Developing new enzymatic catalysts by resurrecting ancestral alpha/beta hydrolases

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Joanna Mooney

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Advisor: Romas J. Kazlauskas

September 2014

Acknowledgements

I would like to thank everyone in the Kazlauskas lab for sharing their expertise and great research ideas, especially Bryan Jones, Titu Devamani, and Alissa Rauwerdink. I would also like to thank all of the other members of the Kazlauskas lab, including Jun Huang, Hui Lin, David Nedrud, Huey Yee Lim, and Alexis Powell. Romas is an outstanding PI and teacher and it was a great experience being able to work in his lab. I would also like to thank Tony Dean and Larry Wackett for serving on my committee.

I would also like to thank Nick Kotloski, Ben Bonis, Ben Binder, W.C. Solomon, Lori Zacharoff, Jonathan Crain, Eu Han Lee, Dan Stroik, Maya Raghunandan, Brittany Bennett, Eric Kees, Aubree Wilke, Audrey Harris, Misha Golynskiy, Alex Sobeck, Barb Pinch, Alex Lange, and everyone in Gortner Laboratory.

Finally, I would like to thank my friends Mallory Kurkoski, Matt Duscher, Ann Johnson, Sam Farrand, Evan Weitz and Melissa Panzer who have made Minneapolis a wonderful place to live. Last, but certainly not least, I would like to thank Bella, Scoops, and Puce Jr.

Dedication

This thesis is dedicated to everyone who is interested in green chemistry.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction to α/β hydrolases

1.1 Introduction

While synthetic chemistry has provided the world with solutions to many problems, it has also increased risks to the environment and to human health. Chemical synthesis has allowed for the production of necessary pharmaceuticals and commodity chemicals. Unfortunately, byproducts of large scale chemical manufacturing processes have often had a severely adverse environmental impact, including polluted water sources and uninhabitable land.[1]

A solution to these would be to replace chemical synthesis with enzymatic synthesis. Chemical processes often require harsh conditions such as high temperature, pH, or organic solvents, all of which pose health risks. In addition to greater safety, enzymes are more selective so they produce fewer unwanted side products.

Many industrially useful reactions are not known to be catalyzed by any yet characterized enzymes. Some of these reactions may occur at low levels or be a promiscuous, or alternative in addition to the natural, activity of existing enzymes.

1.2 What do enzymes do?

All life is based on the chemical conversion of compounds into other forms. This includes the processing of carbon dioxide into sugars using energy from sunlight or the conversion of food into energy, fat and muscle. Many of these reactions could take years to occur spontaneously without the help of enzymes.

Enzymes, on the other hand, catalyze hundreds or thousands of reactions per second.[2]

Enzymes work by increasing the rate of a reaction. The rate of a reaction is a measure of how rapidly a product is produced by the enzyme. For an enzyme to catalyze the conversion of a substrate into product, the substrate must first interact with the active site of the enzyme. The binding of an enzyme (E) to a substrate (S) is reversible – if a substrate binds irreversibly to the enzyme the reaction cannot proceed. Once the enzyme-substrate complex (ES) forms, the reaction the conversion of the substrate into product is completed and the product is released from the active site of the enzyme. This is what occurs in the case of irreversible inhibitors that block the active site. The affinity constant, K_M, further described below, represents how well the substrate binds to the active site of an enzyme. When the catalytic step is complete, the enzyme is regenerated when the product (P) is released. This part of the reaction is not reversible because the equilibrium of the reaction drives the production of the product. The reaction will only proceed if the reaction is favorable. The following is a schematic summarizing the steps needed for the enzymatic catalysis of a substrate into products:

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

While an efficient enzyme will speed of the rate of a reaction, all reactions also have a chemical equilibrium position. A reaction is said to have reached equilibrium when the concentrations of reactants and products no longer change. This occurs when the forward and reverse reactions proceed at the same rate. Product will continue to be produced until equilibrium is reached. Regardless of how fast an enzyme catalyzes a reaction, it cannot change the equilibrium of a reaction.

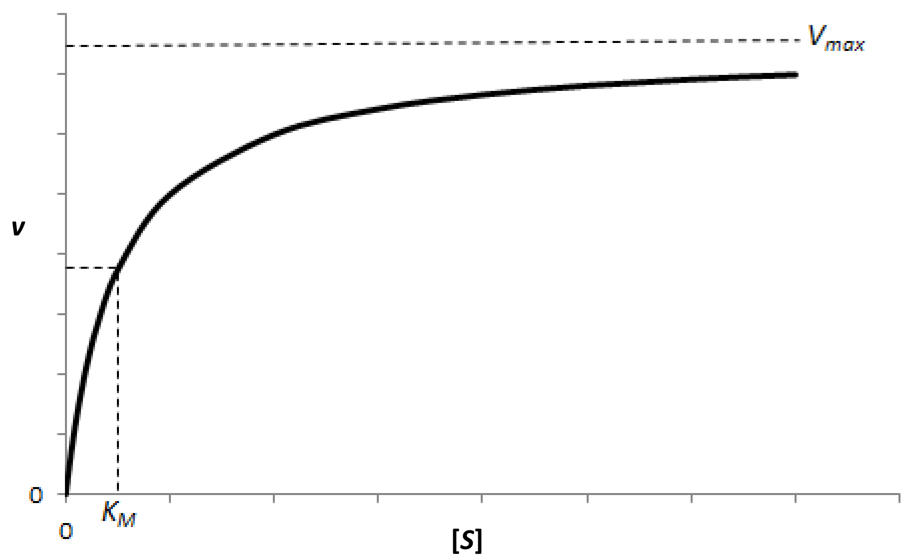


Figure 1.1 How reaction rate (*v*) varies with substrate concentration ([*S*]). The figure above illustrates the concepts of V_{max} and K_M. The V_{max} is the asymptotic maximum velocity (v) that can be reached by the reaction. K_M is the x value ([S]) that is halfway between V_{max} and 0 on the y axis (v).

Enzyme kinetics is expressed as number of reactions catalyzed per enzyme molecule per second, called k_{cat}.[3] Another commonly used measure of enzyme activity is V_{max}, the maximum velocity of enzyme catalyzed reaction, or the change in concentration of the catalyzed products at a given substrate concentration. The V_{max} is how fast an enzyme can catalyze a reaction when every active site in a solution is occupied. K_M is defined as the substrate concentration at which the enzyme catalyzes the reaction at half V_{max}, or the concentration of substrate that occupies half of the active sites of an enzyme. In general, the lower the K_M value, the higher the affinity the enzyme has for the substrate. In short, the stronger the interaction in the correct orientation between the enzyme active site and the substrate, the faster the substrate be catalyzed to product.

The following equations illustrate the relationship of V_{max}, K_M, and k_{cat}.

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

[S] is the concentration of substrate and [E] is the concentration of enzyme. V_{max} is the maximum velocity of a reaction at a given enzyme concentration $[E]_t$ at a given time (t).

$$v = \frac{k_{cat}[E]_t[S]}{K_M + [S]}$$

$[E]_t$ can be simplified to $[E]_0$ if we use the same substrate concentration for all measurements. V_{max} is given in units of moles of product/time. V_{max} is therefore equal to k_{cat}, which usually has units of sec^{-1} or min^{-1}, multiplied by the moles of enzyme used in the reaction.

$$V_{max} = K_M[E]_0$$

Enzymes increase the rate of reaction by reducing the energy of activation needed to catalyze the reaction. A number of factors may enable an enzyme to reduce activation energy. In 1894 Emil Fischer first introduced the "Lock and key" model to explain the phenomenon of enzyme specificity – in order for an enzyme to be specific to one substrate, the substrate must fit exactly into the active site of the enzyme. In the 1930s, Haldane and Pauling expanded upon this theory and hypothesized that enzymes increased reaction rates by fitting to the transition states of substrates.[4,5] The Haldane-Pauling hypothesis suggests a static enzyme structure, however, it is now becoming clear that protein flexibility may play a role in the catalytic efficiency of an enzyme.[4,5] Enzyme flexibility allows the enzyme to alter its conformation – this may need to occur for a substrate to fit into the active site, for the catalytic machinery to reach the substrate, or for the enzyme to be released.

1.3 Basic enzyme structure

Enzymes are special types of proteins that catalyze chemical reactions. While not all proteins are enzymes, all enzymes are proteins (though some catalytic units consist completely or partially of ribonucleic acid (RNA) as well – proteins that include RNA are called ribozymes). Proteins are polymers of amino acids. Amino acids contain three functional groups: an amine group, a carboxyl group, and a side chain. The side chain distinguishes each amino acid. Side chains can be basic, acidic, or hydrophobic. Some contain large aromatic groups, while others contain very small side chains. While the majority of proteins are constructed from just 20 amino acids, there are nearly 500 naturally occurring amino acids.[6]

The order in which these principal twenty amino acids occur in the polymer is called the **primary structure** of a protein. The primary sequence of a protein is determined by the sequence of the nucleic acids in the corresponding DNA gene. Nucleic acids code for amino acids using codons. This means that three nucleic acids code for one amino acid. After DNA is transcribed to mRNA by RNA polymerase, ribosomes translate the sequence of codons into a peptide chain of amino acids. The primary sequence determines how the protein folds into a three-dimensional shape.

