chemical reaction. On the other hand, an enzyme is said to be catalytic promiscuous when it is

able to accept different substrates, but catalyze different chemical reactions that involve different transition states and mechanisms.[56] This ancestral enzyme could be promiscuous in both ways.

As suggested by several groups, ancestral enzymes are the branch points for divergent evolution of novel functions. Model of divergent evolution, such as innovation, amplification and divergence (IAD) model, suggests that branch points are catalytically promiscuous. Ancestral enzyme may possess the ability to catalyze other reactions other than the main reaction. Such catalytic versatility is needed in the ancestral cell with limited enzyme resources.[24] When there is a change in ecological niche, one of the side reactions that the enzyme catalyzes may become important. Selection favors an increase of the new activity (Fig. 1.8).[25] Amplification of the same gene leads to more copies of the gene that encodes for new function since gene amplification is more likely than point mutation. New combinations between gene copies, made possible by recombination, can further improve functionality of some copies. The improved, extra copy is being subjected to selection to further improve its activity until one of the extra copies is able to carry out activity on its own while the other copy retains the original function. Original function may be lost if it imposes fitness cost. The remaining copies may also be lost due to drift or inactivating mutation. Therefore, ancestral enzymes are good places to find enzyme promiscuity. Promiscuous activity, albeit the level may be low, coupled with higher stability (which is generally true for reconstructed ancestral enzymes) makes it a good starting point for protein engineering if the extra function has a strong biotechnological interest. [25]

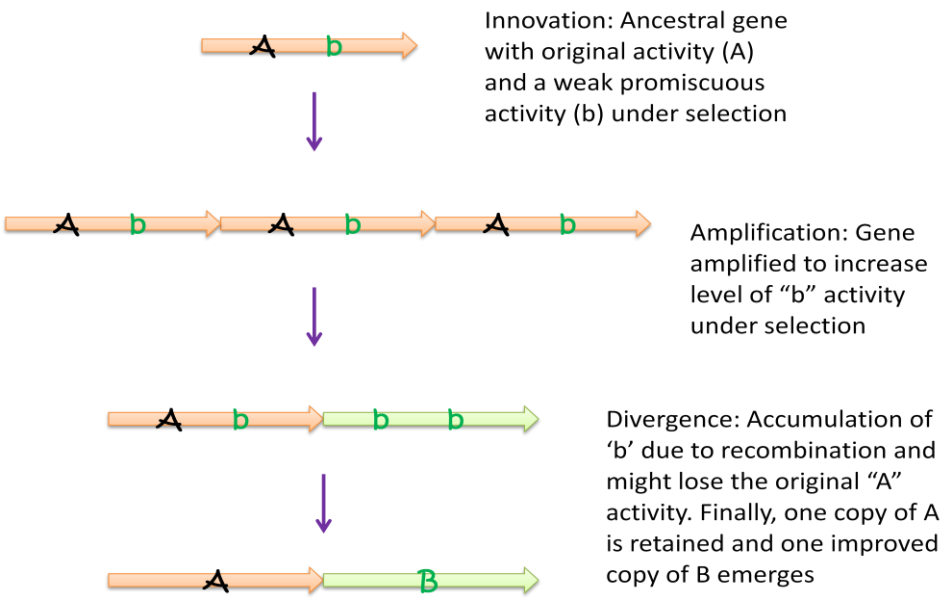


Figure 1.8 Innovation, amplification and divergence (IAD) model. Diagram shows how a promiscuous ancestral enzyme could eventually gain new function through IAD model. Diagram was adapted from Näsvall et al., 2012 with modification.

To test the hypothesis that ancestral enzymes are more promiscuous, ancestral β-lactamases were reconstructed. These ancestral β-lactamases corresponded to 2-3 billion year-old (Precambrian) nodes have ~ 100 amino acid replacements compared to the extant enzymes. They were found to be catalyzing the hydrolysis of various antibiotics with an average catalytic efficiency compared to the modern enzyme, which is a penicillin specialist.[27] Reconstructed ancestral α-glucosidases from modern isomaltase and maltase could accept a broader range of maltose and isomaltose as substrates.[28]

There are two major methods in building phylogenetics tree and reconstructing ancestral enzymes. Distance methods (also known as algorithmic methods) compute distance matrix that relates the number of differences between each pair of DNA or protein and the numerical data from the matrix is used to construct the tree.[30] Neighbor joining is one example of this method. On the other hand, character-based methods utilize sequence alignment and model of evolution to

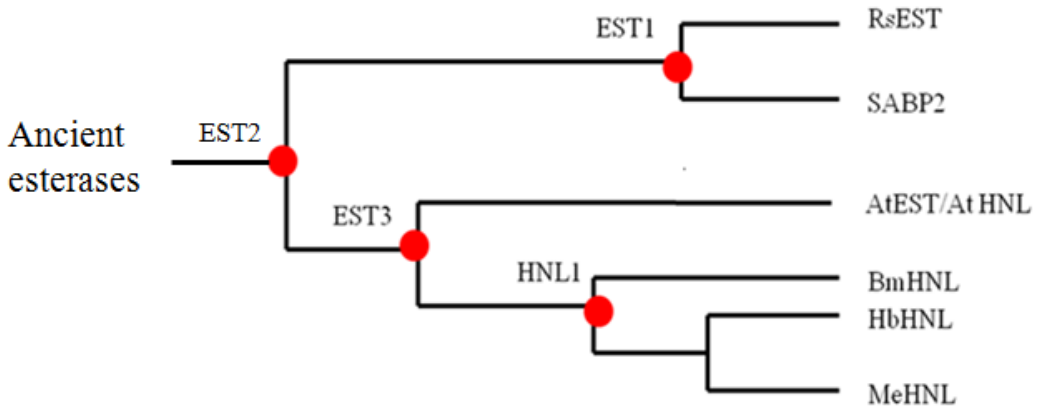
look for most probable tree.[30] Maximum parsimony and maximum likelihood are two most common character-based approaches. Maximum parsimony finds the tree by least amount of substitutions among sequences, assuming less evolutionary changes. Its performance can be compromised greatly when this assumption is not true. Maximum likelihood is a more robust method that takes specific probabilistic evolution models into account and search for most probable tree under these models.[30]

Professor Antony Dean used maximum parsimony, maximum likelihood or the combination of both as well as neighbor joining to reconstruct a few different versions of EST1, EST2, EST3 and HNL1 based on 3200 protein sequences that are 150-400 amino acids long and sharing at least 30% sequence identity with HbHNL.[29] A cladogram is shown at Fig. 1.9a, highlighting a few well-characterized modern enzymes. Phylogenetic reconstruction shows that these HNLs occupies smaller portion compared to esterases, suggesting that the HNLs might have evolved from esterases about 100 million years ago during the diversification of flowering plants.

My project is centered around HNL1, which is the common ancestor constructed from HbHNL, MeHNL and BmHNL. HNL1 is a transitional enzyme between HNLs and esterases and was constructed using maximum parsimony and maximum likelihood. HNL1 was expected to behave more like an HNL than an esterase since it shares higher sequence identity with HbHNL (79%). HNL1 was successfully crystallized by another graduate student, Bryan Jones, from our lab (not published). Structural alignment between HNL1 and HbHNL shows high similarity and it is not surprising since both have high sequence identity (Fig. 1.9b). Previous work has shown that it is the most promiscuous ancestral enzyme compared to all the other ancestral enzymes reconstructed by Professor Dean. It catalyzes aldehyde lyase and ester hydrolase reactions when tested against different ester and cyanohydrins as substrates.[29] It also catalyzes nitro-aldol reaction, which is an unnatural reaction, twice as good than HbHNL.

We hypothesized that it is easier to improve a new function in a promiscuous ancestral enzyme than in a modern enzyme. To test our hypothesis, HNl1 was chosen to represent the promiscuous ancestral enzyme since it was found to be more promiscuous than any other promiscuous enzymes our group has tested. We speculated that HNL1 is easier to improve in terms of nitroaldol activity, which represents the new function, compared to HbHNL. Methods of testing this hypothesis will be discussed in Chapter 3.

(a)



(b)

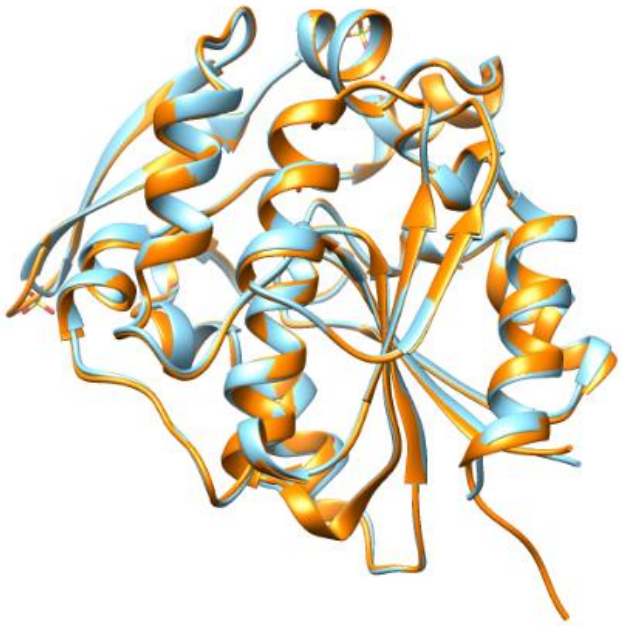


Figure 1.9. Cladogram shows relationship between modern and reconstructed ancestral enzymes and structural alignment between HNL1 and HbHNL. (a) Cladogram highlights four major reconstructed ancestral enzymes in red dots (HNL1, EST1, EST2 and EST3)evolving from ancient esterases belonging to α/β hydrolase family. Length of phylogenetic tree does not reflect time. (b) Structural alignment shows high similarity between HbHNL(blue) and HNL1(orange).

Chapter 2. Directed evolution of hydroxynitrile lyases

2.1 Introduction to protein engineering approaches

It takes millions of years of evolution for enzymes to be finely tuned to carry out their natural functions effectively in their hosts. Enzymes have various attractive properties that could be utilized in industrial applications. However, these enzymes may not work as effectively as desired outside of their natural physiological environment. There are many ways to modify an enzyme to further improve and obtain the desired properties such as stability, shelf life, tolerance to organic solvents, activity, binding, enantioselectivity to suit manufacturing processes. In general, two main strategies commonly employed to engineer proteins are rational design and direction evolution. The third strategy involves combining both strategies, is also called "focused directed evolution" or "semi-rational" approach.[33]

2.2 Rational design

One way to engineer a protein is through rational design. For example, structural-guided consensus approach and active site redesign are commonly employed to engineer a protein for desired properties. Structural guided consensus approach combines sequence and structural data and it is based on the hypothesis that sequences are conserved because they contribute to the fitness of the protein.[33] Therefore, changing a non-consensus amino acid to a conserved amino acid may improve the stability of the enzyme. It has been used to improve the unfolding temperature of a fungal phytase from 60.8 °C to 90.4 °C without compromising activity.[33,34] Aldolase activity was successfully introduced into a PLP-dependent racemase from *Geobacillus stearothermophilus* by redesigning the active site based on careful analysis of the mechanism. One mutation, Y265A was introduced to create more space for the aldolase substrate to bind.[35]

2.3 Directed evolution

In the mid and late 1990s, Pim Stemmer and Frances Arnold invented directed evolution where molecular biology techniques are used to modify enzymes in lab by mimicking Darwinian

evolution.[1] Random amino acid changes are introduced, followed by selection or screening for desired properties and the process is repeated through multiple rounds to accumulate beneficial mutations to obtain superior enzymes.[33] Another directed evolution method is gene shuffling, where a set of related sequences are randomly recombined.[33] One advantage of random approach is no prior structural knowledge of the enzyme is required and unpredictable positive outcome can usually be obtained. [33]

While random mutagenesis has been shown to be an effective method to improve an enzyme, it is highly labor intensive and an effective high throughput screening method is often required. [1,33] One could calculate the numbers of possible variants in a library using the formula below (where V_k= number of variants, L= number of amino acid in the protein, and k= number of amino acid substitution in the protein) to find out how many transformants need to be screened assuming that the library is unbiased.[31]

$$V_k = \frac{L!}{(L-k)!\,k!}19^k$$

According to the calculation, there are 5700 possible variants for one substitution in a protein with 300 amino acids but the number dramatically increases to 16,190,850 when there are two substitutions in the same protein. Usually, mutation rate is kept low so that activity of enzyme is not killed instantly. Low mutation rate produces smaller library with more redundancy (same variants need to be screened). However, bigger changes in protein properties usually require more amino acid changes.[32] If a huge improvement in a protein function is wanted, the mutation rate should be set higher to be assured that every clone that he/she screens is a unique variant if higher mutation rate is chosen. [32] One major drawback with random mutagenesis is that it may not be possible to screen every single variant in the library or find the best improved variants due to the lack of a good high throughput screening method. Usually, screening is

stopped when satisfying improved variants are found and those improved variant are subjected to another round of random mutagenesis, assuming that good mutations have additive effect.

To ensure that every single mutation in the protein is tested, one also can do an exhaustive search of improved variant using "complete saturation mutagenesis." Changing each site of a 214-amino acid cutinase to other 19 amino acids and 24 sites yielded improvement of 2-11 fold in stability in the presence of anionic surfactant dioctyl sulfosuccinate sodium salt (AOT) without losing activity compared to wild type. [36] This type of mutagenesis is highly labor-intensive. More often than not, an improved variant will be found because every single possibility is tested.

2.4 Semi-rational approach

One can eliminate the need to screen for a large library by using semi rational approach where both rational and random approaches are combined. Thermostability of *Bacillus subtilin* lipase A was improved by >490 times in terms of half life at 55 ℃ by using iterative saturation mutagenesis (ISM) to create "smart libraries" that significantly reduced the number of clones to screen.[37] Flexible regions of an enzyme is often targeted for mutagenesis as it indicates the weak interaction between side chains that may cause them to unfold in solution. Since thermophilic enzyme is more rigid than its mesophilic counterpart, it was hypothesized that thermostability of the mesophilic enzyme can be improved by increasing rigidity at certain flexible areas. These flexible areas of the protein can be found using B-factor, which is an atomic replacement parameter of the X-ray data. Residues with higher B factor (or B-value) have higher flexibility and thermal motion is shown as atomic electron densities smearing. [37] These residues were targeted to be mutated to other 19 amino acids by using NNK degenerate codons. Best hits were used as templates for subsequent saturation mutagenesis at a position that gave second best hit from the first round. The best hits with two mutations from the first two best positions was used

as a template for the subsequent SSM on the third best hit from the first round and the process was repeated iteratively until an improved variant with seven mutations that improved the half life by >490 was found.[37] By doing this, they were able to accumulate all the beneficial mutations that eventually led to a much improved variant. The assumption that beneficial mutations are additive worked well in this case.

2.5 Construction of SSM libraries and screening effort

To construct a library to randomize one specific spot to test all possible 20 amino acids, degenerate codons can be used. The most straightforward degenerate codons are NNN, where N represents each of the nucleotide, which consists of 64 codons that encode for all 20 amino acids.[33] Three of the codons are stop codons. It is highly redundant and increases screening effort. One method to reduce redundancy by half is to use NNK (K= T/G) or NNS (S=G/C) degenerate codons. Number of codons is 32 instead of 64 (Fig. 2.1) and the stop codon only occurs once. It greatly reduces screening effort. Only one set of primers is needed. It is one of the most common ways to construct SSM library. Other way to reduce screening effort with no redundancy is to use 20 different sets of primers for each library but it can get costly when more than one library is constructed. For my project, I needed to construct 12 different SSM libraries (six for each enzyme) so NNK degenerate codons were obvious choice for me.