The **secondary structure** of a protein is local segments of three-dimensional structure that are formed by hydrogen bonds in the peptide backbone. While side chains are not involved in these bonds, the characteristics of side chains influence the kind of secondary structure that they will be likely to form.[7] α-Helices are formed most easily by unbranched amino acids like methionine, alanine, leucine, glutamic acid, and lysine. β-sheets, on the other hand, are preferred by large aromatic residues (phenylalanine, tyrosine, tryptophan) and branched amino acids (isoleucine, threonine, and valine). Besides the α-helix, the most common helix structure, there are two other possible structures known as the 3_{10} helix and the π helix. An α-helix has 3.6 residues per turn and requires at least four residues per helix. The 3_{10} helix is tighter and only requires three residues. The π helix on the other hand is looser and requires five and has 4.4 residues per turn.

Tertiary structure consists of the three-dimensional shape of a protein as the secondary structural elements arrange in space. There are many common tertiary structural folds. These range from the very simple lone α-helix fold of some small peptides to much more complex structures containing alternating α-helix and β-sheet subunits.[8,9] The folding of a protein into its native tertiary structure is driven by the interactions of residues with the solvent (entropic effects) and with one another (enthalpic effects).[10–12]

As the Gibbs free energy equation describes, increased disorder (entropy) and loss of energy through heat (enthalpy) lead to a negative (favorable) Gibbs free energy. In the equation below, G is Gibbs free energy, H is enthalpy, T is temperature, and S is entropy.[13–15]

$$\Delta G° = \Delta H° - T\Delta S°$$

Enthalpic effects result from energy lowering interactions between resides, such as hydrogen bonds, ionic bonds (salt bridges), Van der Waals forces and covalent bonds (disulfide bridges). This decrease in enthalpy partially compensates for the loss of entropy that results from folding the free-moving peptide chain into a defined conformation.[13–15]

The major driver of folding, however, is the hydrophobic effect.[10–12,16] The hydrophobic effect stems from the tendency of water to form a hydrogen bond network around a hydrophobic surface. To reduce the number of water molecules organized at the surface interface and increase the entropy of the system, hydrophobic molecules aggregate. In the case of proteins, the molecule will fold so that hydrophobic surfaces are on the interior of the protein and hydrophilic residues face outward and interact with the solvent.

Some proteins have a **quaternary structure**. The quaternary structure of a protein is the interaction of multiple folded proteins to form a multi-subunit complex. Hemoglobin is a tetramer containing two α subunits and two β subunits. These subunits cooperate: when one subunit binds its substrate, oxygen, the next subunit will change shape, increasing its affinity for oxygen.[17] Quaternary structure can be held together by covalent means, such as disulfide bonds, but more commonly by non-covalent means such as hydrogen bonds. Often, hydrophobic surfaces are driven together by the hydrophobic effect previously described.[16]

Though a great body of knowledge exists regarding factors that drive protein folding, folding dynamics and stability of proteins remain very difficult to extrapolate from sequence alone.[13] Homology modeling – basing tertiary and quaternary structure on evolutionarily related structures – is a common method for aiding these predictions.[18] The ancestral enzymes studied in this work are based on homology models because no crystal structure is yet available.

1.4 Hydroxynitrile lyases

This thesis will focus primarily on one particular group of enzymes: hydroxynitrile lyases. Hydroxynitrile lyases (HNLs) are plant enzymes that catalyze the formation of hydrogen cyanide and an aldehyde from cyanohydrins as a protection against predators. There are at least 2500 known cyanogenic plant species in the world. Cyanogenic plants synthesize cyanohydrins from aliphatic and aromatic amino acids. Cytochrome P450 enzymes convert the amine group of the amino acid precursor into a cyano group. The final step of the synthetic reaction, catalyzed by UDP-transferase, adds a sugar group to the cyanohydrin, creating a cyanogenic glucoside. These cyanogenic glucosides are stored in vacuoles – when the plant tissue is injured by predators, cyanogenic glucosides are cleaved into cyanohydrins by β-glucosidase, the cyanohydrins are released, and hydrogen cyanide is released from the cyanohydrin. The release of hydrogen cyanide can either occur spontaneously at high pH or it can be catalyzed by HNLs.[19,20]

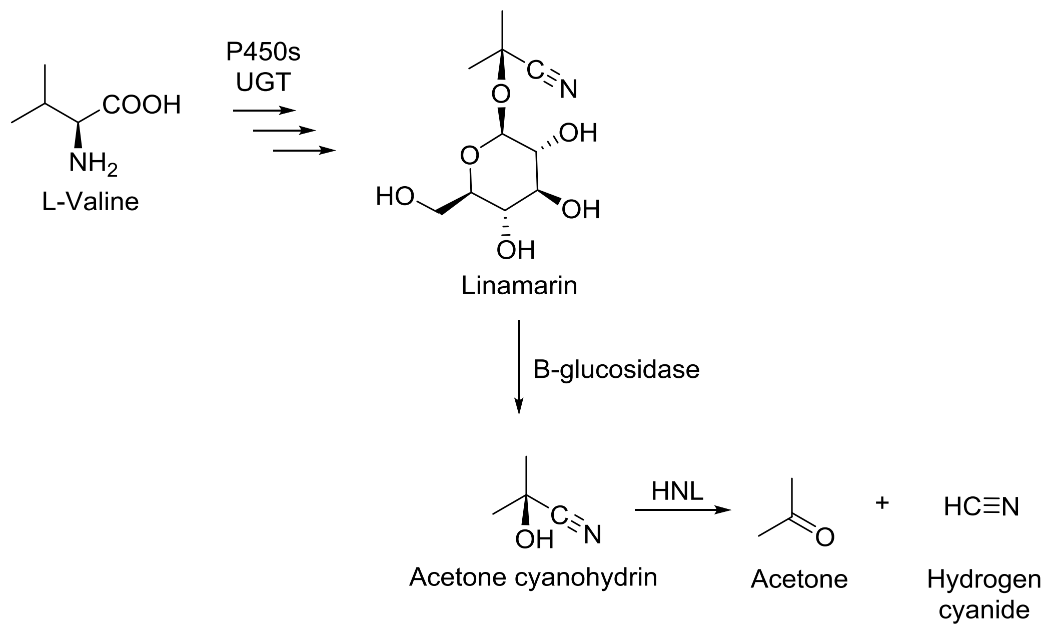


Figure 1.2 Cyanogenesis in plants. Cyanogenic glucosides are synthesized from amino acid precursors by P450s and UGTs. β-glucosidase cleaves the cyanohydrin from the glucoside to prepare the substrate for cleavage by HNLs.

Many common plants including bitter almonds (*Prunus amigdalis*), rubber trees (*Hevea brasiliensis*), and cassava (*Manihot esculenta*), contain cyanogenic glycosides.[20,21] The natural substrate of *P. amigdalis* HNL is mandelonitrile. The natural substrate of *Hevea brasiliensis* HNL (*Hb*HNL) is the much smaller acetone cyanohydrin. Shown below are common cyanogenic glycosides, their amino acid precursors and the corresponding cyanohydrins[19,20,22]:

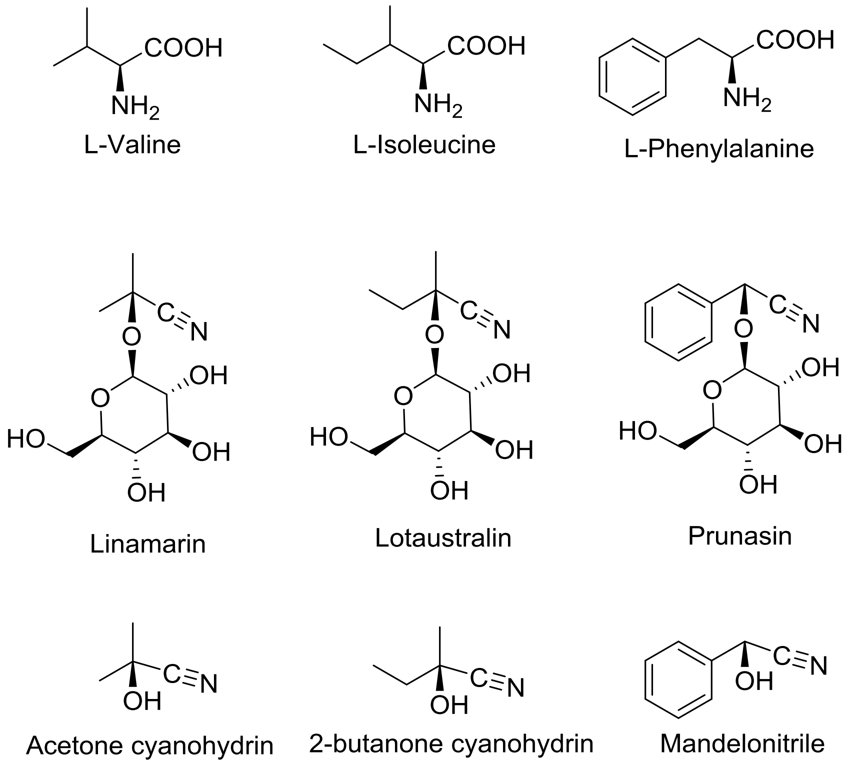


Figure 1.3 Cyanogenic glucosides and substrates of HNLs. Acetone cyanohydrin, 2-butanone cyanohydrin, and mandelonitrile are substrates of HNLs.

For other HNLs, the natural substrate remains unknown. Though *Arabidopsis thaliana* HNL can efficiently catalyze the release of hydrogen cyanide from mandelonitrile, but cyanogenic glycosides have not been detected in that plant, so its natural function is unknown.

1.5 Esterases

Esterases are another group of hydrolases in the α/β hydrolase family. Esterases split esters into an alcohol and an acid. For example, Salicylic Acid Binding Proteins (SABPs) produce salicylic acid from methyl salicylate. Plants produce Salicylic Acid to induce systemic acquired resistance (SAR), local resistance (LR) against invading pathogens, and hormone signaling.[23–25] Precursors to salicylic acid include phenylalanine and shikimate that are converted into salicylic acid by different pathways.[26] Salicylic acid levels increase in the presence of invading pathogens.[23–27] This increase appears to be dependent on levels of precursors produced through the shikimate.[26] Many SABPs are thought to have arisen from a common ancestor to the group of HNLs with which it shares similarity. The HNLs and esterases that this work centers on belong to the α/β hydrolase family.[28–30]

1.6 The α/β hydrolase fold

Apart from HNLs and SABPs, enzymes with the α/β hydrolase fold include the serine proteases, esterases and lipases among other enzymes including structural proteins.[8] The α/β hydrolase fold is one of the most common folds found in nature – a central β sheet surrounded by α-helices.[8] α/β hydrolases are comprised of eight β-sheets surrounded by six α-helices. Flexible loops between secondary structural elements vary in length and complexity to modify the shape and specificity of the active site.

Within all α/β hydrolases, the nucleophile is conserved within a strand-turn-helix motif near an "oxyanion hole," comprised of two backbone nitrogen atoms that stabilize the negatively charged transition state of the substrate. While catalytic activities among this family differ greatly, there is conservation among the catalytic residues with all members of this family including a histidine in the active site and a serine, cysteine, or aspartic acid nucleophile.[8]

Differences in the primary sequence can determine the catalytic activity of α/β hydrolases. In esterases, for example, the oxyanion hole is available to stabilize the oxyanion of the transition state of the substrate. This orients the serine near the carbonyl carbon, positioning it to act as a nucleophile. In hydroxynitrile lyases, on the other hand, the oxyanion hole is blocked by bulky residues - threonine in *Hb*HNL and asparagine in *Arabidopsis thaliana* HNL (*At*HNL).[31] In esterases, this position is occupied by a glycine, the smallest amino acid, leaving the oxyanion hole exposed.[32] Interestingly, Andexer *et al*. propose that AtHNL uses the oxyanion hole in HNL catalysis, though AtHNL is an inefficient esterase.[31]

The differences between esterases and HNLs, though these enzymes are closely related, greatly alter enzymatic activity. In *Hb*HNL, blocking the oxyanion hole prevents orients the substrate so that the hydroxyl hydrogen of the substrate hydrogen bonds with the serine.[31] This single amino acid difference changes the serine from a nucleophile to a hydrogen donor. Padhi *et al*. showed that changing a few residues can switch an esterase into a functional, albeit inefficient, hydroxynitrile lyase (**Figure 1.4**).[32]

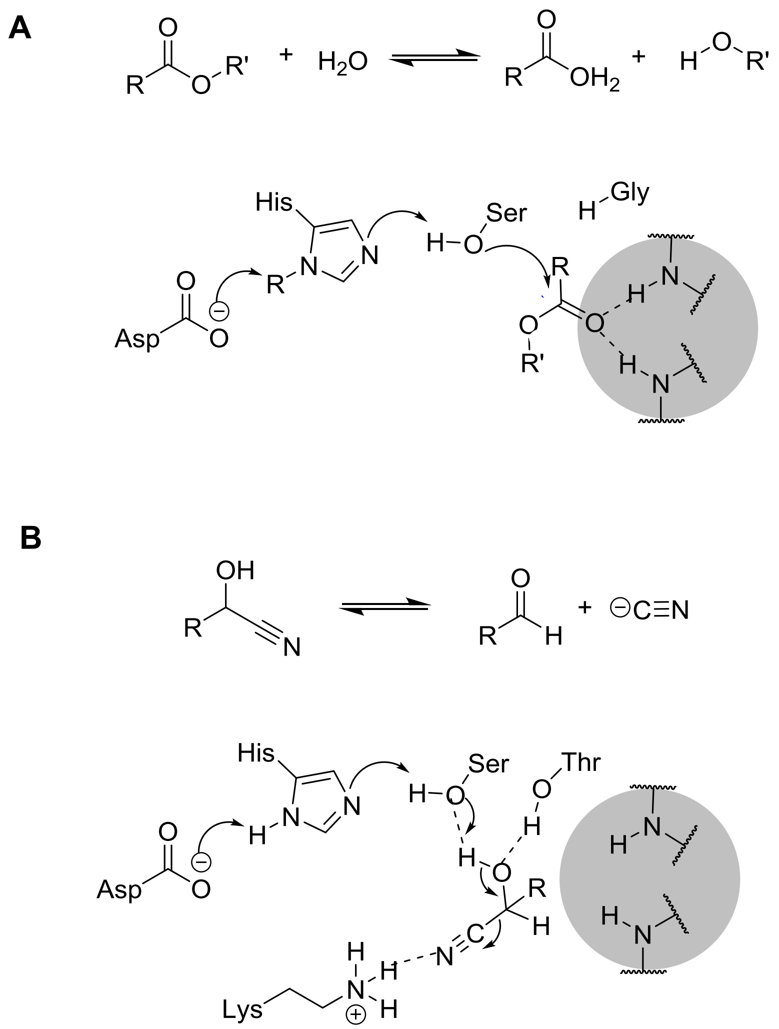


Figure 1.4 Esterase and HNL catalytic mechanisms. General reaction mechanism of A. esterase and B. HNL. The gray circle represents the oxyanion hole, which is blocked by threonine in the HNL mechanism. Diagram adapted from Steiner *et al*.[33]

1.7 Catalytic mechanisms of hydroxynitrile lyases

As with all members of the α/β hydrolase family, HNLs use the catalytic triad to produce hydrogen cyanide from cyanohydrins. With the exception of *At*HNL, most characterized hydroxynitrile lyases in the α/β hydrolase family are

(S)-selective.[31,34,35] In the catalytic mechanism of *Hb*HNL, the negative charge of the cyanide leaving group is stabilized by the positively charged Lys_{236} and the position of the aldehyde is stabilized by hydrogen bonds to the active site Thr_{11} and catalytic Ser_{80}.[28,30,31,36,37] In *At*HNL, however, Thr11 is replaced by asparagine and Lys_{236} is substituted with the uncharged methionine.[31] Andexer *et al.* suggest that these substitutions account for the (R)-selectivity of *At*HNL compared to the strong (S)-selectivity of *Hb*HNL.[31]

Catalysis occurs in *Hb*HNL when Ser80 forms a hydrogen bond to the hydroxyl group of the cyanohydrin. Ser80 removes the proton from the cyanohydrin while simultaneously donating its own proton to His235. The cyanide group, oriented by Lys236, acts as the leaving group (**Figure 1.4**). Orientation of active site residues can be seen in **Figure 1.5**.