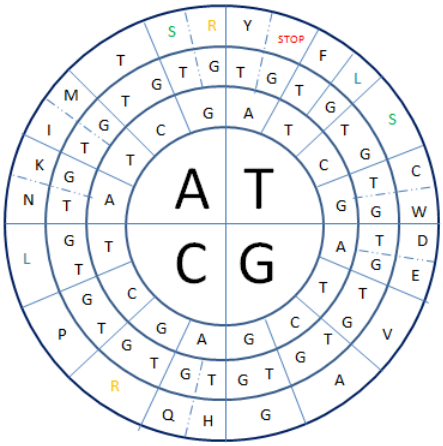


Figure 2.1: NNK degenerate codons encoding 20 amino acids shown in wheel format. Figure was modified from Kille et al., 2012

Screening effort refers to the number of clones needs to be screened to analyze the 'protein sequence space.' To know how many clones to screen, we first have to know the library coverage and the probability of screening all of them. Library coverage tells us about the total number of possible variants in a library. These factors have been defined mathematically and the formula is shown below:

$$L = -C^m \ln(1 - F)$$

where L= number of clones to be screened, C= total number of codons, m=number of site being saturated and F= completeness of the library in terms of proportion.[38] For example, to have 99% probability of covering even the rarest amino acid in a library with one NNK degenerate codon, 147 clones [$L = -32^1 \ln(0.01)$] need to be screened.

2.6 Improving nitroaldol activity of HNLs using both directed evolution and semi-rational methods

I used both directed evolution and site saturation mutagenesis methods to improve nitroaldol activity of HbHNL and HNL1. Nitroaldol reaction (also known as Henry reaction) is an unnatural reaction catalyzed by both HbHNL and HNL1. Nitroalkanes have similar in molecular sizes and pk_a (pk_a ≈ 9) compared to cyanohydrins. Both should have similar mechanism.[17]

HbHNL catalyzes nitroaldol reaction using 1-nitro-2-phenylethanol (NPE) as substrate with 100 times lower catalytic efficiency. However, HNL1 is 2 times better than HbHNL in catalyzing this unnatural reaction with 10 times reduction on its HNL activity using mandelonitrile (MNN) as substrate, making it a more specific nitroaldolase as a whole.

Table 2.1. Catalytic efficiency (k_{cat}/K_M) for HbHNL and HNL1 using mandelonitrile (MNN) as HNL substrate and 2-nitro-1-phenylethanol (NPE) as nitro-aldol substrate.

Enzyme	MNN k_{cat}/K_M ($mM^{-1}\cdot S^{-1}$)	NPE k_{cat}/K_M ($mM^{-1}\cdot S^{-1}$)
HbHNL	14700	109
HNL1	1200	209

A promiscuous enzyme usually has higher plasticity (meaning that it is easier to confer new activity just by a few amino acid changes) and easier to evolve.[51,52] Plastic residues are also more likely to be found at the active site region[51], giving us more reason to choose active site resides to mutate. We hypothesized that residues in the active site are important in improving a promiscuous activity such as nitroaldol activity since active site residues have more contact with the substrate and are probably more plastic to accommodate the substrate for binding. Alissa, the ex-post doc from my lab has tried copying residues from the active site of HNL1 to HbHNL in the hope of improving its nitroaldol activity using NPE as substrate. The attempt was fruitless. Hence, using semi-rational approach was an obvious step for us to increase our chances of getting improved variants. We also hypothesized that HNL1 is a better starting point to evolve a new function because it is a more promiscuous enzyme.

Six positions (11, 14, 79, 106, 153 and 178) within 6 Å in the active site were selected to be mutated to other 19 amino acids (Figure 2.2). Changes at position 11 make more space for the substrate. Position 14 is close to where CN or NO_2 group binds so a different residue might also make more space for the NO_2 group. Removing a negatively charged group from position 79

might also help the NO_2 group to bind better. Other positions do not interact directly with the substrate. They were chosen because these residues differ between HbHNL and HNL1.

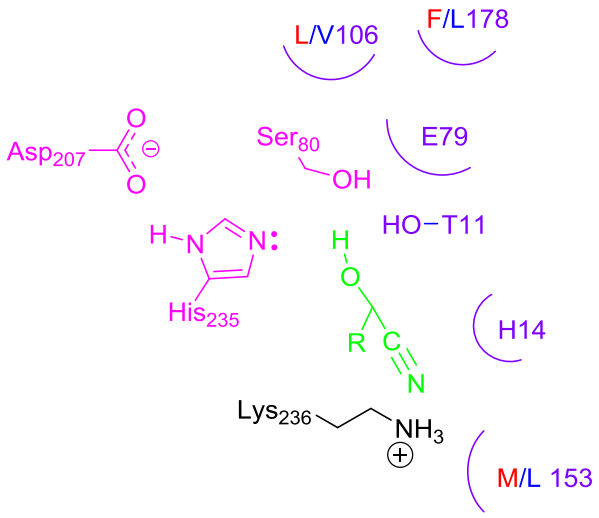


Figure 2.2. The active site of HbHNL and HNL1. Residues colored in purple were subjected to SSM. Red and blue residues represent the original residues in HbHNL and HNL1 respectively. Catalytic triad was colored in magenta while the substrate is colored in green.

To construct site saturation mutagenesis libraries, I used degenerate NNK codons (N= A/T/C/G and K= T/G) and screened about 130 clones for HNL1 and 168 clones for HbHNL SSM libraries. Screening was carried out at different times. I started screening HNL1 first and then HbHNL because of its stability (or more specifically, solubility) problem. I oversampled HbHNL by screening 2 plates instead of 1.5 plates. I was able to cover at least 95% of the library by screening 130-168 clones for each library.

Error prone PCR libraries were also constructed to make mutations all over the enzymes including the active site. I decided to keep the mutation rate to about 2-3 mutations/gene so that most of the variants were still active. We were curious to know if residues farther from active sites could be contributing to improving nitroaldol activity as well.

Chapter 3. High throughput screening methods

3.1 Selection

After one gets passed the molecular cloning to construct a huge library of mutants, it is time to screen them for the desired activities in a high throughput manner. The fastest and most high throughput way is selection. Selection links the growth of the host with acquired enzymatic activities.[53] One common selection example is to clone β lactamase that confers ampicillin resistance on plasmids by breaking down ampicillin to enable the hosts to survive on growth medium containing ampicilin.[40] Each colony harbors the plasmid being transformed in the host because the ones that do not harbor the plasmid will not grow. Another interesting example to improve enantioselectivity of a Candida antartica B (CalB) by utilizing a model reaction involving kinetic resolution of isopropylidede glycerol (IPG) has been reported. CalB mutant that favors formation of the desired enantiomer (in this case, the R-enantiomer) will survive since R-enantiomer can be used as a carbon source. However, if a mutant favors the formation of S-enantiomer, a poisonous side product (fluoroacetic acid) will form and kill the host cell.[39] A good selection technique allows one to screen large libraries in short period of time but it often takes very long time to design a useful selection method as it often requires manipulation of the metabolism of the host organism. So far, no selection method has been reported to detect HNL or nitroaldol activity.

3.2 Agar plate screening

It is difficult to create a good selection method since it normally involves complicated cloning processes to link growth and enzymatic reaction. Another way is to detect activity directly at the colony level, also known as agar plate screening. One can grow a few hundred of colonies expressing protein with desired enzyme activity and test it directly on an agar plate. Upon reaction with substrate, active or improved variants should display color or fluorescence

that can be easily detected. A high throughput agar plate screening for hydroxynitrile lyase at colony level based on detection of hydrocyanic acid (HCN) instead of detection of aldehyde has been reported.[41] The advantage of this approach is that multiple substrates can be tested as long as they produce HCN. A membrane with blotted colonies is placed on substrate solution. Then an HCN sensitive detection paper is pressed on membrane blotted colonies with a nylon tissue in between. Dark blue spots will appear when HCN is liberated from the enzymatic reaction with aldehyde.[41]

One limitation with this method is that the enzymes may not be active enough to show activity that result in change in color at the colony level, even if T7 expression system that is known to be leaky is used. More often than not, enzymes still need to be induced with IPTG to show detectable activity on the plate. It is also difficult to control protein expression level since not all variants are expressed at the same level. Some variants that express better may be mistaken as hits. Sometimes, the enzyme may not be secreted outside of the cells so the cells need to be lysed before incubating them with the substrate. A replicate plate needs to be made because it is not possible to recover lysed cells. Agar plate screening is good in identifying active clones but is not effective in estimating catalytic rate of enzyme variants. It is useful to be used as a primary screen as one can go through many clones in relatively short amount of time to eliminate "dead" clones for subsequent screening. Therefore, an agar plate screening to detect nitroaldol activity was not developed due to these drawbacks.

3.3 Microtiter plate (MTP) screening

The most common and practical form of screening is microtiter plate screening. Single colonies are picked and grown in 96 well plates. Some bacterial culture is usually transferred to a

second plate while the first plate is kept as master plate or back-up. The second plate is tested with substrate after protein induction and cell lysis (Figure 3.1).

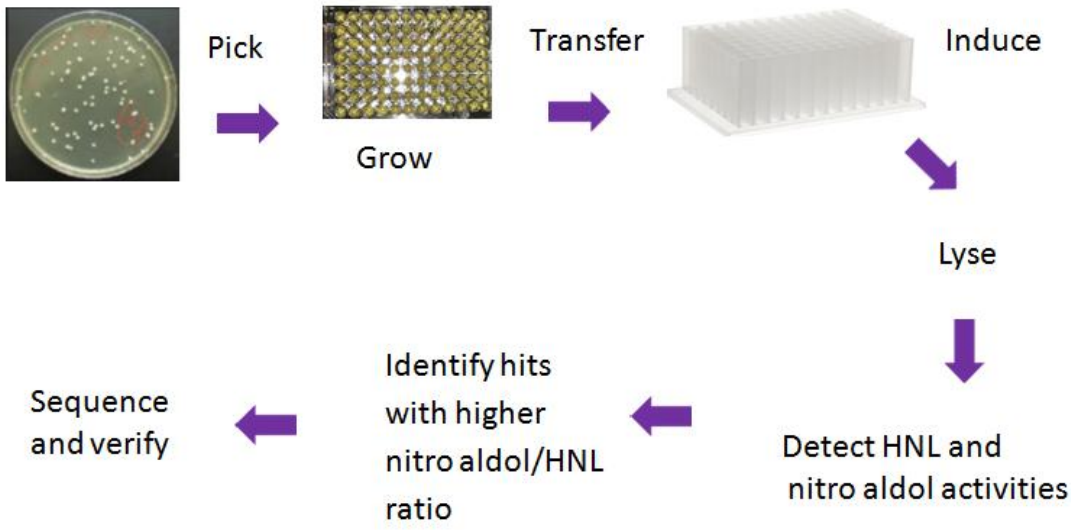


Figure 3.1 Work flow for the high throughput screening in this project.

Screening MeHNL for HNL activity using this method was reported.[42] Cell free extract/supernatant was obtained by centrifuging lysed cells. Mandelonitrile was used as substrate and formation of benzaldehyde was monitored using UV spectrophotometer at 280 nm. This method only worked as MeHNL exhibits high HNL activity so a little cell free extract was enough to detect increased activity. Since protein also absorbs at 280 nm, this method may not work as well if more protein is used for the screen as Beer Lambert law only applies when the OD is less than 1. However, detecting benzaldehyde formation remains one of the most convenient and straightforward ways for high throughput screening for HNL activity.

Another high throughput screening method for HNL that allows detection of wider range of HNL substrates was reported.[43] Cyanohydrins that have been tested for the study are acetaldehyde cyanohydrin, propionaldehyde cyanohydrin, benzaldehyde cyanohydrin, acetone

cyanohydrin, cyclohexane cyanohydrin and 3-phenoxybenzaldehyde cyanohydrin.[43] Cyanohydrin is added to enzyme solution and is allowed to incubate at room temperature for 5 minutes. Mixture of N-chlorosuccinimide/succinimide is added to oxidize CN^- to CN^+. Then, mixture of isonicotinic acid/barbituric acid is added for colorimetric detection. The color formation from dialdehyde is monitored spectrophotometrically over 20 minutes at 600 nm.[43] This method is useful to measure HNL activity but cannot be used to detect nitroaldol activity since the reaction does not produce cyanide.

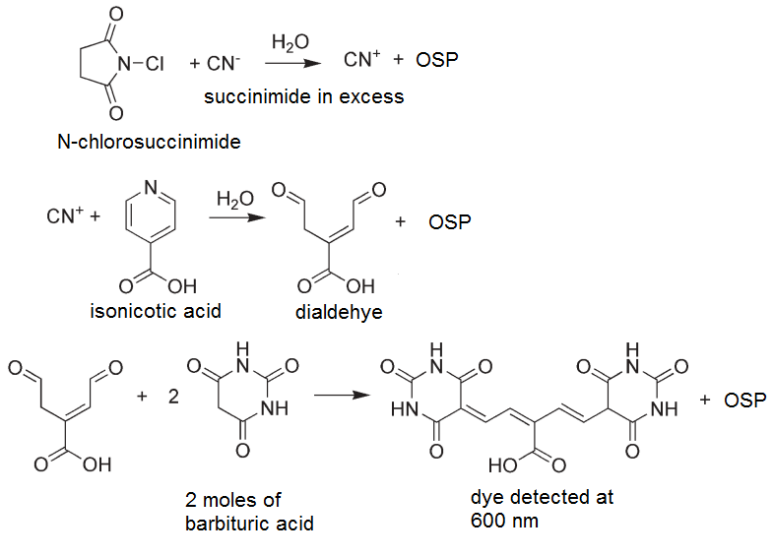


Figure 3.2. An agar plate screening method that detects CN^+ formation. Diagram adapted from Andexer et al., 2006 shows the chemical reaction that leads to CN detection using different types of substrates.[43] OSP = other side product

3.4 Screening method for HNL and nitroaldol activity

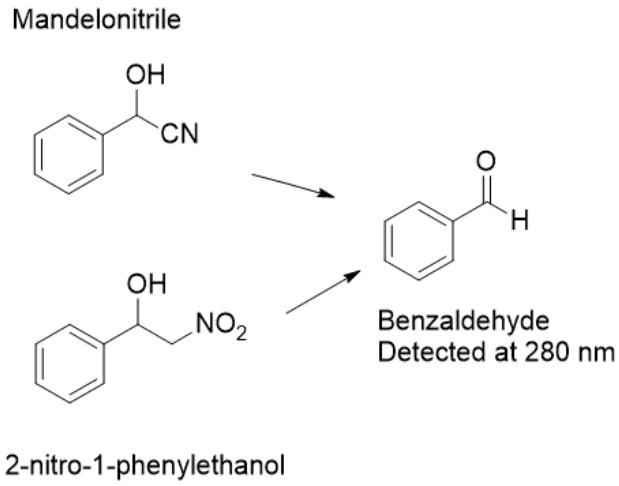
Since both HNL and nitroaldol reactions lead to same product, the same activity detection method could be used to identify clones that exhibit higher activity than WT. Mandelonitrile is a good substrate (k_{cat} =1530 min^{-1}) for WT HbHNL. Both 2-nitro-1-phenylethanol and mandelonitrile lead to the same product, benzaldehyde that can be detected at 280 nm (Fig. 3.3a).