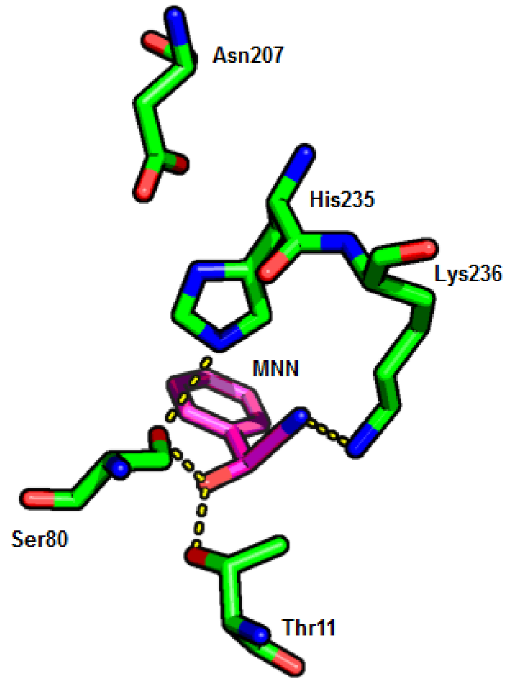


Figure 1.5 Active site amino acids of *Hb*HNL. Mandelonitrile (MNN) is shown in pink. Blue represents nitrogen and red represents oxygen. Green represents carbon atoms in the enzyme and magenta represents carbon in MNN. Shown are Lys236, His235, Asp207, Ser80, and Thr11. Hydrogen bonds are shown as yellow dashes. Hydrogen atoms are not shown.

One of the major questions remaining in the field is what the natural function of *At*HNL is. Originally characterized as an esterase based on sequence homology, *At*HNL shows very poor activity with esters. Likewise, the catalytic mechanism of *At*HNL is thought to be much different from that of *Hb*HNL, since *At*HNL is (*R*)-selective. In constrast to *Hb*HNL, no co-crystalization of *At*HNL and cyanohydrin has been accomplished, which makes understanding how the (*R*)-selective mechanism differs from the (*S*)-selective mechanism more difficult. Interestingly, there have been no cyanohydrins or cyanogenic glucosides found in *At*HNL, which implies that the natural function of the enzyme may not be that of a hydroxynitrile lyase.

1.8 Evolution of new activities through gene duplication

The catalytic diversity across members of the α/β hydrolase family is often attributed to differences in the loop regions that shape the active site. Many of these loop differences are found among the cap domains of the enzymes. The cap domain is a section of the enzyme of variable length and structure that projects from the canonical α/β hydrolase structure. In **Figure 1.6**, the black shaded area is the cap domain structure, which extends over the eight β-sheets, shaded gray, and the six α-helices, shaded white. How did so many different catalytic activities within this family evolve?

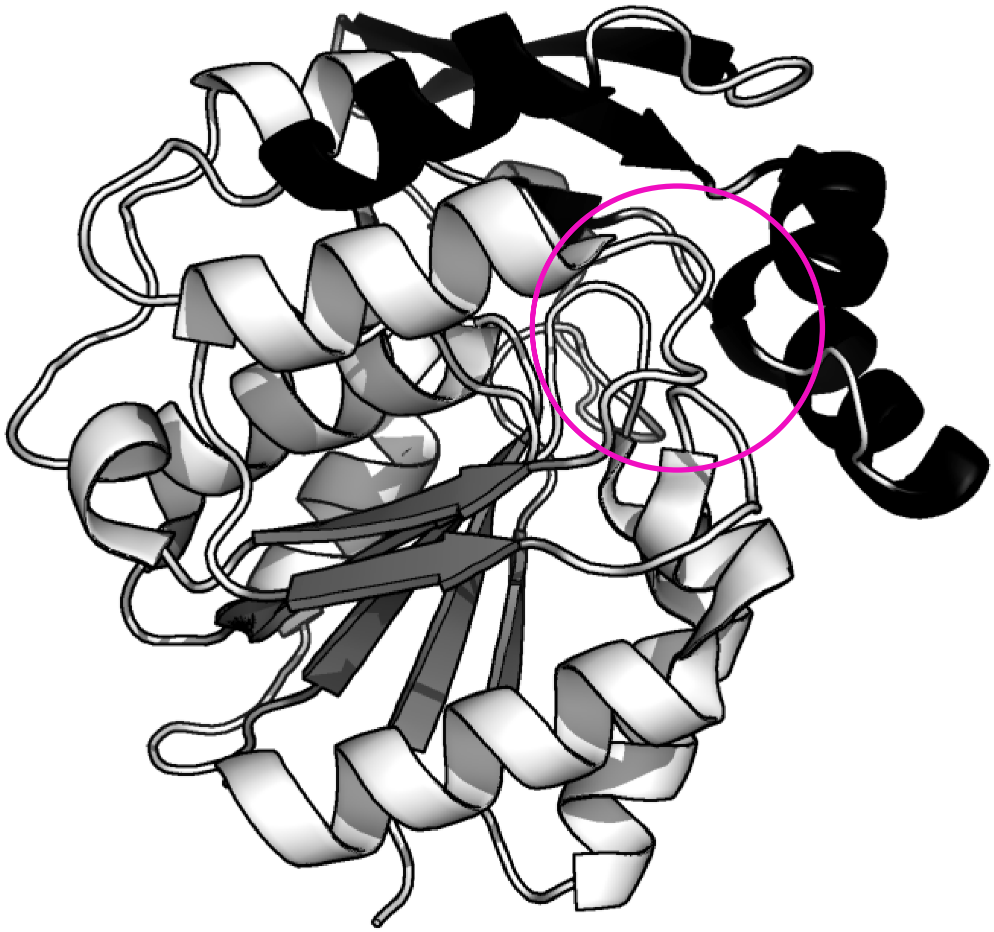


Figure 1.6 Structure of an α/β hydrolase (*Hb*HNL). The six α-helices (white) and eight β-sheets (gray) comprise the canonical α/β hydrolase fold. The cap domain is shaded in black and loops over the active site area (circled). PDB: 1YB6.

One possible explanation is that genes encoding α/β-hydrolases were duplicated in early organisms. If only one copy of the gene was required to allow the organism to survive, mutations in the duplicated copy would not be detrimental to the organism. When selective pressures changed, if the duplicated copy of the gene had accumulated beneficial mutations while not under selective pressure, the organism would survive.[38] This is known as the mutation during non-functionality (MDN) hypothesis proposed by Susumo Ohno. It states that when a gene is released from selective pressure – for example, if there are two copies of the gene and only one is needed – previously "forbidden" or highly detrimental mutations would be allowed to take place. These forbidden mutations may result in an enzyme with a new activity.[39] For the MDN hypothesis to be correct, however, the mutated gene and its regulatory elements must not acquire so many mutations that they are no longer functional.[40]

Jensen built on Ohno's work by hypothesizing that the transition of a duplicated gene from coding for a protein of one activity to a protein catalyzing a new activity might be more likely if the original gene encoded a protein that was already multifunctional.[38,40] According to this hypothesis, the ancestor of modern α/β-hydrolases might have catalyzed many reactions within the same active site. This non-specific enzyme was considered promiscuous either because it could catalyze the same reaction with multiple substrates, could catalyze multiple types of reactions or both. Because a promiscuous enzyme already displays low levels of activity for multiple reactions, it is possible that minor mutations in the gene would result in increased activity levels for some of these promiscuous reactions. Over time, enzymes that were specific to different reactions (lipase, esterase, HNL, etc.) developed. These changes may have been energetically favorable or may have been kept due to new pressures in the environment. Bergthorrson *et al*. proposed an additional step between promiscuity and specialization: amplification of the gene. This model of evolution is referred to as the innovation, amplification, and divergence (IAD) model. Bergthorrson *et al*. hypothesized that in a case where a promiscuous activity catalyzed by an enzyme becomes important (innovation), such as when there is an environmental change, selective pressure then drives this gene to duplicate, which increases catalysis in the cell (amplification). Finally, the various copies of this gene can collect mutations over time (divergence) – some mutations might be improvements, while others will result in defective proteins or "junk DNA." [40] An improved copy of the gene might then be evolutionarily favored and go through the same selection process again, beginning with amplification.

This theory predicts that ancestral enzymes were more promiscuous than modern day enzymes. Voordeckers *et al.* showed that reconstructed maltase enzymes showed improved specific activities with other sugar substrates when compared to modern day maltases.[41] Likewise, reconstructed Precambrian β-lactamases showed improved catalysis of third generation antibiotics.[42] Even modern day enzymes catalyze low levels of promiscuous reactions, including HNLs, which catalyze the Henry reaction.[29,30,43] The Henry reaction will be discussed in more detail in **Chapter 3**.

1.9 Summary

The focus of this thesis is the characterization of natural HNL and esterase activities in both modern day and reconstructed α/β-hydrolases. Our hypothesis is that ancestral enzymes are more promiscuous and more evolvable than modern day enzymes. This is because modern day specialist enzymes may have arisen from ancient generalist (promiscuous) enzymes, whose genes, when duplicated, were free to accumulate mutations.