However, after a few trials with these substrates, I concluded that this method is not suitable for high throughput screening because both enzymes have weak nitroaldol activity. A lot more enzyme was needed for nitroaldol activity detection. Since more unpurified cell free extract was used in screening, it led to high initial OD because E.coli proteins absorb at 280 nm as well. Beer Lambert law only applies in the range of 0-1 but the absorbance was outside the range. So, it was impossible to tell which clone was more active than the other. Since assay was run in low pH, some of more unstable proteins may precipitate during the screen, leading to uninterpretable results as precipitation increases the OD at 280 nm.

A more sensitive assay that is not affected by protein concentration was needed for high throughput screening. So bigger substrates such as 2-hydroxy-2-(6-methoxynaphthlen-2-yl)acetonitrile and 1-(6-methoxynaphthen-2-yl)-2-nitroethanol-1-ol were used instead. Both produce a highly conjugated, fluorescent product 6-methoxy-2-napthaldehyde, which can be easily detected at excitation wavelength of 311 nm and emission wavelength of 448 nm (Fig. 3.3b). This compound can also be detected at 315 nm using UV spectrophotometer.

This is not the first fluorimetric assay that detects 6-methoxy-2-napthaldehyde. An epoxide hydrolase was reported to catalyze the epoxide hydrolysis to its corresponding diol, liberated a cyanohydrin intermediate that quickly decomposed into 6-methoxy-2-napthaldehyde that was detected in the same way.[44] This compound has high stokes shift (difference between emission and excitation spectra maxima) that provides good detection with low noise. It is also stable in water so the fluorescence signal is reliable for the duration of a kinetic experiment (<15 min). Not only the fluorescence assay was more sensitive, it was less likely to be affected by the amount of protein in supernatant. So, all library screening and characterization were done by detecting the 6-methoxy-2-napthaldehyde using the fluorescent plate reader.

(a)



(b)

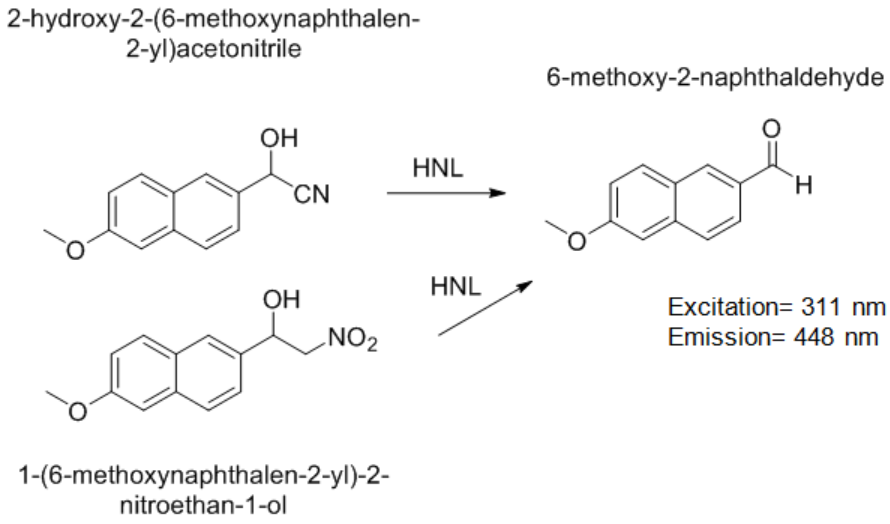


Figure 3.3. Substrates used in high throughput screening. (a) Initial choice of substrates for HNL and nitroaldol reactions that lead to production of benzaldehyde. (b) Final choice of substrates used for HNL and nitroaldol activity detection that lead to production of 6-methoxy-2-napthaldehyde.

After optimizing that growth condition and screening method for HNL1, I realized that I was not able to detect any activity with WT HbHNL with the same growth condition and screening method due to lack of solubility of HbHNL. I improved the solubility of the enzyme by a single mutation, H103V. After solubility problem was fixed, HNL and nitroaldol activities could be easily detected by using the bigger substrates. Details will be covered in Chapter 5.1.

Chapter 4. Experimental section

4.1 Site Saturation Mutagenesis on position H103 for HbHNL

Site saturation mutagenesis (SSM) was carried out on position H103 in HbHNL to improve solubility of the enzyme for high throughput screening. Partially overlapping forward primer, 5'- GCT GTT TTC NNK AAT TCA GTA TTG CCA GAC-3' and reverse primer, 5'- CTG AAT TMN NGA AAA CAG CAG CTG CAA TC -3' were designed according Zheng et al. 2004. The reaction mixture for PCR reaction contained 5X Q5 reaction buffer, 10 mM of each dATP, dGTP, dCTP and dTTP, 10 μM of each forward and reverse primer, Q5 High Fidelity DNA polymerase (New England Biolabs) and 8 ng of plasmid DNA in a 50 μL reaction. Thirty cycles of PCR were carried out at 98 °C for 10 s, 50-70 °C for 30 s, 72 °C for 5 min with final extension at 72 °C for 10 min. PCR product was digested with *dpn*1(New England BioLabs) at 37 °C for four hours to remove parental template and transformed into chemically competent *E.coli* BL21(DE3). About 80 colonies were picked and screened for higher activities in both HNL and nitroaldol activities for each library. Improved variants were purified and their solubility was confirmed on mini protein gel. For HbHNL, H103V was chosen to be used as template for SSM to improve nitroaldol activity.

4.2 Site-Saturation Mutagenesis of HbHNL-H103V and HNL1

Site saturation mutagenesis was carried out on positions 11, 14, 79, 106, 153 and 178 of HbHNL-H103V and HNL1. Partially overlapping mutagenic primers (Table 4.1) were designed according to Zheng et al., 2004 except for position L106 for HNL1 where completely overlapping primers were used.[55] The PCR and subsequent digestion and transformation were the same as described in SSM for position 103 for HbHNL above.

Table 4.1. Primers used in SSM libraries for the six positions for each enzyme. (a) Forward and reverse primers for SSM of HbHNL-H103V (b) Forward and reverse primers for site saturation mutagenesis of HNL1.

(a)

Primer	Sequence (5'-3')
T11X F	CATTTTGTTCTTATTCATNNKATATGCCACGGTGCATG
T11X R	GTGGCATATMNNATGAATAAGAACAAAATGAGCGAATGC
H14X F	CTTATTCATACCATATGCNNKGGTGCATGGATTTGGC
H14X R	GCACCMNNGCATATGGTATGAATAAGAACAAAATGAGC
E79X F	GATTCTGGTTGGCNNKAGCTGTGGAGGAC
E79X R	CAGCTMNNGCCAACCAGAATCACCTTTTC
V106X F	GTTTTCGTGAATTCANNKTTGCCAGACACCGAG
V106X R	CTGGCAAMNNTGAATTCACGAAAACAGCAGCTG
L153X F	CTTCACGCTTNNKAGGGAAAATTTATATACCCTTTGCG
L153X R	GTATATAAATTTTCCCTMNNAAGCGTGAAGCCCAGTTTT
L178X F	GACAAGGAAGGGATCANNKTTTCAAAATATTTTAGC
L178X R	GAAAMNNTGATCCCTTCCTTGTCAACATCTTCG

(b)

Primer	Sequence (5'-3')
T11X F	CATTTTGTATTAATTCACNNKATATGCCACGGCGCATGG
T11X R	CGTGGCATATM NNGTGAATTAATACAAAATGAGCGGTTGCC
H14X F	GTATTAATTCACACTATATGC NNKGGCGCATGGATT TGGTATAAG
H14X R	CATGCGCCMNNGCATATAGTGTGAATTAATACAAAATGAGC
E79X F	GTAATACTTGTTGGTNNKTCTTGTGGAGGACTG
E79X R	CACAAGAMNNACCAACAAGTATTACCTTTTCTCCTTGG
L106X F	GGTGTTTCATAATGCCNNKATGCCGGATACAGAAC
L106X R	GTTCTGTATCCGGCATMNNGGCATTATGAAACACC
M153X F	GCTTTAAATTGNNKCGGGAAAACTTGTACACAAACTGCCCAATCG
M153X R	GTGTACAAGTTTTCCCGMNNCAATTTAAAGCCAAGTTTCATGCCG
F178X F	GACGCGTAAAGGGAGTNNKTTTCAGAACGATTTGGC
F178X R	CTGAAAMNNACTCCCTTTACGCGTCAGCATTTTTGC

4.3 Error prone PCR for HNL1 and HbHNL

Same primer set (non-overlapping forward and reverse primers) was designed for HbHNL and HNL1 to flank the gene of interest. Forward primer is 5'-

CCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAG-3' and reverse primer is 5'-CAGCAAAAAACCCCTCAAGACCCGTTTAG-3'. The reaction mixture for PCR reaction contained 10X Mutazyme II reaction buffer, 10 mM of each dATP, dGTP, dCTP and dTTP, 2.5 μM of each forward and reverse primer, Mutazyme II DNA polymerase (New England Biolabs) and 44 ng of plasmid DNA in a 50 μL reaction. Thirty cycles of PCR were carried out at 95 °C for 60 s, 55 °C for 60 s, 72 °C for 1 min with final extension at 72 °C for 10 min. PCR product was digested with *dpn*1(New England BioLabs) at 37 °C for four hours to remove parental template.

4.4 Gibson assembly to ligate vector and insert containing random mutations

Compared to traditional cloning method to ligate vector and insert, Gibson Assembly has shown to be a more successful ligation method based on my experience. HbHNL plasmid was cut by two different restriction enzymes, StyI-HF and NdeI (New England BioLabs) to create an empty vector with sticky ends. PCR product was not cut as the primer design has included the complementary sites that helped PCR products to stick to vector. To assemble cut vector and PCR product, 2X Gibson Assembly Master Mix (New England Biolabs) was used. Reaction mix was incubated for 30 minutes at 50 °C. The Gibson Assembly Master Mix contains an exonuclease that degrades double stranded DNA from 5' end and leads to a 3' overhang that allowed complementary fragment to anneal together. The enzyme was quickly inactivated at 50 °C to avoid over-'chewing'. DNA polymerase replaced the nucleotide within the gap in the annealed fragments. Finally, a ligase repaired the nick and yielded complete plasmids harboring the mutated insert. The reaction mixture was transferred to 0.025 μm membrane (EMD Millipore) to perform micro-dialysis to remove excess salt prior to electroporation. Reaction mixture was dialyzed for 1-2 hours and final product was transformed into BL21 cells using Micropulser electroporator (Bio-Rad Laboratories, Inc) with 'EC2' or 'EC3' setting.

4.5 Cell Culture in 96 well plates

Transformants were plated onto Luria-Bertani (LB) agar plates supplemented with 100 μg/mL of ampicillin and incubated at 37 °C overnight. Ninety six clones including WT (occupied 12 wells) were inoculated into 96 well plates containing LB with 100 μg/mL of ampicillin. Plates were shaken (700 rpm) at 37 °C overnight for bacterial growth. Twenty microliter of bacterial culture from the 96 well plates, which were also master plates, was transferred into 96 deep well plates containing 1 mL of terrific broth (TB) with 100 μg/mL of ampicillin in each well and was grown for about 3 h at 37 °C. Master plates were kept in freezer after adding 25% glycerol as cryoprotectant. Final concentration of 80 μM of isopropyl-β-D-thiogalactopyranoside (IPTG) was added into each well in 96 deep well plate to induce protein expression. Cells were shaken (700 rpm) at 17 °C overnight. Cells were centrifuged, washed with phosphate buffered saline (PBS) once, spun down at 3000 rpm for 20 min and the supernatant was removed. The cells were then lysed with 100 μL Bugbuster protein extraction reagent (Novagen, Merck Millipore) for 1 h. Six hundred microliter and 400 μL of 50 mM citrate phosphate buffer (pH 6.0) was added to the HNL1 and HbHNL lysates respectively. Lysate was then centrifuged at 300 rpm for 20 minutes to obtain clear supernatant containing the desired protein for screening.

4.6 Screening for hydroxynitrile lyase and nitroaldol activities and selection of hits

Supernatant (7 μL for HNL1, 5 μL for HbHNL) was added into reaction containing 100 μM 2-hydroxy-2-(6-methoxynaphthalen-2-yl)acetonitrile and sodium citrate buffer (50 mM, pH 5.0) to detect HNL activity. To detect nitroaldol activity, 1-(6-methoxynapthlen-2-yl)-2-nitroethanol with final concentration of 100 μM was added into 120 μL of supernatant (pH 6.0). Both reactions led to fluorescence of 6-methoxy-2-napthaldehyde, which was measured at excitation wavelength at 311 nm and emission wavelength at 448 nm using Spectra Max Gemini XS microplate reader (Molecular Devices) for 10 min. Activity was measured in mRFU/min.

Ratio of nitroaldol activity to HNL activity was calculated for each clone. Clones that exhibited 3*standard deviation of WT ratio or higher were selected as hits of the library. Clones that have noticeable higher nitroaldol activity (regardless of whether it has higher ratio) are also selected to be tested just in case they are improved variants for both reactions. Basically, three out of four categories of variants were selected as hits and the reasoning is summarized in the Table 4.2.

Table 4.2. Four different categories of clones. Hits were selected from all categories except "low ratio" and "low nitroaldol activity" category. Low nitroaldol activity is between 0.2 and up to 1+3 standard deviation (data was normalized to WT. Value of 1=average WT activity). Any value higher than 1+3 standard deviation is considered "high nitroaldol activity." Ratio is also defined in the same way. Ratio lower than 1+3 standard deviation is also considered low and vice versa.

	High nitroaldol activity	**Low nitroaldol activity**
High ratio	High activity indicated good expression/solubility or actual increase of high nitroaldol activity. High ratio indicated loss of HNL activity. Clones in this category were selected as hits.	Low activity might due to low expression/solubility or actual decrease of nitroaldol activity. To be sure no hit will be missed due to lack of expression/solubility, clones fell in this category were still selected as hits unless it was completely inactive (<20% WT activity).
Low ratio	High activity indicated good expression/solubility. Low ratio may be due to increase of HNL activity, which makes it an overall better enzyme. Clones in this category were selected as hits.	Clones in this category were not selected as hits. Most clones fell in this category.

4.7 Enzyme purification

Hits were streaked on LB-Amp agar plates from the frozen master plates. Colonies were picked and inoculated in 3 mL of LB with 100 μg/mL ampicillin in culture tubes. Cells were grown overnight at 37 °C and 200 rpm. One milliliter of culture was transferred into 100 mL terrific broth supplemented with 100 μg/mL ampicillin and grown at 37 °C until its optical density reached 1. Cell culture temperature was reduced to 17 °C and 800 μM IPTG was added to induce protein expression for 20 h. Cells were harvested at 8000 rpm for 20 min at 4 °C and were resuspended in lysis buffer (50 mM sodium phosphate, 500 mM NaCl and 10 mM imidazole, pH 8.0). Cells were lysed using a sonicator (Branson Digital Sonicator) in ice bath with 3 s of pulses alternating with 5 s of cooling time at 40% amplitude for 5 minutes. Lysed cells were centrifuged at 18,000 rpm for 40 min to obtain supernatant containing the protein of interest. Using gravity columns, enzymes were purified from their supernatants using 1-2 mL regenerated Ni-NTA resin. Columns were washed with 50 mL of 10 mM, 20mM and 50 mM of imidazole in buffer containing 50 mM sodium phosphate, 500 mM NaCl at pH 8.0. Protein was eluted at 150 mM of imidazole containing the same salts at pH 8.0. Enzymes were then concentrated and the buffer was exchanged with 5 mM N,N-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid (BES) buffer using centrifugal filter units (Millipore).

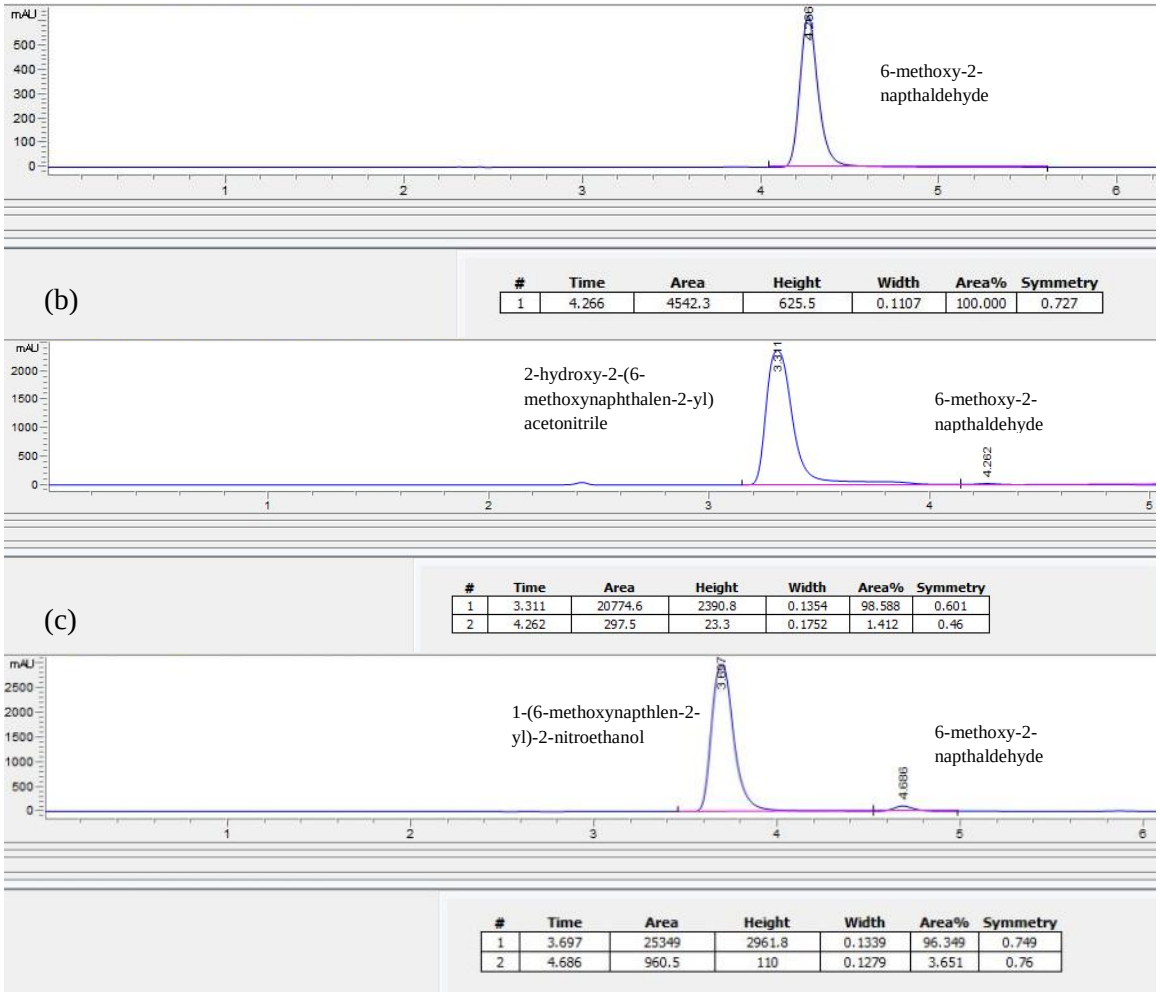
4.8 2-hydroxy-2-(6-methoxynaphthalen-2-yl) acetonitrile and 1-(6-methoxynapthlen-2-yl)-2-nitroethanol synthesis

Substrates used in experiment are not commercially available so they had to be synthesized. Synthesis of 2-hydroxy-2-(6-methoxynaphthalen-2-yl) acetonitrile was adapted from Bunyan et al., 2009.[46] About 3-5 equivalent of NaCN was added directly to 1 equivalent of 6-methoxy-2-naphthaldehyde dissolved in acetic acid. Reaction was carried in the closed fume hood at room temperature (22 °C) for a day and stopped when more product was present

compared to the starting material, as indicated by TLC. Reaction mixture was neutralized using $NaHCO_3$ until ~pH 5 and was extracted using ether/dichloromethane and dried using Na_2SO_4. The compound was then purified using silica column chromatography using 80% of hexane and 20% of ethyl acetate. The purity of product was examined using TLC (which has shown to be unreliable for this compound since it reverts back to starting material on the plate itself) and HPLC. The purity was about 96% based on HPLC (Figure 4.1b). NMR also revealed no starting material. In later preparation, I realized that if reaction was left running for much longer time (2 weeks), relatively pure product (97%) formed and no purification was needed (result not shown). It is better not to purify the compound because this compound is highly unstable and it tends to revert back to starting material during purification that usually takes more than 10 hours, resulting in at least 90% loss of yield.

The protocol of synthesizing 1-(6-methoxynapthlen-2-yl)-2-nitroethanol has not been reported in literature and was improvised by Titu, a chemist from my lab. The starting material 6-methoxy-2-napthaldehyde was dissolved in 10 equivalent of nitromethane in a round bottom flask. Triethylamine (1 equivalent) was added to the reaction and reaction was carried out in cold room overnight. The product formation was monitored and confirmed using TLC plate. Reaction was stopped before a side product (dehydrated compound) was formed. The pH of the reaction was adjusted to 7 using HCL before extracting the final product using ether. The compound was purified the same way using silica gel chromatography. The purity of the compound was 99% according to HPLC (Fig. 4.1c).

(a)



#	Time	Area	Height	Width	Area%	Symmetry
1	4.266	4542.3	625.5	0.1107	100.000	0.727

#	Time	Area	Height	Width	Area%	Symmetry
1	3.311	20774.6	2390.8	0.1354	98.588	0.601
2	4.262	297.5	23.3	0.1752	1.412	0.46

#	Time	Area	Height	Width	Area%	Symmetry
1	3.697	25349	2961.8	0.1339	96.349	0.749
2	4.686	960.5	110	0.1279	3.651	0.76

Figure 4.1. HPLC profile for the synthesized substrates. (a) 6-methoxy-2-napthaldehyde, the starting material (b) Synthesized 2-hydroxy-2-(6-methoxynaphthalen-2-yl) acetonitrile with 99% purity (c) Synthesized 1-(6-methoxynapthlen-2-yl)-2-nitroethanol with 96% purity

4.9 Enzyme characterization for nitroaldol activity

Formation of 6-methoxy-2-napthaldehyde was measured at excitation wavelength at 311 nm and emission wavelength at 448 nm at room temperature using Spectra Max Gemini XS microplate reader. Assay was performed with total reaction volume of 100 µL consisting of 50 mM sodium citrate buffer, 2-10 µg enzyme, 0-900 µM of 1-(6-methoxynapthlen-2-yl)-2-

nitroethanol and 7% DMSO on black 96 well plates (Greiner) at pH 5.75 for nitro-aldol activity. Experiment was run in triplicate and a blank with no enzyme to account for spontaneous reaction. A standard curve of 6-methoxy-2-napthaldehyde was plotted and its slope (RFU/µM) was used to determine product formation over time (µmol/min or U) from mRFU/min (Fig. 4.2). Higher substrate concentration was not used in these assays as fluorescence signal was be out of range of the detection limit (out of what can be predicted by the standard curve) with higher substrate concentration. Standard curve was only straight up to 90 µM, so additional data points at higher concentration was omitted. It was assumed that no more than 90 µM will be made in the duration of characterization.

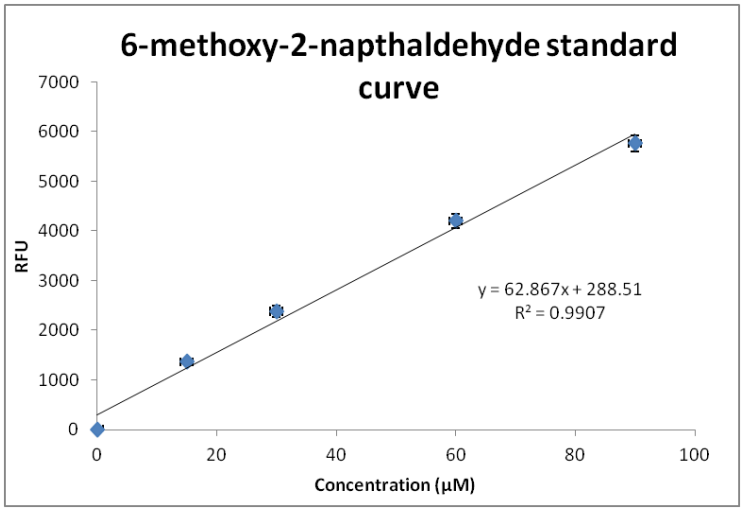


Figure 4.2. Plot shows 6-methoxy-2-napthaldehyde standard curve in the range of 0-90 µM concentrations. Data points at higher concentration were omitted because the curve tapers off at a concentration more than 100 µM.

Steady state kinetic constants were determined by fitting data non-linearly to Michelis Menten equation using OriginPro 7.5 (OriginLab). For enzymes that were inhibited by the substrates, data was fitted using equation below. Besides k_{cat} and K_M, K_I value was also obtained by fitting the data to substrate inhibition curve. These values were reported in the result section.

$$V_0 = \frac{V_m[S]}{K_M + [S] + \frac{[S]^2}{K_I}}$$

4.10 Statistical Analysis of Improved Variants of SSM and epPCR libraries using JMP

All results were analyzed using JMP. JMP contains built-in functions that analyze data using ANOVA statistical method as well as t-test fitted by standard least square personality with the emphasis on effect leverage. Both methods provide information on whether two enzymes are different from each other in terms of positions being mutated and further inference on whether one enzyme is superior than the other one could be made. Both nitroaldol to HNL activity ratio and nitroaldol activity for SSM libraries at each position were analyzed to determine if one is different than the other. Improved variants of SSM libraries and epPCR libraries were also analyzed with position nested within enzyme. Data was transformed into log scale to yield a normal distribution so that model can be applied for analysis.

Chapter 5. Results and Discussion

5.1 Fixing solubility of HbHNL by H103V

Before starting my SSM experiment, I tested both HNL1 and HbHNL enzymes expressed in 96 deep well plates with the substrates that lead to fluorescent product. HbHNL WT did not show any activity, even though HNL1 did. I ran their soluble fractions (supernatant) and insoluble fractions (inclusion bodies) on a mini protein gel (Fig. 5.1). Both HbhNL and HNL1 have similar molecular weight (about 30 kDa) and are identified by a band in between the 4^{th} and 5^{th} marker. HNL1 supernatant had visible bands at 30 kDa in its supernatant and that band was missing from the HbHNL supernatant but was present in the inclusion body. This showed that HbHNL was expressed, but did not fold correctly under the growth condition and most of it ended up in inclusion bodies. This explained why no activity could be detected using the WT HbHNL. Our group had noticed this problem prior to this experiment, but it was not a big concern since we could always grow twice or thrice as much bacterial culture to compensate for the lack of yield. However, I was limited to only 1-2 mL culture for each clone since I had to carry out the experiment in a higher throughput way in order to screen through a few thousand clones in a relatively short time. This called for a need to improve its solubility before any SSM libraries for HbHNL could be constructed.

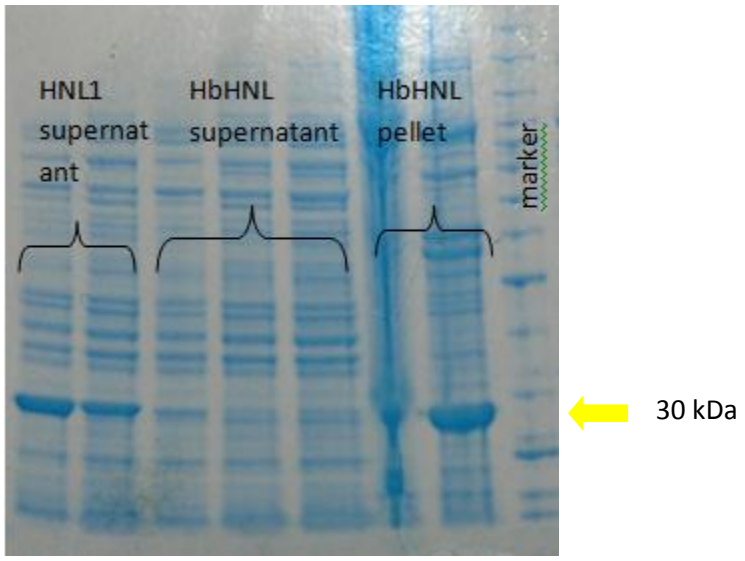


Figure 5.1. Protein gel shows low solubility of HbHNL WT. Supernatant from lysed cells in 96 deep well plate was run on a minigel showing over-expressed enzymes (HNL1 or HbHNL) at 30 kDa. Bands were present in HNL1 supernatant but were absent in HbHNL supernatant, which explains lack of activity in the screening.

Asano group reported that a few of H103 substitutions that increased solubility of MeHNL, which shares about 77% identity with HbHNL. Out of ~80 clones screened for improved activities for both reactions, four improved variants (H103L, H103V, H103M and H103I) were found and were consistent with Asano's finding.[45] Purified enzymes of these variants showed that the substitutions improved solubility significantly compared to WT HbHNL (Figure 3.6). They yielded at least 10 times more enzyme compared to WT.

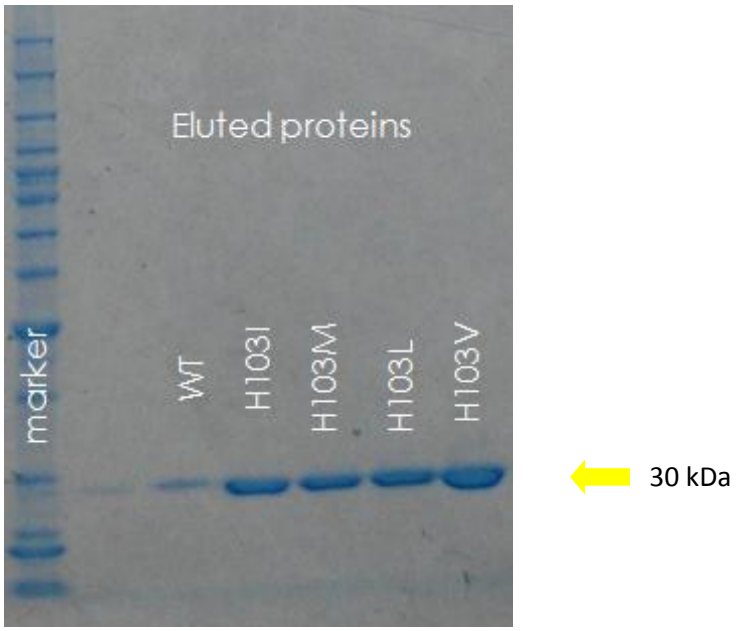


Figure 5.2. H103X substitution increased yield. Protein gel loaded with purified WT HbHNL, H103I. H103M, H103L and H103V, showing huge difference in amount of protein after purification.

All variants were tested against the 2-hydroxy-2-(6-methoxynaphthalen-2-yl)acetonitrile and 1-(6-methoxynapthlen-2-yl)-2-nitroethanol. The enzymes did not get saturated with the assay condition so the slope (V/K) was used to estimate activity. All variants had better nitroaldol activity but slightly worse HNL activity. This showed that they might be better starting point compared to the WT. Since all the variants were pretty similar in activities, I chose H103V as the template for site saturation mutagenesis because it gave the highest yield.