Chapter 2 focuses on the characterization and screening of ancestral enzymes compared with modern day (extant) enzymes. The modern day enzymes screened include HNLs from *Hevea brasiliensis* (*Hb*HNL), *Manihot esculenta* (*Me*HNL) and *Arabidopsis thaliana* (*At*HNL) and the esterase SABP2 from *Nicotiana tabacum* – therefore we initially screened the reconstructions for esterase and HNL activities. We additionally screened for unnatural activities including the nitroaldol (Henry) and an aldol reaction. The ancestral enzymes are reconstructions based on sequences of extant HNLs and esterases, including *Hb*HNL, *At*HNL, and SABP2. **Figures 2.1** and **3.2** show a phylogenetic tree and a summary of activity among reconstructed and modern day enzymes.

Chapter 3 focuses on further characterization of the nitroaldol mechanism of an ancestral enzyme. This ancestral enzyme, known as PHYL2, is most closely related to *Hb*HNL and *Me*HNL, which are both efficient HNLs. However, while PHYL2 can catalyze cyanohydrin cleavage, it is twice as efficient a nitroaldolase as *Hb*HNL. To identify regions of the enzyme that are important for nitroaldolase activity, we created several chimeric proteins by swapping residues from PHYL2 into *Hb*HNL. We found that this decreased the HNL activity of HbHNL but also generally decreased nitroaldol activity. We identified one substitution in *Hb*HNL, V43I, which improves nitroaldol activity. We also found that the site that is most detrimental to HbHNL activity with both the nitroaldol and the HNL reaction is not situated near the active site, but is located near the beginning of the cap domain.

Chapter 4 describes the investigation of the aldol reaction, a reaction similar to the nitroaldol, in ancestral enzyme PHYL4. PHYL4 is at an evolutionary node (the site of an inferred most recent common ancestor) between *Hb*HNL and *At*HNL, making it older than PHYL2. The aldol reaction often occurs through a Schiff base mechanism, usually through a neutrally charged lysine residue. Screening of PHYL4 with the aldol substrate suggested that this enzyme may catalyze this reaction. Using molecular dynamics simulations, we identified a lysine (Lys237) in PHYL4 with a pK_a (~7) lower than for the corresponding lysine residue (~10) in *Hb*HNL. UV/vis spectrometry detected a possible Schiff base formed between PHYL4 and pyridoxal phosphate (PLP), but not between PHYL4-K237M and PLP.

Chapter 2. Characterization of reconstructed ancestral α/β hydrolases

2.1 Esterases and HNLs belong to the α/β hydrolase family

Extant esterases and HNLs are modern day enzyme specialists, meaning that they catalyze a single reaction with high specificity. These modern day specialists possibly evolved from earlier generalist enzymes that demonstrated low levels of promiscuous activity.[38,41,42] Enzymes with broad substrate specificity would have allowed primitive organisms the ability to catalyze many necessary reactions. With increasing levels of metabolic complexity, however, it would have become advantageous for enzymes to specialize in order to more tightly regulate signals and focus energy on necessary production, rather than additional products.[38,39,41,42] In order for differentiation and specialization to occur, more than one copy of the original gene needs to be present in the genome.[39] Over time, accumulated mutations create specialization among copies of the gene.[38,39,41,42]

Here, we set out to reconstruct ancestral α/β hydrolases at the nodes where esterases and HNLs diverged from an ancient common ancestor. Within the α/β hydrolase family, there is a great variety of catalytic versatility. Members of the α/β hydrolase family, including esterases and HNLs, evolved by divergent evolution.[8,28,32,44,45] In general, α/β hydrolases are enzymes that catalyze nucleophilic addition to carbonyl compounds.[8]

We call these enzymes α/β hydrolases because of the characteristic alternating six α-helices on the outside and eight β-sheets at the core of the structure.[8,44] While amino acid sequences vary greatly among members of this family, this folds differ very little. All variants within this group have the same catalytic triad.[8,44] This catalytic triad consists of a polar/acidic nucleophile, a histidine, and an aspartate.[32] In addition the catalytic triad, the cap domain is a characteristic structure of the α/β hydrolase family modifies which substrates will fit within the active site.[8] Cap domains vary greatly among members of the α/β hydrolase family. For more information on cap domains, refer to **Figure 1.6**. Number of residues, α-helices, and β-sheets in the cap domain vary from enzyme to enzyme. This domain may even move to assist with catalysis, which has been shown to occur in lipases.[8,46]

The esterase and HNL groups are two of the most similar families in the α/β hydrolase superfamily. The catalytic Ser-His-Asp triad is conserved in both groups.[28,32,33,37] In esterases, substrates orient in the active site through their interaction with the oxyanion hole, which uses two N-H groups from the main chain. This orientation allows the nucleophilic serine attacks the carbonyl carbon, forming a tetrahedral intermediate.[32,33] The carbonyl then reforms, cleaving the ester bond, after which water attacks the newly formed carbonyl freeing the enzyme.[32,33] In hydroxynitrile lyases, however, substrates orient the carbonyl oxygen instead of the carbon towards the catalytic serine, resulting in the serine acting as a hydrogen donor instead of a nucleophile.

Interestingly, a few changes can interconvert the activities of esterases and HNLs. As Padhi *et al*. found, they could turn esterase SABP2 into an HNL by simply swapping two amino acids.[32] The mutant completely lost the ability to catalyze ester hydrolysis and gained the ability to cleave mandelonitrile.[32] This two amino acid substitution can convert an esterase into an HNL because the substitution blocks the oxyanion hole with a bulky side chain residue, whereas in esterases, the amino acid in this position is a glycine, leaving the oxyanion hole open (refer to **Figure 1.4**).[32,33]

Another difference between *Hb*HNL and SABP2 is at position 79, which affects the reactivity of the lysine.[47] In *Hb*HNL, it is a glutamate, while in SABP2, it is a histidine. In *Hb*HNL, the negative charge of Glu79 raises the pK_a of Lys236 by stabilizing the positive charge of the protonated lysine over the neutral charge of the non-protonated lysine.[48–50] This positively charged lysine can then hydrogen bond to the cyano group. In esterases, the histidine at this position has the opposite effect on Lys236. Because both histidine and lysine are basic residues, the positive charge of the lysine becomes destabilized and the neutral form of the residue is favored.[48,51] This deprotonated lysine cannot hydrogen bond as well to the the cyano group. In fact, Nedrud *et al*. found that in addition to making the oxyanion hole available, mutating Glu79 to histidine was necessary to convert *Hb*HNL into an esterase.[45,47]

2.2 Resurrecting ancestral α/β hydrolases

We hypothesized that ancestral α/β hydrolases would be more promiscuous than modern day enzymes. In essence, the ancestral α/β hydrolase from which esterases and HNLs descended would show some activity with both cyanohydrin and ester substrates.[41,42] Likewise, assuming that the ancestral reconstructions are representative of historical enzymes at these nodes, these reconstructions could provide information on the pathway of evolution that was taken by α/β hydrolase esterases and HNLs at the molecular level.

A simplified phylogenetic tree of ancestral reconstructions of α/β hydrolases is shown in **Figure 2.1**. Dr. Antony Dean reconstructed enzymes at each of these four nodes, here referred to as PHYL2, PHYL4, PHYL6, and PHYL8, using all available, non-redundant, sequences of hydroxynitrile lyases. For simplicity, only well characterized extant enzymes that were used in the reconstructions are shown.

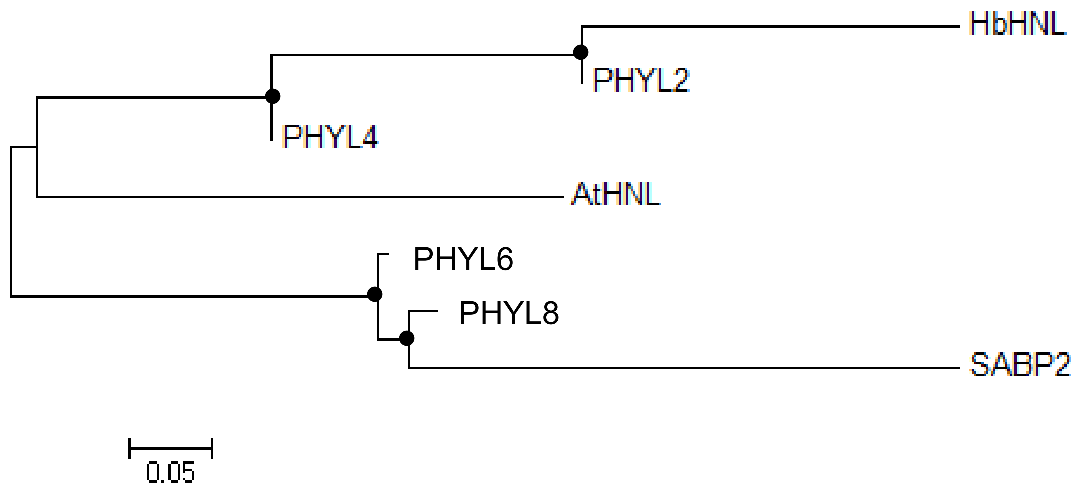


Figure 2.1 Simplified phylogenetic tree of modern day enzymes and ancestors. Black circles represent ancestral evolutionary nodes (sites of inferred most recent common ancestors).

PHYL2 and PHYL4 were expected to behave more like HNLs based on the sequence similarity between PHYL2, PHYL4, and known HNLs. While this reconstruction could potentially provide information on the timeline along which *R* enantioselectivity developed, critical mutational information has been lost throughout time, possibly rendering any reconstruction this ancient non-functional. PHYL6 and PHYL8 were hypothesized to be esterases, as their sequences are more similar to known esterases.

There are a variety of statistical methods commonly used for inferring the sequences of ancestral enzymes. These approaches use modern day sequences to estimate parameters of the model.[52] In this model, the parameters would be the phylogenetic tree, the rate of evolution, and the difficulty of transformation.[52] From these parameters, the sequences of ancient ancestral enzymes can be inferred. The three most commonly used methods for sequence based gene reconstruction are Maximum Parsimony, Maximum Likelihood, and Poisson.

The first of these is the **Maximum Parsimony** method which is the most commonly used method by evolutionary biologists to construct phylogenetic trees.[53] “Parsimony” means the simplest explanation; therefore, a maximum parsimony tree is the model in which the fewest genetic events need to take place. **The Maximum Likelihood** method differs from the maximum parsimony method in how it determines which model fits the data the best. “Likelihood” refers to the probability of calculating the observed data points given the model generated from those data.[52] The **Poisson** method predicts the number of mutation events that are likely to occur given a fixed rate of evolution. [54]

2.3 Percent identity of ancestors with modern day enzymes

Percent identity measures the exactly matching amino acids between two proteins. Percent similarity measures the number of amino acids that are an exact match or are chemically similar between two proteins. Percent identity and similarity can be used to gauge the relatedness of two proteins and to infer their evolutionary history.

PHYL6 and PHYL8 share 73% and 74% identity with SABP2 respectively, while PHYL2 and PHYL4 only share 51% and 56%. When we compare the ancestral enzymes to *Hb*HNL, however, the trend reverses. PHYL2 shares 81% identity with *Hb*HNL, while PHYL2 shares 69%, PHYL6 shares 51%, and PHYL8 shares 50%. Interestingly, PHYL2 shares only 58% identity with AtHNL, while PHYL4, the other ancestral HNL, shares 67%. In fact, *At*HNL shares more identity with the esterase ancestors, PHYL6 and PHYL8 (62% and 60% respectively) than it does with PHYL2 (**Table 2.1**). This data supports our interpretation that PHYL2 and PHYL4 are more closely related to extant HNLs, though PHYL2 is a more recent ancestor. The similarity of PHYL6 and PHYL8 to modern day esterases suggest that these ancestral reconstructions are more closely related to esterases than HNLs.

Table 2.1. Percent identity of ancestral enzymes compared with modern day enzymes

	PHYL2	PHYL4	PHYL6	PHYL8
*Hb*HNL	81%	69%	51%	50%
*At*HNL	58%	67%	62%	60%
SABP2	51%	56%	73%	74%

All four ancestral enzymes contain the α/β hydrolase serine-histidine-aspartate catalytic triad. Like all esterases in the α/β hydrolase family, PHYL6 and PHYL8 have a glycine instead of a bulky residue near the oxyanion hole (asparagine in PHYL4 and *At*HNL, threonine in PHYL2 and *Hb*HNL). The glycine does not block the oxyanion hole like the bulkier residues, which allows the enzymes to function as esterases. PHYL4 shares the highest percent identity of the ancestral α/β-hydrolases with *At*HNL, though it shares slightly higher percent identity with *Hb*HNL. Like *At*HNL, PHYL4 shares many similarities with the esterases as well. Like SABP2, PHYL6, and PHYL8, PHYL4 contains a histidine at position 79 instead of a glutamate. However, like *Hb*HNL and PHYL2, it contains a lysine at position 237 instead of a methionine. Unlike the esterases, PHYL4 and *At*HNL have a bulky asparagine in place of a glycine at position 12.

2.4 HNL and esterase activity of reconstructed α/β hydrolases

To characterize the activity of the ancestral α/β hydrolases, enzymes were expressed and screened for activity with cyanohydrins and esters. All ancestral reconstructions expressed and were soluble. As predicted, HNL activity of PHYL2 and PHYL4, closer to the HNL node of the α/β hydrolase phylogenetic tree, shows more activity for cleavage of cyanohydrins than for ester substrates. Of the ancestral reconstructions, PHYL2 has the highest activity for cyanohydrin cleavage.

PHYL2 was predicted to be (*S)*-selective in terms of enantioselectivity since enzymes at this node are representative of ancestors of *S* selective HNLs. PHYL4, however, is at a node between *R* and *S*-selective HNLs. Though PHYL4 should be an HNL based on its sequence, this enzyme shows very low activity for cyanohydrin cleavage (Table 2.3). Because this enzyme is reconstructed from both (*S)* and (*R*) selective HNLs, which contain very different catalytic mechanisms, the resulting PHYL4 may represent a combination of catalytic mechanisms that do not function well together. It is possible that over time, the genetic information that would explain the divergence of (*S)* and (*R)* specific HNLs was lost. Many rapid mutations without gene duplication to preserve the precursors would result in a loss of information that coded for the original enzyme. PHYL6 and PHYL8 both show higher esterase activity than HNL activity, also as predicted. All activities were lower than the modern day α/β hydrolases. If the ancestral enzymes are generalists, it would be expected that they would catalyze activities at lower rates than specialists.[41,42,45,55] This will be determined in further investigations into the promiscuity of the reconstructed ancestral enzymes.

Table 2.2 HNL activity and esterase activity of ancestral enzymes measured in k_{cat}, min^{-1}. Esterase activity was measured as hydrolysis of p-nitrophenyl acetate at pH 7.2 in BES buffer with 10% acetonitrile. HNL activity is measured as release of benzaldehyde from racemic mandelonitrile and pH 5.0. The detection limit of the assay is 0.0002 s^{-1}.

Enzyme	Esterase activity	HNL activity
*Hb*HNL	0.06	110 [56]
*Me*HNL	0.01	1300 [57]
SABP2	7.8	<0.0002
PHYL2	<0.0002	35
PHYL4	<0.0002	0.4
PHYL6	10	0.4
PHYL8	5	0.2

2.5 Discussion

While all of the ancestors showed activity similar to the activity the modern day enzymes that descended from that predicted node, there appeared to be little reaction promiscuity, the ability to catalyze a variety of different reactions using different chemistry, among both modern and reconstructed esterases and HNLs with the substrates tested. Another possibility is increased substrate promiscuity, or the ability to catalyze the same reaction from a variety of reactants. Voordeckers *et al.* showed that reconstructed ancestral maltases have increased substrate promiscuity for both sugars of maltose similarity and isomaltose similarity.[41] Likewise, Risso *et al.* showed increased substrate promiscuity towards third generation antibiotics in reconstructed β-lactamases.[42]

Other findings by Titu Devamani have shown that *Hb*HNL and *At*HNL are both highly enantioselective (>99%), while the ancestral enzymes at these nodes (PHYL2 and PHYL4) are all less enantioselective.[34,58] Not surprisingly, all of the enzymes constructed at node 2 (reconstructed common ancestor of *Me*HNL and *Hb*HNL) were (*S*)-selective. While *Hb*HNL is (*R*)-selective and *At*HNL is (*S)* selective, it is not known how these differences evolved or why.[31] The natural substrate of *Hb*HNL is acetone cyanohydrin, a non-chiral molecule. Moreover, cyanohydrins do not even exist in *Arabidopsis thaliana*.[59] We wanted to determine the enantioselectivity of ancestral HNLs, PHYL2 and PHYL4. Because the “descendants” of PHYL2 are *Hb*HNL and *Me*HNL, two highly (*S*)-selective HNLs, we hypothesized that PHYL2 would also be (*S*)-selective. The reaction catalyzed by PHYL4 would be expected to have no enantioselectivity, given that 1) it catalyzes the reaction poorly, 2) its active site is a mixture of an (*S)*-selective and an (*R)*-selective enzyme.

If our ancestral reconstructions are accurate, PHYL2 is an (*S*)-selective HNL, suggesting that the node from which *Me*HNL and *Hb*HNL evolved was already an HNL and already (*S*)-selective. The original PHYL4 reconstruction on the other hand showed very low esterase and HNL activities. It is possible that this is because the enzymes used to construct this ancestor have opposing enantioselectivities, and when used as a template, resulted in an inactive enzyme.