Table 5.1. Comparison of kinetic parameters of purified WT HbHNL and H103X variants. V/K, which is the slope for the kinetic curve, was used for comparison since the enzymes did not get saturated with substrate in the given assay condition.

	Henry	HNL			
Variants	V/K (s^{-1}·mM^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1}·mM^{-1})	V/K (s^{-1}·mM^{-1})
WT HbHNL	0.013	0.065	0.3109	0.21	
H103I	0.067				0.1
H103M	0.05				0.1

H103L	0.067	0.15
H103V	0.05	0.15

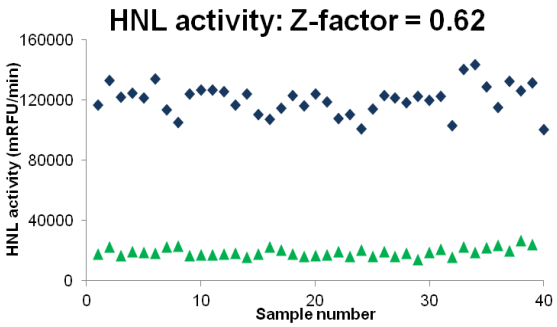
5.2 Quality assessment of screening using Z-factor

To determine whether it is possible to detect the fluorescent product at an acceptable level, wild type HNL1 and WT HbHNL (used as negative control as it was not expressed solubly in solution. Since it was not present in the supernatant, soluble fraction contains only crude proteins of E. coli cells) were tested using condition stated above. The separation of signal and noise levels was quantified using Z-factor.[47] The equation of Z-factor is shown below.

$$Z = 1 - \frac{3SD\ of\ sample + 3SD\ of\ control}{|mean\ of\ sample - mean\ of\ control|}$$

Z-factor was calculated to be 0.6 for HNL assay and 0.7 for nitroaldol assay (Figure 5.3). Z-factor in the range of 0.5- 1.0 is considered "excellent assay".[47] After 'fixing' HbHNL solubility problem with single mutation H103V, the same test was carried on HbHNL-H103V. Z-factor for both assays for HbHNL with H103V was similar, which is around 0.6-0.7. This showed that it is possible to differentiate activity from the noise of the screen so high throughput screening was carried out on the HbHNL-H103V SSM libraries.

(a)



(b)

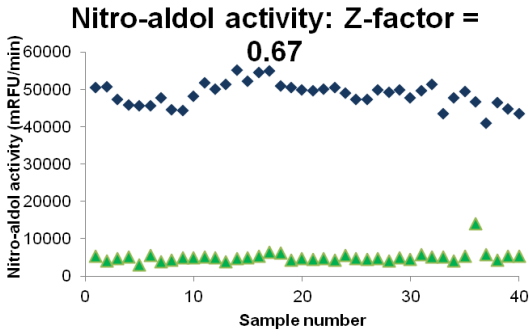


Figure 5.3. Diagrams showing good separation in activities between background (crude protein of E.coli) and positive control (WT HNL1). for (a) HNL activity and (b) nitroaldol activity.

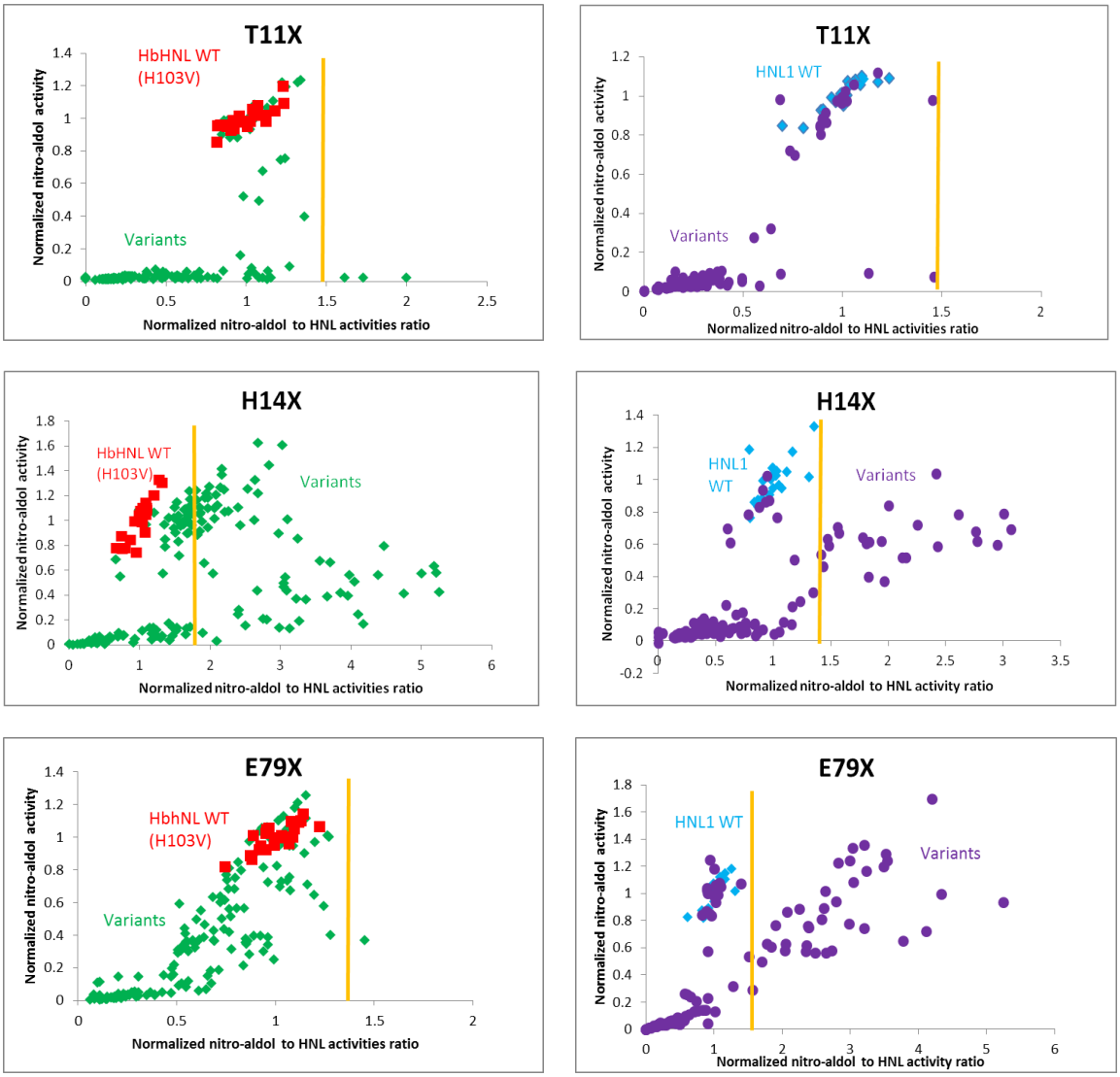
5.3 Hits obtained from SSM libraries

Clones with higher nitroaldol to HNL activity ratio were considered hits. SSM on positions 11, 14, 79, 106, 153, 178 yielded 0%, 43%, 0%, 12%, 14%, 41% hit rate respectively for HbHNL and 0%, 14%, 30%, 10%, 13%, 77% hit rate respectively for HNL1 (Figure 5.4). Sequencing result revealed that many of them are repeated hits for both enzymes due to oversampling. Nitroaldol to HNL activity ratio was taken into consideration instead of just nitroaldol activity because potential uneven protein expression or solubility across 96 deep well plates needed to be accounted. Clones that have high activities for both reactions may be due to higher expression/solubility, not real improvement for activities. However, clones that have higher nitroaldol activity but lower HNL activity indicate that it has good expression/solubility and so the improvement in nitroaldol activity is real.

Almost all T11X variants in both enzymes had no activity. The ones with activity were sent for sequencing but all of them were identified as WT. In HbHNL-H103V, mutation on H14X and L178X seemed to have more improvement in ratio as well as nitroaldol activity compared to other libraries. In HNL1, majority of F178X variants improved the ratio. H14X yielded some

clones with higher ratio but most of them have lower nitroaldol activity compared to WT in HNL1. E79X yield no hit in HbHNL-H103V but yielded some hits in HNL1. Mutation at position 153 yielded hits with higher ratio but lower if not similar nitroaldol activity compared to WT for both enzymes.

(a) (b)



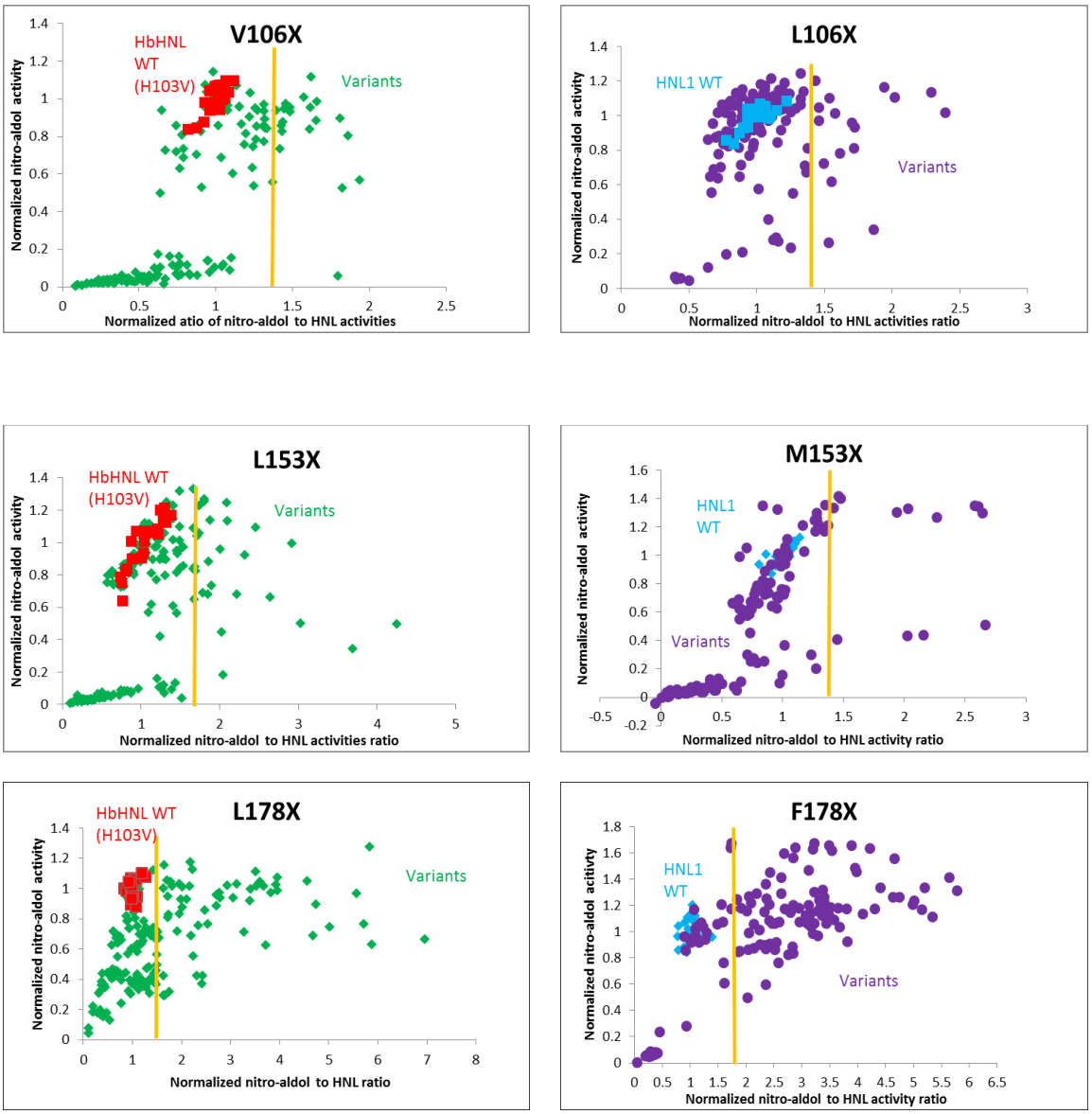


Figure 5.4. Site saturation mutagenesis libraries show hits on the right of the yellow line. (a) SSM on six different positions on HbHNL with H103V. (b) SSM on six different positions on HNL1.

To determine whether two enzymes are different when it comes to obtain a new enzymatic function, which is nitroaldolase activity in this case, each SSM library was analyzed using statistical software called JMP. Data were tested using ANOVA model. Results are

reported in terms of p-values in JMP to show whether two enzymes are significantly different (at α level= 0.05) from each other for SSM for each position (Table 5.2). Analysis was done on both the nitroaldol to HNL activity ratio and nitroaldol activity. Both enzymes are different for all positions except position 11 in terms of ratio. Enzymes are different for position 14, position 106 and position 178 but not different for position 14, 106 and 178 for nitroaldol activity. This analysis did not, however, provide any information about which enzyme is better.

Table 5.2: ANOVA analysis on both parameters: nitroaldol to HNL activity ratio and nitroaldol activity. Tick represents significant difference between the libraries at each position while cross shows no significant difference betweem the libraries.

Position	Nitroaldol to HNL activity ratio	Nitroaldol activity
11	✗	✗
14	✓	✓
79	✓	✗
106	✓	✓
153	✓	✗
178	✓	✓

5.4 Catalytic efficiency comparison between SSM hits and WT HNLs

All hits were purified and their k_{cat}/K_M values for nitroaldol activity were determined and only improved variants (> 2-fold in nitroaldol activity) were included in Table 5.3. HbHNL yielded 13 hits and HNL1 yielded 10 hits in terms of k_{cat}/K_M. For HbHNL, H14X yielded 2 variants, H14C and H14G that improved nitroaldol activity for 8 and 12-fold respectively. Even though there is a high hit rate (41%) in L178X library for HbHNL-H103V, the best variant L178V from the library improved the activity only by 10-fold. The same substitution, F178V in HNL1 led to almost 82-fold improvement in HNL1. However, it is important to note that HNL1 is a worse starting point as it has a worse catalytic efficiency compared to HbHNL.

While two libraries (H14X and L178X libraries) yielded more HbHNL variants with up to 10-fold improvement, only one library (F178X library) yielded improved variants for HNL1.

F178M and F178V improve nitroaldol activity by almost 80-fold whereas F178A improves it by 15-fold. In HbHNL, position 178 is occupied by leucine instead of phenylalanine. F178L improved nitroaldol activity in HNL1 to a lesser extent and has a k_{cat}/K_M value that is comparable to HbHNL-H103V. The two best variants of HbHNL have better k_{cat} value in general but overall improvement on k_{cat}/K_M is not as substantial as HNL1 variants because HbHNL-H103V is 3-fold better than HNL1 before mutagenesis. Variants from other libraries did not yield much improvement (<10-fold) in nitroaldol activity, which is not unexpected according to the results from SSM screening since most hits did not possess higher in nitroaldol activity based on the screening results. Unfortunately, no real hit in terms of nitroaldol activity was found for HNL1 in E79X library.

Table 5.3. The catalytic efficiency (k_{cat}/K_M) of each improved variants were compared to catalytic efficiency of WT HbHNL and HNL1. V/K is the slope of Michelis Menten curves when enzyme did not get saturated with substrate, which is common for HNL1 and its variants. Higher substrate concentration was not used in the assay due to substrate insolubility in water. K_i was also reported when curve was fitted using the substrate inhibition curve.