Likewise, PHYL6 and PHYL8 have been characterized as esterases based on sequence and activity. Further investigation into the activities of these enzymes may provide insight into the diversification of α/β hydrolases into esterases and hydroxynitrile lyases. It will be interesting to see if ancestral esterases can catalyze a broader range of reactions than modern day esterases.

We hypothesized that ancestral enzymes are more promiscuous than modern day enzymes. To test this hypothesis, we decided to screen the ancestral enzymes and modern day enzymes for activities with a variety of substrates. We chose substrates that might react through similar reaction mechanisms due to similarities in chemical properties and molecular size. In preliminary data, both ancestral enzymes and modern day enzymes showed varying levels of activity several substrates. PHYL2 showed especially good activity with 1-nitro-2-phenylethanol. The cleavage of this molecule into nitromethane and benzaldehyde is known as the Henry reaction. This reaction had previously been identified in *Me*HNL and *Hb*HNL, and will be covered in more detail in **Chapter 3** (See **Table 2.2**).[29,30,43] As expected, modern day esterase SABP2 was a poor catalyst for HNL activity, as were the ancestral enzymes PHYL6 and PHYL8. HbHNL and MeHNL as well as ancestral HNL PHYL2 show HNL activity but very low esterase activity. While PHYL4 is more similar to HNLs, it showed poor activity with both esters and cyanohydrins.

Even within modern day enzymes, there are low levels of promiscuous activity that can be detected. These promiscuous activities are not seen because the enzyme does not naturally come into contact with the substrate of the reaction. This could be because other enzymes have a higher affinity for this substrate or because it is an unnatural substrate that would not interact with the enzyme under natural conditions.[60]

Future directions include characterizing the flexibility of the ancestral reconstructions. Other research suggests that the flexibility of an enzyme may allow the enzyme to adapt multiple conformations, enhancing its ability to carry out a variety of activities.[42,61] B factor calculations on crystal structures of modern day and ancestral β-lactamases designed by Risso *et al*. suggest that ancestral β-lactamases are indeed more flexible than their modern day counterparts.[42] Even in modern day hydroxynitrile lyases, flexible residues are important for allowing larger substrates, such as *m*-phenoxy-mandelonitrile for *Hb*HNL, into the active site.[36]

2.6 Materials and Methods

Reagents. Reagents were purchased from commercial suppliers. Protein concentrations determined by Bradford assay from Bio-Rad per the manufacturers' instructions. Bovine serum albumin (BSA) from New England Biolabs was used as a standard. Pre-made sodium dodecyl sulfate polyacrylamide gradient gels (NuPage 4-12% Bis-Tris gel) and SafeStain Coomassie Blue Reagent were purchased from Invitrogen to visualize expressed proteins.

Plasmids. All constructs were cloned into a pET-28 vector (Novagen), which contain a 6X His-tag, a T7 promotor, and T7 polymerase.

Percent Identity Calculations. Percent identity calculations were carried out using NCBI BLAST with amino acid sequences. *Hb*HNL (PDB: 1YB6_A), *At*HNL (PDB: 3DQZ_C), *Me*HNL (GenBank: AAV52632.1), and SABP2 (UniProtKB/Swiss-Prot: Q6RYA0.1) FASTA sequences were downloaded from the NCBI database. Below are the protein sequences of the ancestral enzymes:

PHYL2:

MATAHFVLIHTICHGAWIWYKLKPLLEAAGHKVTALDLAASGIDPRQIEQINTFDE
YSEPLLTFMESLPQGEKVILVGESCGGLNIALAADKYPEKISAAVFHNALMPDTE
HSPSYVVDKFMEVFPDWKDTEFSTYTSNNETITGMKLGFKLMRENLYTNCPIED

YELAKMLTRKGSFFQNDLAQRPKFTEEGYGSIKRVYVWTDEDKIFPPEFQLWQI

ENYKPDKVYRVQGGDHKLQLSKTNELAEILQEVADTYADLLAVAGGGHHHHHH

PHYL4:

MQAKAHFVLVHNICHGAWIWYKLKPLLEAAGHKVTAIDLAASGIDPRQIEQVNTF

DEYSEPLLEFMESLPQNEKVILVGHSFGGLNIALAADKFPEKISVAVFLNALMPD

TEHSPSYVVDKYMEVPPGWRDTEFSPYGSPNETMTSMKLGFKLMRANLYQNC

PIEDYELAKMLVRQGSFFQEDLAKRKKFTEEGYGSVKRVYVMTNEDKAFPPEF

QLWQIENYNPNKVYEVKGGDHKVQLSKTQELADILQEVADNYADLLDVLGGGH

HHHHH

PHYL6:

MAEMKQQTKHFVLVHGACHGAWIWYKLKPLLEAAGHRVTALDLAASGINPRKI

EEVHTFDEYSEPLMELMASLPPNEKVILVGHSFGGLNLALAMEKFPEKISVAVFL

TAFMPDTEHRPSYVLEKYNERTPAEAWLDTQFSPYGMPEEPLTSMLFGPKFM

ANKLYQNCPIEDLELAKMLVRPGSLFIEDLSKAKKFSDEGYGSVQRVYIVCNED

KAIPEEFQRWMIENSGVNKVMEIKGADHMPMFSKPQELCQCLLEIANKYAKAG

DPLGGGHHHHHH

PHYL8:

MAEMKNRTKHFVLVHGACHGAWVWYKLKPLLEAAGHRVTALDLAASGINPKKI

EEVHTFDEYSEPLMELMASLPPNEKVILVGHSLGGLNLALAMEKFPEKISVAVFL

TAFMPDTEHRPSYVLEKYNERTPAEAWLDTQFSPYGNPEEPLTSMLFGPKFMA

NKLYQLSPIEDLELAKMLVRPGSLFIEDLSKAKKFSDEGYGSVPRVYIVCNEDKA

IPEEFQRWMIENSGVNEVMEIKGADHMPMFSKPQELCQCLLEIANKYAKAGDP LGGGHHHHHH

Site directed mutagenesis. For site directed mutagenesis, 0.5 μl template plasmid (50-200 ng/μl), 0.5 μl primers each of forward and reverse primers (2.5 μM stock), 0.25 μl of dNTPs (40 mM, 10 mM each dNTP), 1.25 μl Cloned *Pfu* buffer (Agilent Technologies), and 0.25 μl Cloned *Pfu* DNA Polymerase were combined and brought up to a final volume of 12.5 μl with sterilized deionized water. PCR was carried out according to the following program: 95°C for 5 minutes, followed by 35 cycles of 95°C for 50 seconds, 60°C for 50 seconds, and 68°C for 10 minutes, followed by 7 minutes at 68°C. After PCR, 0.25 μl *Dpn*I was added directly to the 12.5 μl of reaction mixture and incubated at 37°C for 3 hours to digest methylated template DNA. 2.5 μl of each reaction was transformed into 50 μl of chemically competent DH5α *E. coli* and plated onto LB plates containing 100 mM ampicillin. Colonies were grown overnight in LB with ampicillin and purified with the Qiagen Miniprep kit following the manufacturer's protocol. Plasmids were sequenced by ACGT, Inc. (Wheeling, IL). This protocol was adapted from Loening (2005) and Zheng et al. (2004).[62,63]

Protein expression and purification. Purified, sequenced plasmids were transformed in to chemically competent BL21 *E. coli*. Single colonies were picked and grown in an overnight LB pre-culture containing ampicillin. Pre-cultures were used to inoculate 300 ml TB containing 100 mg/mL ampicillin. Cultures were grown at 37°C to an OD_{600} of 0.8, after which the cultures were cooled to 17°C.

Once the cultures were cooled, IPTG was added to a final concentration of 0.5 mM and shaken overnight at 17°C. Cells were harvested by centrifugation and resuspended in Buffer A (5 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.0) and lysed by sonication (400 W, 40% amplitude for 6 min). Lysates were centrifuged at 13,000 rpm for 20 min at 4°C to remove cell debris.

The cell-free extract was loaded onto a column containing 3 mL of Ni-NTA resin (Qiagen) pre-equilibrated with 25 mL of buffer A. The column was washed sequentially with 20 column volumes of buffer A, buffer B (50 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.0), buffer C B (80 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.0) and finally eluted with 15 mL of elution buffer (125 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.0). The elution was concentrated to 1 mL and exchanged with BES buffer (5 mM, pH 7.0).