Variants	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($mM^{-1}\cdot s^{-1}$)	V/K	Fold better	Variants	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($mM^{-1}\cdot s^{-1}$)	V/K	Fold better
WT (H103V)				0.03		**WT**				0.012	
H14N	0.035	0.18	0.19		7	**H14A**	0.03	0.60	0.03		3
H14C	0.13	0.56 (K_i=0.56)	0.23		8	**H14C**				0.02	2
H14G	0.15	0.4 (K_i=0.62)	0.36		12	**H14G**				0.02	2
H14T	0.11	0.5	0.20		7						
V106R	0.03	0.4	0.06		2						
V106L	0.04	0.5	0.07		2						
L153G	0.06	0.4	0.13		4	**M153**				0.02	2

						A				
L153F	0.04	0.5	0.09		3	**M153F**			0.02	2
L153S				0.05	2					
L153T	0.07	0.6	0.10		3					
L178V	0.07	0.2	0.29		10	**F178K**	0.01	0.29	0.02	2
L178Q	0.19	1.1 (K_i=0.13)	0.17		6	**F178L**	0.02	0.38	0.05	4
L178G	0.03	0.28	0.11		4	**F178A**	0.06	0.31	0.18	15
						F178M	0.08	0.08	1.00	82
						F178V	0.06	0.06	1.01	82

Data from Table 5.3 was also analyzed using JMP to determine if one enzyme is better by mutating the same positions on the enzymes. Results showed that there was a significant difference in terms of position and position within enzyme, but not the enzymes themselves (Table 5.4a and 5.4b). Therefore, the position is important to explain the difference in improvement in nitroaldol activity of the enzymes. Results from JMP was reported in ANOVA table, effect tests as well as Tukey HSD test that show that position 178 in HNL1 is significantly difference than any other positions (Table 5.4c) and is put in category A. Positions 178, 153 and 14 of HbHNL and position 153 in HNL1 are not significantly different than each other (category B). Some overlap is observed between category B and C, making them less unique than each other. The result is not surprising since we already know that position 178 from both HNL1 and HbHNL produced variants with improvement up to 80 times and 10 times respectively. None of the other positions has produced better improved variants compared to this position.

Table 5.4. JMP statistical analysis results on SSM libraries between HbHNL and HNL1. (a) ANOVA table shows that results are significant different. (b) Effect tests shows that only position and position nested in enzyme (position * enzyme) are important to explain the difference (c) Tukey HSD shows that position 178 is the specific position that is significantly different than other positions.

(a)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	5.951504	1.19030	9.9907
Error	53	6.314488	0.11914	Prob > F
C. Total	58	12.265992		<.0001*

(b)

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Position	2	2	3.1408017	13.1810	<.0001*
Enzyme	1	1	0.0085141	0.0715	0.7903
Position*Enzyme	2	2	3.7722915	15.8312	<.0001*

(c)

Level			Least Sq Mean
178,HNL1	A		1.3520617
153,HbHNL		B	0.7711251
178,HbHNL		B	C 0.6276875
14,HbHNL		B	C 0.6095072
153,HNL1		B	C 0.3010300
14,HNL1			C 0.2798885

Levels not connected by same letter are significantly different.

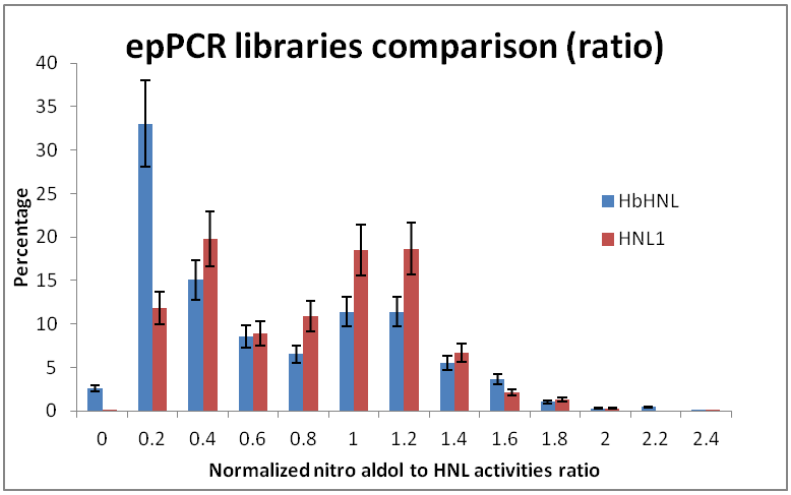
5.5 Hits from error prone PCR libraries

The mutation rates for both HbHNL and HNL1 epPCR libraries were set to be similar, which is about 2-3 mutations/gene. The mutation rate was controlled by varying the amount of DNA used in the epPCR. About 40 ng of plasmids in a 50 µL reaction is needed to obtain this mutation rate. About 8 plates were screened for each library. To be specific, 658 clones were screened for HbHNL and 617 clones were screened for HNL1. Screening was discontinued because no substantial improvement was observed.

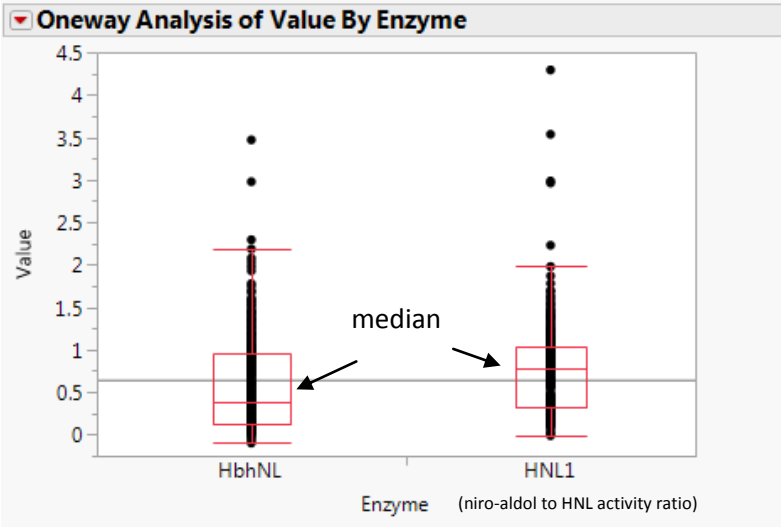
Libraries were compared by using bar charts to show percentages in each category after being normalized to WT in terms nitroaldol to HNL activity ratio as well as nitroaldol activity (Fig 5.5a and Fig. 5.5c). Judging from the bar charts, at least 50% of HbHNL variants fell in the 0-0.4 categories in terms of ratio or nitroaldol activity, whereas less than 40% of HNL fell in 0-0.4 categories. Only about 30% of HbHNL and almost 50% of HNL1 variants maintained ratio or nitroaldol activity level compared to WT (0.8-1.2 categories) assuming that error rate is about 20%. Error bars represent the average coefficient of variance of the WT HbHNL and HNL1 in

each library. HNL1 has an overall higher median compared to HbHNL in both nitroaldol to HNL activity ratio and nitroaldol activity (Fig. 5.5b and 5.5d)

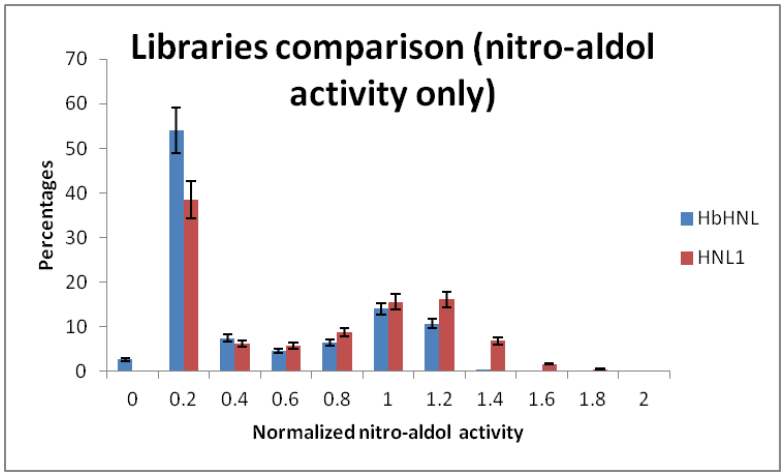
(a)



(b)



(c)



(d)

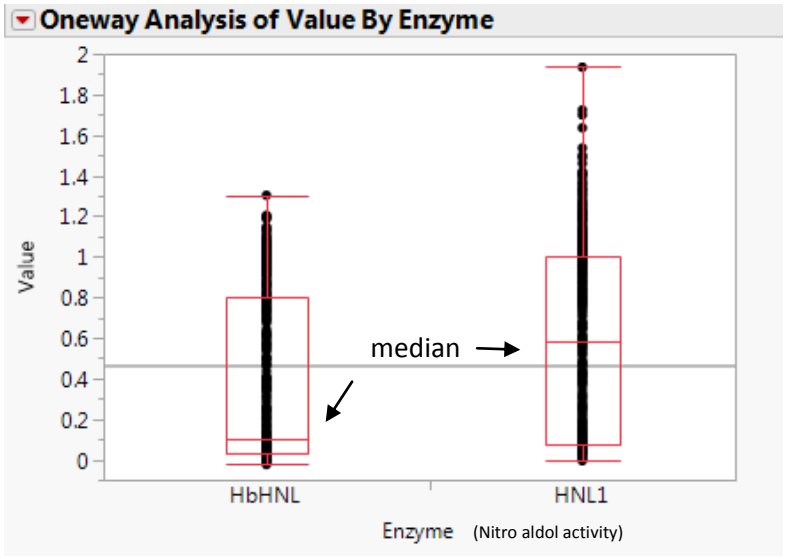
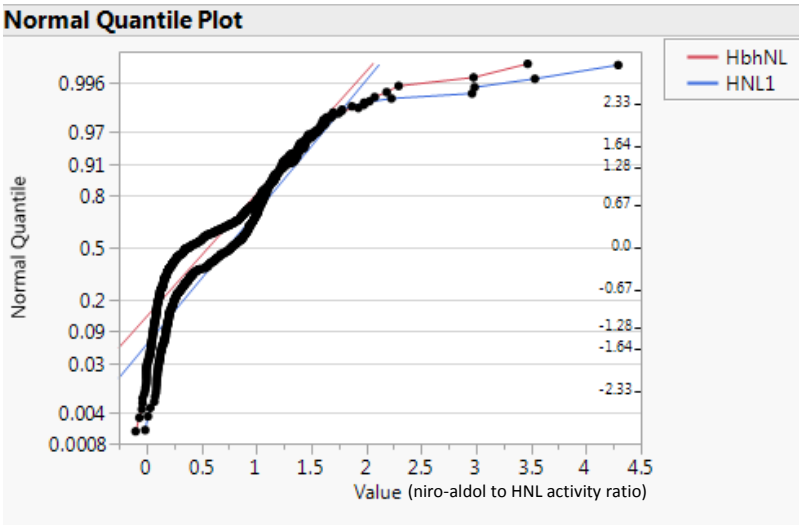


Figure 5.5. Comparison of distribution between HbHNL and HNL1 epPCR libraries. (a) Bar chart shows distribution of HbHNL and HNL1 libraries in terms of nitroaldol to HNL activity ratio. (b) Boxplot shows that HNL1 has a higher median compared to HbHNL in terms of nitroaldol to HNL activity ratio. (c) Bar chart shows distribution of HbHNL and HNL1 libraries in terms of nitroaldol activity only. (d) Boxplot shows that HNL1 has a higher median compared to HbHNL in terms of nitroaldol activity.

Data was analyzed by fitting values of normalized ratio or nitroaldol activity by enzymes (x by y). Distributions of the libraries were examined for its normality using normal quantile plots (Fig. 5.6). Both distributions were not normal as many data points are not on the straight line. It is

also pretty evident that the distributions are not normal just by looking at the bar charts themselves. Hence, ANOVA and t-test could not be carried out.

(a)



(b)

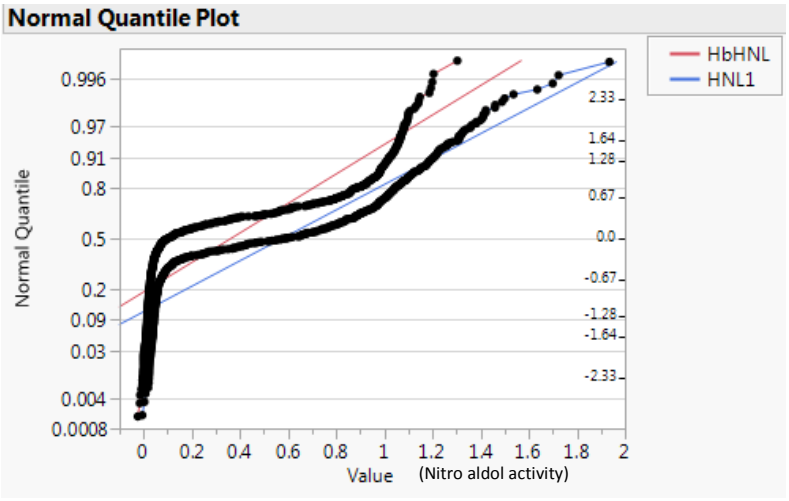


Figure 5.6. Normality of distributions of both enzymes tested using normal quantile plots. (a) Normal quantile plot shows that distributions of both HbHNL and HNL1 for nitroaldol to HNL activity ratio is not normal (b) Normal quantile plot shows that distributions of both HbHNL and HNL1 for nitroaldol activity is not normal

Instead, non-parametric Wilcoxon test (also known as Mann-Whitney test), which is commonly used for non-normal data that have similar distributions, was used to test if both enzymes are different from each other. Median of the data was being compared instead of the mean. Wilcoxon test showed that these enzymes are different at significance level of 0.05 based on two sample t-test as well as single sample chi square test. HNL1 has an overall higher score means, for both ratio and nitroaldol activity (Table 5.5).

Table 5.5. Wilcoxon test carried in JMP. (a) Wilcoxon test on nitroaldol to HNL activity ratio. (b) Wilcoxon test on nitroaldol activity.

(a)

Wilcoxon / Kruskal-Wallis Tests (Rank Sums) (Ratio)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
HbhNL	657	367377	418509	559.174	-7.800
HNL1	616	443524	392392	720.006	7.800

2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
443524	7.80048	<.0001*

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
60.8487	1	<.0001*

(b)

Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
HbHNL	657	349944	418509	532.639	-10.460
HNL1	616	460957	392392	748.307	10.460

2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
460957	10.46001	<.0001*

1-Way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
109.4135	1	<.0001*

More than 20 clones were selected as hits for HbHNL-H103V and HNL1 respectively and were characterized. A small number of them are false hits (nitroaldol activity could not be detected) and were not included on the table below (Table 5.6). Only six for HbHNL and five for HNL1 turned out to be improved nitroaldolase in terms of k_{cat}/K_M. Most of them have improvement in terms of K_M but not in terms of k_{cat}/K_M. Both HbHNL-H103V and HNL1 have a big K_M bigger than 1 mM that cannot be measured so the slope of the kinetic curve was used (V/K). Non-improved variants were not sent for sequencing, but left on the table for reference and to be used in JMP analysis.

Table 5.6. Catalytic efficiency (k_{cat}/K_M) values of all improved variants were compared to WT HbHNL and HNL1. (a) Data for all the hits characterized for nitroaldol activity and reported in terms of k_{cat}/K_M for HbHNL-H103V. (b) Data for all the hits characterized for nitroaldol activity and reported in terms of k_{cat}/K_M for HNL1. K_i values were also reported for variants that were fitted against substrate inhibition curve instead of Michealis Menten curve.

(a)

Variants	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($mM^{-1}.s^{-1}$)	V/K ($mM^{-1}\cdot s^{-1}$)	Improvement in fold
HbHNL-H103V (WT)				0.03	
L152I, A253S	**0.03**	**0.25**	**0.13**		**4.3**
E95V	**0.05**	**0.84**	**0.06**		**2.0**
A35T, E142G, L153M, T159S	**0.06**	**0.40**	**0.15**		**5.0**
I197T, D234E	**0.03**	**0.42**	**0.06**		**2.0**
A35V, K96N, A253V	**0.04**	**0.82**	**0.05**		**1.5**
T63M	**0.04**	**0.47**	**0.08**		**2.7**
5-5A	0.02	0.53	0.03		0.9
5-5E	0.02	0.93	0.02		0.5
5-6D	0.02	0.89	0.02		0.7
6-6F	0.01	0.61	0.02		0.7
7-4E	0.01	0.46	0.02		0.7
7-5C	0.02	0.80	0.02		0.7
5-8G	0.02	0.94	0.02		0.6

(b)

Variants	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($mM^{-1}.s^{-1}$)	V/K ($mM^{-1}\cdot s^{-1}$)	Improvement in fold
HNL1 WT				0.012	
G78S/F178L	**0.024**	**0.27**	**0.09**		**9**
F54L/D109N/C161S/K188E	**0.02**	**0.30 (K_i=0.47)**	**0.08**		**8**
H10R/K23R/L76V	**0.006**	**0.17**	**0.035**		**4**
L61H/F178L	**0.05**	**0.21**	**0.23**		**23**
P114L/N156I				0.01	1
Y57F/Q70H/F125S	0.002	0.33	0.006		0.5
V17E/R154Q/T159I/G232C	1.17E-03	0.60	0.002		0.2
5-6D	0.062	8.9 (K_i=0.055)	0.007		0.7
8-10F				0.01	1
3-4A	0.006	0.51	0.012		1.2

3-12D	0.8	81 (K_i=5.2)	0.009	0.9

The same analysis that was done on SSM hit variants was also done on the epPCR hit variants by including all the worse enzymes in terms of k_{cat}/K_M to determine if the enzymes are different. Since data distribution is not normal, data were transformed into log scale to yield a normal distribution before ANOVA test was done on them. Unfortunately, data analysis shows that these two enzymes are not significantly different than each other with a p-value of 0.34 (Table 5.7).

Table 5.7. ANOVA table shows no significant difference between hits from HbHNL and HNL1.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.2385584	0.238558	0.9616
Error	23	5.7061476	0.248093	Prob > F
C. Total	24	5.9447060		0.3370

5.6 Discussion on hit rate and actual hits

The number of improved nitroaldolase is relatively small given that many clones seemed to have improved the nitroaldol to HNL activity ratio in SSM libraries screening. My selection of hits was pretty loose (Table 4.2) because I did not want to miss any potential hits. So, it was inevitable to get some false positives. Most hits from SSM and epPCR libraries fell into "high ratio" and "low nitroaldol activity" category and the characterization result from purified hits reflected that. Some of the purified hits retained the original nitroaldol activity or have less nitroaldol activity but lost HNL activity completely (higher ratio compared to WT), which makes them non-improved nitroaldolase or worse enzymes overall. Since we were not interested in worse enzymes but the ones with improved nitroaldol activity, these enzymes were discarded. Very few clones actually fell into "high ratio" and "high nitroaldol activity" category. Based on

the screening results, none of the clones displayed more than 2X activity compared to WT during the screen. Only H14X from HbHNL library and F178X library from HNL1 have shown mild improvement in nitroaldol activity. However, some of these clones showed much better activity (up to 80-fold improvement in HNL1) after being purified and tested. This may mean that these clones were expressed solubly to a lesser extent when they were grown in 96 deep well plates. None of the clones actually fell in the "low ratio" and "high nitroaldol activity." Most of the clones fell in "low ratio" and "low nitroaldol activity" category, were never selected and tested. Hence, the number of improved variants from SSM libraries is relatively small in comparison to actual number of hits obtained from the libraries.

Using nitroaldol activity to HNL activity ratio to normalize protein expression/solubility seemed to have worked but it is not the best way to select for improved nitroaldolase given the amount of time and effort to characterize many non-improved variants only to find out that they were not truly improved in the desired activity. At the end of 2015, Bornscheuer group reported a split-GFP technology that can be used to normalize protein expression/solubility level. A 16-amino acid GFP11 tag is attached to either N or C terminus of desired protein or as part of an exposed flexible loop.[48] Total amount of protein is quantified by detecting the GFP11 tag that is co-expressed using a fluorescence plate reader. Activity of variants is then be normalized by dividing the fluorescence values by the GFP11 tag, resulting in more true hits. By using split-GFP technology, one does not have to rely on the ratio to select the hits since ratio (one activity divided by another activity) has an overall higher error rate. It may allow more stringent selection based on nitroaldol activity instead of the ratio. It may also lead to less wasted effort and time and also result in higher true hits.

Most real hits do not actually reflect hit rate calculated from screening. L178X for HbHNL only yielded 3 improved hits despite of 44% hit rate. Many clones were sent for

sequencing but most of them turned out to be the same variant, L178G. This is due to oversampling to cover at least 95% of the library. It is also due to our decision to forgo worse, specific enzymes, which might have high nitroaldol to HNL activity ratio. For example, E79X library that yielded 30% hit rate and but none of the hits were actually improved nitroaldolase. This could mean that they are overall worse enzymes with lower nitroaldol activity and even worse HNL activity that gave them higher ratio compared to the WT. These enzymes were discarded and were never considered hits.

It is also important to note that NNK degenerate codons are not perfect. It is commonly used because only one set of primers is needed and it has much lower redundancy compared to NNN degenerate codons. The redundancy is not completely omitted since it still contains three codons for serine, arginine and leucine, two codons for valine, alanine, proline, glycine and threonine that are responsible for a biased library. Screening for 146 clones covers most of the clones, including the rarest amino acid but it might lead to different numbers of hit rate. For example, if leucine (encoded by three different codons) happened to be a hit in a library, it will be found more times compared to asparagine (encoded by one codon) in another library. This leads to ~9% of hit rate for the former library but ~ 3% hit rate for the latter library. Therefore, hit rate is not the best parameter to use to determine which enzyme is more capable of producing better nitroaldolase because of codon bias of NNK. It is only good as a quick reference to observe the overall pattern of improvement in each library. Subsequent purification and characterization is the only way to determine which library is capable to producing better clones.

A novel way to further reduce redundancy of genetic code from 32 from NNK to 22 by using three different set of primers called "22c-trick" was reported. One degenerate primer, NDT (D=no C) encodes for 12 codons, VHG (V= no T, H= no G) for another nine and TGG for tryptophan.[38] With this method, no stop codon is encoded and only two amino acids (valine and

leucine) are encoded twice. To cover even the rarest amino acid, one only has to screen ~66 clones instead of 146. Since 18 amino acids are only encoded once, hit rate of the library is more likely to reflect the number of actual hits after purification and characterization. This method is extremely helpful when more than one site is being mutated because it greatly reduces the screening effort. NNK degenerate codons are not perfect but they are convenient to use when one site is being mutated and a good screening method is present. If I were to mutate more than one site for each enzyme, 22c-trick is definitely a better option for me.

5.7 Bigger active sites are responsible for nitroaldol activity improvement

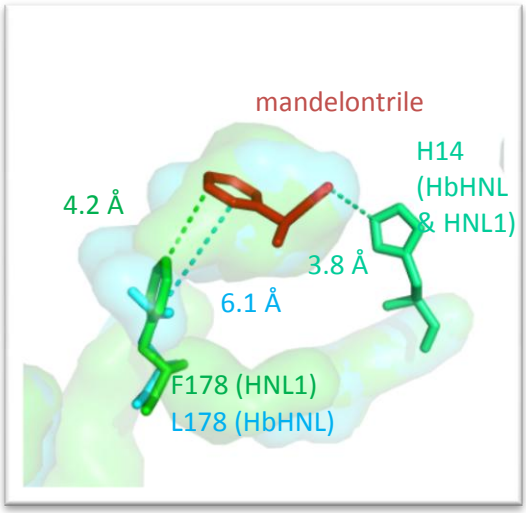
Most improved variants seemed to suggest that making more space in active site improved nitroaldol activity in both HbHNL and HNL1. I tried to mutate positions 14 and 178 in HbHNL and HNL1 to see how new residues fit in their active sites in Pymol by using a crystal structure (PDB code: 1YB6) of HbHNL previously complexed with mandelonitrile and HNL1 that was not complexed with mandelonitrile (Fig. 5.7a). Two different structures were structurally aligned in Pymol and positions 14 and 178 were selected to be substituted by the best amino acids for both enzymes. I assumed that 1-(6-methoxynapthlen-2-yl)-2-nitroethanol binds the same way despite of its bigger size. I manually docked the substrate in the active site (Fig. 5.7b) by orienting the hydroxyl group to the hydroxyl group of mandelonitrile. The NO_2 group also points to the same direction where the CN group points. Residues 14 and 178 are much closer to the substrate because of its bigger size.

All the good mutations of position 14 and position 178 have similar effect, which is to make the binding pocket slightly larger. This potentially leads to better binding since the substrate is bigger. Position 14 is close to where NO_2 group binds (Fig 5.7b), so it is not surprising that replacing it with glycine (Fig. 5.7d) may have created more space (from 2.1 Å to 4.9 Å) for the NO_2 group to bind. In HbHNL, position 178 is occupied by leucine (Fig. 5.7a and 5.7b) but same

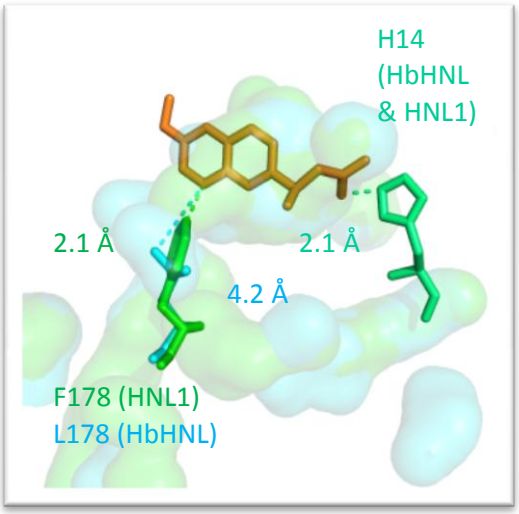
position is occupied by phenylalanine in HNL1 (Fig 5.7a and 5.7b). As shown in diagrams below, phenylalanine is a big amino acid and might have reduced the space in the active site in HNL1. This may be the reason why F178L increased the nitroaldol activity by 4-fold in HNL1 (not shown in picture). Other substitution such as valine (Fig. 5.7c) greatly increased nitroaldol activity in HNL1 also created more space in active site of HNL1 (from 2.1 Å to 4.8 Å) but this substitution did not have a huge impact in HbHNL because they are more similar in sizes (from 4.2 Å to 4.7 Å). This explains why the same substitution did not improve HbHNL dramatically.

SSM results also suggest that position 11 is the worst place to mutate because it is essential for enzyme functionality. T11 provides hydrogen bond to the carbonyl part of the substrate. Removing it may cause the substrate to not able to bind effectively.

(a)



(b)



(c) (d)

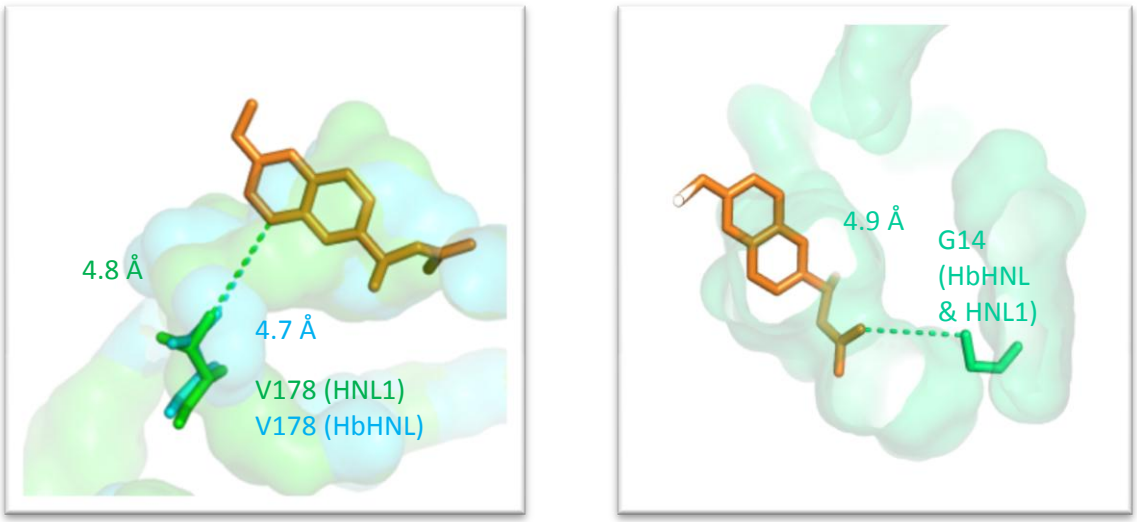


Figure 5.7. Diagrams show residues in both HbHNL and HNL1 that are structurally aligned. Green color represents residue from HNL1 while blue color represents residue from HbHNL. Cyan color represents residues from both enzymes. Dotted lines also show the distance of two closest carbons of the substrates and the residues.(a) Mandelonitrile was complexed with HbHNL crystal structure (PDB code: 1YB6). Diagram shows F178 leads to smaller space in the active site. (b) Diagram shows how 1-(6-methoxynapthlen-2-yl)-2-nitroethanol fits by fixing the position of hydroxyl group using mandelonitrile as a reference. (c) Mutation on position 178 makes more space in the active for both enzymes. (d) Mutation on position 14 makes more space in the active site for both enzymes.

5.8 Hits from epPCR library

Although HNL1 variants seemed to have higher nitroaldol to HNL ratio (suggesting that they might be more specific) and nitroaldol activity based on the analysis result, the characterization of pure enzyme hits did not seem to reflect that. Analysis on the purified hits showed that they are not significantly different. Many of the hits were found to have lower nitroaldol activity in terms of k_{cat}/K_M compared to the WT. However, most of them have lower K_M compared to WT, which may explain why they appeared as hits in screening when tested with low concentration of substrate (~100 µM).

Again, my loose selection might also be responsible for high false rate. Most hits I selected are from "high ratio" and "low nitroaldol activity" category so it actually was not surprising to obtain hits that are poorer but more specific nitroaldolase. WT carryover is another common problem in error prone PCR library especially when mutation rate is kept low or there is an imperfect dpN1 digestion. However, WT carryover might not be huge problem here since the percentages of clones around 0.8-1.2 categories for both enzymes in nitroaldol activity are similar, showing that there are similar percentages of WT-like clones/WT itself even if carryover is an issue. These enzymes have bigger difference in terms of ratio in the 0.8-1.2 categories, showing that variants other than WT were more likely to contribute in the differences in distributions.

5.9 Comparison of best hits from SSM and epPCR libraries

SSM libraries have produced more improved nitroaldolases compared to epPCR libraries. Since almost 50% of epPCR library is inactive, only about 300 active clones were actually tested. Random mutagenesis requires a lot of time and effort as well as resources. It is impossible to test every single possible variant with 2-3 mutations in a 300 amino acid long protein. I only managed to sample a very small fraction of the endless possibilities to see if one enzyme performs better than the other. With the help of computational knowledge, one can look at the active site of the enzyme and speculate what residues to mutate to obtain new function. Combining both directed evolution and SSM worked pretty well for my experiment since I was able to find better improved variants compared to random mutagenesis alone.

However, it is interesting to note that epPCR library was able to find the same substitutions on certain regions despite the fact that only a small number of clones from both libraries were screened. For HbHNL, substitutions of position 35 and position 253 were found twice. For HNL1, F178L was also found twice. F178L is not the best variant in the F178X family but it was picked up by the screening repeatedly. Random mutagenesis has produced interesting

results that we would never have speculated to be beneficial for the enzyme, such as A35 or A235 for HbHNL since both are far from active site and their interaction with the substrates is much harder to predict. If we had no prior knowledge about the structure and mechanism, A35 and A235 in HbHNL as well as F178 in HNL1 would be good starting points for mutagenesis.

5.8 Conclusions

The number of improved variants is similar in both enzymes for both SSM and epPCR libraries. Statistical analysis on the improved hits also has concluded that there is no difference between two enzymes. It may seem that ancestral enzyme is no better in evolving new function at first glance but it is also important to note that it was impossible to start screening HbHNL without improving its solubility. Asano et al., 2011 has speculated that H103 is a bad residue in a hydrophobic pocket, leading to instability of the MeHNL (Fig. 5.6).[42] Mutating it to other hydrophobic residues helps the enzyme to fold correctly in order to be functional. The H103 substitution has also benefited another *Baliospermum montanum* (BmHNL) since H103 is a highly conserved among the modern HNLs.[49]

Our group has also shown that HbHNL is a much less stable enzyme compared to HNL1 in terms of thermal stability. HNL1 unfolding temperature is about 80 °C while HbHNL unfolding temperature is only 54 °C.[29] Screening with HbHNL was only possible after its solubility was improved by H103V substitution.

A protein is only functional when it is able to fold to a 'thermodynamically stable' structure and remains folded at its physiological condition.[50] In our case, HbHNL WT is not able to stay folded once it is expressed and most of it remains in inclusion bodies. Because of this, 'evolution' cannot take place since it is impossible to select for nitroaldol activity. Robustness

conferred by stability is prerequisite to gaining new functions in many proteins because stable proteins are able to accept destabilizing mutations that increase activity.[50,51]

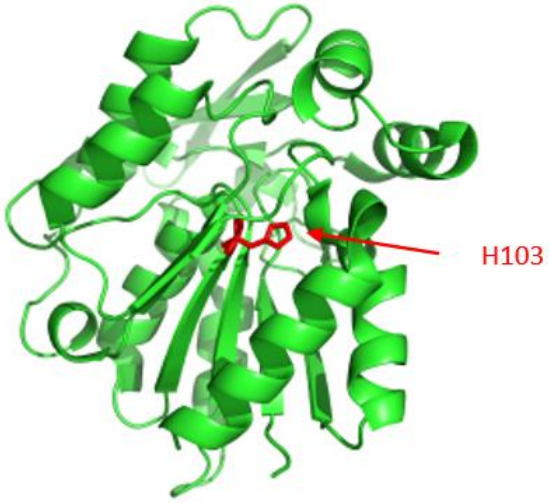


Figure 5.6. HbHNL crystal structure (PDB code: 1YB6). H103, residue colored in red, is a hydrophilic residue buried in a hydrophobic core.

HNL1 might potentially be a better starting point just based on the fact that it is a much more stable or "robust" enzyme compared to WT HbHNL. Error prone PCR results shown in bar chart also suggested that HNL1 is more likely to retain activity compared to HbHNL. Therefore, HNL1 is more 'evolvable' compared to WT HbHNL. However, once the solubility of HbHNL was improved, HNL1 is not better at improving new catalytic function compared to HbHNL-H103V. My results also showed that focusing on the active site is a good strategy to improve an unnatural reaction. Improved variants might be good starting points for the next round of mutagenesis (either by combining good mutations in the active site or random mutagenesis) to further improve the enzyme.

Bibliography

1. 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.

2. Aehle, Wolfgang, and Onno Misset. "Enzymes for industrial applications."Biotechnology Set, Second Edition (1999): 189-216.

3. Kokkinidis, M., Glykos, N. M., & Fadouloglou, V. E. (2012). Protein flexibility and enzymatic catalysis. Adv Protein Chem Struct Biol, 87, 181-218.

4. Zhang, X., & Houk, K. N. (2005). Why enzymes are proficient catalysts: beyond the Pauling paradigm. Accounts of chemical research, 38(5), 379-385.

5. Nelson, D. L. & Cox, M. M. (2004). *Lehninger Principles of Biochemistry*. New York: W.H. Freeman and Company

6. Sanchez, S., & Demain, A. L. (2010). Enzymes and bioconversions of industrial, pharmaceutical, and biotechnological significance. *Organic Process Research & Development*, *15*(1), 224-230.

7. Holmquist, M. (2000). Alpha beta-hydrolase fold enzymes structures, functions and mechanisms. Current Protein and Peptide Science, 1(2), 209-235.

8. Nardini, M., & Dijkstra, B. W. (1999). α/β Hydrolase fold enzymes: the family keeps growing. Current opinion in structural biology, 9(6), 732-737.

9. Forouhar, F., Yang, Y., Kumar, D., Chen, Y., Fridman, E., Park, S.W., Chiang, Y., Acton, T.B., Montelione, G.T., Pichersky, E. and Klessig, D.F. (2005). Structural and biochemical studies identify tobacco SABP2 as a methyl salicylate esterase and implicate it in plant innate immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5), 1773-1778.

10. Park, Sang-Wook, et al. "Use of a synthetic salicylic acid analog to investigate the roles of methyl salicylate and its esterases in plant disease resistance."*Journal of Biological Chemistry* 284.11 (2009): 7307-7317

11. Hickel, A., Hasslacher, M., & Griengl, H. (1996). Hydroxynitrile lyases: functions and properties. Physiologia Plantarum, 98(4), 891-898.

12. Jensen, N. B., Zagrobelny, M., Hjernø, K., Olsen, C. E., Houghton-Larsen, J., Borch, J., ... & Bak, S. (2011). Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects. Nature communications, 2, 273.

13. Kadow, D., Voß, K., Selmar, D., & Lieberei, R. (2012). The cyanogenic syndrome in rubber tree Hevea brasiliensis: tissue-damage-dependent activation of linamarase and hydroxynitrile lyase accelerates hydrogen cyanide release. Annals of botany, mcs057

14. Ganjewala, D., Kumar, S., Devi, S. A., & Ambika, K. (2010). Advances in cyanogenic glycosides biosynthesis and analyses in plants: A review. *Acta Biologica Szegediensis*, *54*(1), 1-14.

15. Andexer, J. N., Langermann, J. V., Kragl, U., & Pohl, M. (2009). How to overcome limitations in biotechnological processes-examples from hydroxynitrile lyase applications. Trends in biotechnology, 27(10), 599-607.

16. Sharma, M., Sharma, N. N., & Bhalla, T. C. (2005). Hydroxynitrile lyases: at the interface of biology and chemistry. Enzyme and microbial technology, 37(3), 279-294.

17. Purkarthofer, T., Gruber, K., Gruber-Khadjawi, M., Waich, K., Skranc, W., Mink, D., & Griengl, H. (2006). A biocatalytic Henry reaction—the hydroxynitrile lyase from Hevea brasiliensis also catalyzes nitroaldol reactions. *Angewandte Chemie International Edition*, *45*(21), 3454-3456.

18. Gruber-Khadjawi, M., Purkarthofer, T., Skranc, W., & Griengl, H. (2007). Hydroxynitrile Lyase-Catalyzed Enzymatic Nitroaldol (Henry) Reaction. Advanced Synthesis & Catalysis, 349(8-9), 1445-1450.

19. Andexer, J. N., Langermann, J. V., Kragl, U., & Pohl, M. (2009). How to overcome limitations in biotechnological processes-examples from hydroxynitrile lyase applications. Trends in biotechnology, 27(10), 599-607.

20. Gruber, K., Gartler, G., Krammer, B., Schwab, H., & Kratky, C. (2004). Reaction Mechanism of Hydroxynitrile Lyases of the α/β-Hydrolase Superfamily THE THREE-DIMENSIONAL STRUCTURE OF THE TRANSIENT ENZYME-SUBSTRATE COMPLEX CERTIFIES THE CRUCIAL ROLE OF LYS236. *Journal of Biological Chemistry*, *279*(19), 20501-20510.

21. Wagner, U. G., Hasslacher, M., Griengl, H., Schwab, H., & Kratky, C. (1996). Mechanism of cyanogenesis: the crystal structure of hydroxynitrile lyase from Hevea brasiliensis. *Structure*, *4*(7), 811-822.

22. Padhi, Santosh Kumar, et al. "Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions." *Chemistry & biology* 17.8 (2010): 863-871.

23. Nedrud, David M., et al. "Uncovering divergent evolution of α/β-hydrolases: a surprising residue substitution needed to convert Hevea brasiliensis hydroxynitrile lyase into an esterase." *Chemical Science* 5.11 (2014): 4265-4277.

24. Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annual Reviews in Microbiology*, *30*(1), 409-425.

25. Bergthorsson, U., Andersson, D. I., & Roth, J. R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences*, *104*(43), 17004-17009.

26. Näsvall, J., Sun, L., Roth, J. R., & Andersson, D. I. (2012). Real-time evolution of new genes by innovation, amplification, and divergence. *Science*,*338*(6105), 384-387.

27. Risso, Valeria A., et al. "Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases." *Journal of the American Chemical Society* 135.8 (2013): 2899-2902.

28. Voordeckers, K., Brown, C. A., Vanneste, K., van der Zande, E., Voet, A., Maere, S., & Verstrepen, K. J. (2012). Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. PLoS Biol, 10(12), e1001446.

29. Devamani, T., Rauwerdink, A. M., Lunzer, M., Jones, B. J., Mooney, J. L., Tan, M. A. O., Zhang, Z.J., Xu, J. H., Dean, A. M. & Kazlauskas, R. J. (2016). Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *Journal of the American Chemical Society*.

30. Sleator, R. D. (2011). Phylogenetics. Archives of microbiology, 193(4), 235-239.

31. Bosley, Allen D., and Marc Ostermeier. "Mathematical expressions useful in the construction, description and evaluation of protein libraries." *Biomolecular engineering* 22.1 (2005): 57-61.

32. Drummond, D. Allan, et al. "Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins." *Journal of molecular biology*,350.4 (2005): 806-816.

33. Steiner, K., & Schwab, H. (2012). Recent advances in rational approaches for enzyme engineering. *Computational and structural biotechnology journal*, 2(3), 1-12.

34. Lehmann, M., Pasamontes, L., Lassen, S. F., & Wyss, M. (2000). The consensus concept for thermostability engineering of proteins. Biochimica et Biophysica Acta (BBA)-Protein *Structure and Molecular Enzymology*, 1543(2), 408-415.

35. Seebeck, F. P., & Hilvert, D. (2003). Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation. *Journal of the American Chemical Society*, 125(34), 10158-10159.

36. Brissos, V., Eggert, T., Cabral, J. M. S., & Jaeger, K. E. (2008). Improving activity and stability of cutinase towards the anionic detergent AOT by complete saturation mutagenesis. *Protein Engineering Design and Selection*, 21(6), 387-393.

37. Reetz, Manfred T., José D. Carballeira, and Andreas Vogel. "Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability." *Angewandte Chemie International Edition* 45.46 (2006): 7745-7751.\

38. Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z. G., Opperman, D. J., Reetz, M. T., & Acevedo, J. P. (2012). Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS synthetic biology*, *2*(2), 83-92.

39. Reetz, Manfred T., et al. "A genetic selection system for evolving enantioselectivity of enzymes." *Chemical Communications* 43 (2008): 5502-5504.

40. Dean, D. (1981). A plasmid cloning vector for the direct selection of strains carrying recombinant plasmids. *Gene*, *15*(1), 99-102.

41. Krammer, B., Rumbold, K., Tschemmernegg, M., Pöchlauer, P., & Schwab, H. (2007). A novel screening assay for hydroxynitrile lyases suitable for high-throughput screening. *Journal of biotechnology*, *129*(1), 151-161.

42. Asano, Yasuhisa, et al. "Functional expression of a plant hydroxynitrile lyase in Escherichia coli by directed evolution: creation and characterization of highly in vivo soluble mutants." *Protein Engineering Design and Selection* 24.8 (2011): 607-616.

43. Andexer, J., Guterl, J. K., Pohl, M., & Eggert, T. (2006). A high-throughput screening assay for hydroxynitrile lyase activity. *Chemical communications*, (40), 4201-4203.

44. Jones, P. D., Wolf, N. M., Morisseau, C., Whetstone, P., Hock, B., & Hammock, B. D. (2005). Fluorescent substrates for soluble epoxide hydrolase and application to inhibition studies. *Analytical biochemistry*, 343(1), 66-75.

45. Asano, Y., Dadashipour, M., Yamazaki, M., Doi, N., & Komeda, H. (2011). Functional expression of a plant hydroxynitrile lyase in Escherichia coli by directed evolution: creation and characterization of highly in vivo soluble mutants. Protein Engineering Design and Selection, 24(8), 607-616.

46. Bhunya, R., Mahapatra, T., & Nanda, S. (2009). Prunusarmeniaca hydroxynitrile lyase (ParsHNL)-catalyzed asymmetric synthesis of cyanohydrins from sterically demanding aromatic aldehydes. *Tetrahedron: Asymmetry*, *20*(13), 1526-1530.

47. Zhang, J. H., Chung, T. D., & Oldenburg, K. R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of biomolecular screening*, *4*(2), 67-73.

48. Santos-Aberturas, J., Dörr, M., Waldo, G. S., & Bornscheuer, U. T. (2015). In-Depth High-Throughput Screening of Protein Engineering Libraries by Split-GFP Direct Crude Cell Extract Data Normalization. *Chemistry & biology*,*22*(10), 1406-1414.

49. Kawahara, N., & Asano, Y. (2015). Mutagenesis of an Asn156 Residue in a Surface Region of S-Selective Hydroxynitrile Lyase from Baliospermum montanum Enhances Catalytic Efficiency and Enantioselectivity. *ChemBioChem*, 16(13), 1891-1895.

50. Bloom, J. D., Labthavikul, S. T., Otey, C. R., & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences*, *103*(15), 5869-5874.

51. Tokuriki, N., & Tawfik, D. S. (2009). Stability effects of mutations and protein evolvability. *Current opinion in structural biology*, *19*(5), 596-604.

52. Khersonsky, O., Roodveldt, C., & Tawfik, D. S. (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. *Current opinion in chemical biology*,*10*(5), 498-508.

53. Leemhuis, H., Kelly, R. M., & Dijkhuizen, L. (2009). Directed evolution of enzymes: library screening strategies. *IUBMB life*, 61(3), 222-228.

54. Rauwerdink, Alissa, Mark Lunzer, Titu Devamani, Bryan Jones, Joanna Mooney, Zhi-Jun Zhang, Jian-He Xu, Romas J. Kazlauskas, and Antony M. Dean (2015). "Evolution of a Catalytic Mechanism." *Molecular biology and evolution*, msv338.

55. Zheng, L., Baumann, U., & Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research*, 32(14), e115-e115.

56. Babtie, A., Tokuriki, N., & Hollfelder, F. (2010). What makes an enzyme promiscuous? Current opinion in chemical biology, 14(2), 200-207.