HNL activity. HNL activity was assayed as described in Nedrud *et al*.[46]

Esterase activity. Esterase activity was measured by hydrolysis of *p*-nitrophenyl acetate as described by Padhi *et al*. and Bernhardt *et al*.[32,64]

Chaper 3. Investigation and of Nitroaldol (Henry) activity in *Hb*HNL and ancestral enzyme PHYL2

3.1 The Henry reaction

The Henry reaction, also known as the nitroaldol addition, is an industrially useful reaction. Synthetic chemists use basic catalysts to synthesize nitroaldols from nitroalkanes, eg. nitromethane, and an aldehyde (**Figure 3.1**). This reaction is useful for the synthesis of a number of pharmaceuticals such as Amprenavir, an HIV protease inhibitor and Taxol, a chemotherapeutic drug,.[65,66] The Henry reaction is also used in the synthesis of antibiotics like L-acosamine and the production of substituted inositols.[65,67,68] Like cyanohydrins, nitroaldols serve as valuable intermediates for synthetic reactions. These intermediates can be further dehydrated or converted to amines.

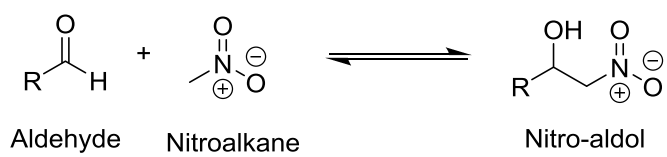


Figure 3.1 Schematic of the nitroaldol reaction. Nitromethane (nucleophile) adds across the carbonyl of an aldehyde to produce a nitroaldol.

Enantioselective methods for producing nitroaldols have long been of interest to synthetic chemists. Several catalysts, such as metallo-salts and non-enzymatic organocatalysts, have been developed to favor the production of one enantiomer over the other.[69] However, the effectiveness of these methods are reduced by long reaction times, harsh reaction conditions, and low final yields caused by the inevitable production of side products.[29,65,67,69]

In addition to cyanohydrin cleavage, *Hb*HNL catalyzes a nitroaldol reaction. Purkarthofer *et al.* hypothesize that in this reaction, nitro-methane replaces hydrogen cyanide as the nucleophile.[29,30,43,69] *Hb*HNL, which is (*S*)-selective for the HNL reaction, is also (*S*)-selective for the nitroaldol reaction. *At*HNL on the other hand is (*R*)-selective for both reactions.[31,35,43] Through docking simulations, Purkarthofer *et al*. predict that Ser80 and Thr11 make hydrogen bonds with the hydroxyl group of the product (*S*)-2-nitro-1-phenylethanol (NPE) in *Hb*HNL. Likewise, the nitro group of the nitroaldol hydrogen bonds to Lys236 – the residue with which the cyanide group of (*S*)-mandelonitrile interacts.[30] In addition to 2-nitro-1-phenylethanol, the formation of a variety of other substrates have been shown to be catalyzed by *Hb*HNL.[29]

Our goals were to 1) investigate the mechanism of the Henry reaction in PHYL2, 2) identify which residues and regions of the enzyme contribute to activity with NPE, and 3) screen mutations that improved the ratio of nitroaldol activity to that of cyanohydrin cleavage.

3.2 Catalytic mechanism of the Henry reaction

Based on sequence, PHYL2 and *Hb*HNL would be expected to have a similar tertiary structure. All of the key catalytic residues in *Hb*HNL are conserved in PHYL2, including the catalytic triad, Thr11 and Lys236. Overall, the enzyme is 81% similar to *Hb*HNL (See **Chapter 2** and **Table 2.1** for more detail on this topic). To identify sites that may be key for the improvement of the Henry reaction, we used molecular modeling to identify residues near the active site and then used site directed mutagenesis to swap residues at these positions from PHYL2 into *Hb*HNL.

Our lab had previously compared the catalytic activity of the ancestral and some extant enzymes with NPE and with MNN. To compare how well enzymes catalyze the nitroaldol reaction in comparison to the HNL reaction, we took the ratio of MNN activity over NPE activity (measured in k_{cat}/K_M).

$$Activity\ Ratio = \frac{Activity\ with\ MNN\ (\frac{kcat}{Km})}{Activity\ with\ NPE\ (\frac{kcat}{Km})}$$

The activity ratio of *Hb*HNL is 172, meaning that the k_{cat}/K_M is 172 times greater for activity with MNN than NPE. The activity ratio is only 24 for PHYL2. Our results show that the k_{cat} of *Hb*HNL for the Henry reaction is 0.12 s^{-1} – significantly lower than the k_{cat} of this enzyme for mandelonitrile cleavage (20.6 s^{-1}). Like *Hb*HNL, *At*HNL and *Me*HNL also catalyze the Henry reaction.[29,30,43,67,70] PHYL2, the reconstructed ancestral enzyme at the node between *Hb*HNL and *Me*HNL, catalyzes the nitroaldol reaction at a rate two-fold faster than *Hb*HNL. This indicates that while PHYL2 is still a better HNL than a nitroaldoase, it is a relatively better nitroaldolase compared to hydroxynitrile lyase than *Hb*HNL. For mutants in HbHNL, we measured the activity ratio of MNN/NPE activities. We decided to test whether swapping single residues in *Hb*HNL for residues in PHYL2 would show improvement for nitroaldol activity compared to HNL activity, as PHYL2 is relatively improved. We suspect that this reaction is being catalyzed by PHYL2 through similar mechanism as what was proposed for *Hb*HNL (**Figure 3.2**).[30]

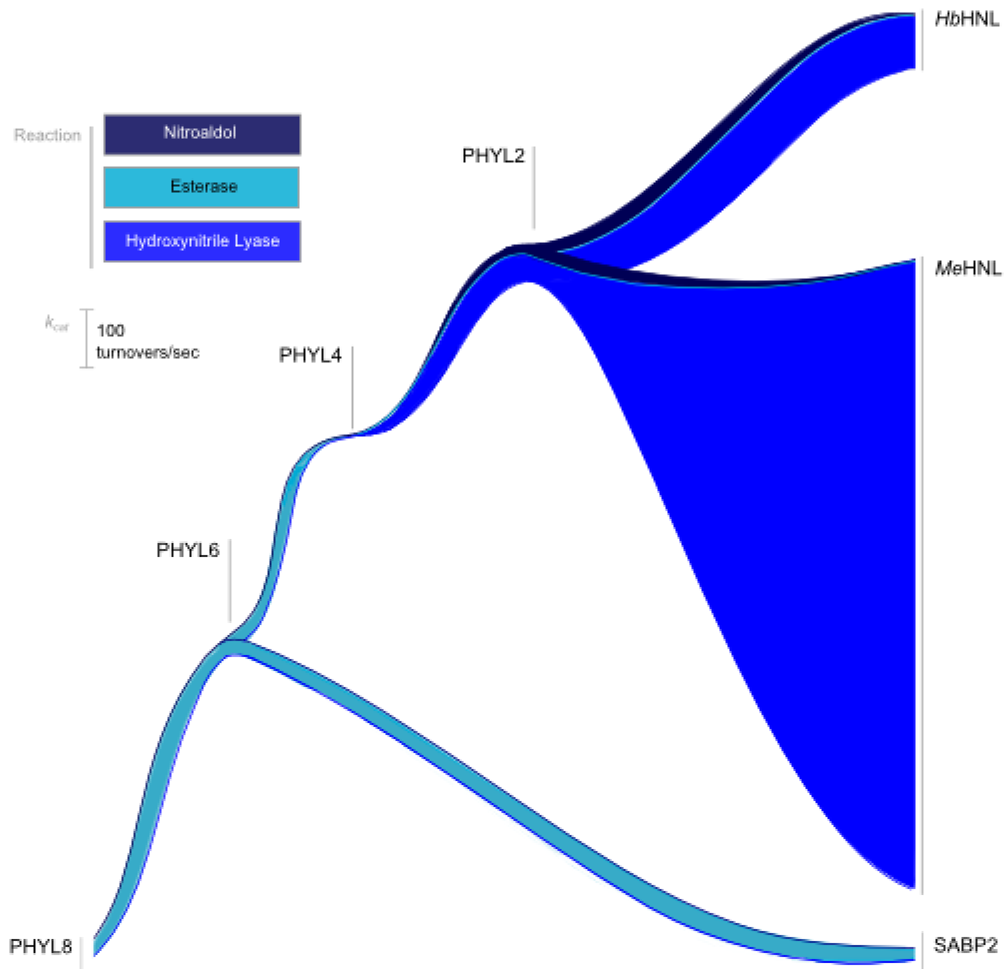


Figure 3.2 HNL, esterase, and nitroaldol activity of modern day and ancestral α/β hydrolases (reverse direction). Esterase activity was measured as hydrolysis of p-nitrophenyl acetate at pH 7.2 in BES buffer with 10% acetonitrile. HNL activity is measured as release of benzaldehyde from racemic mandelonitrile and pH 5.0. Nitroaldolase activity measured as release of benzaldehyde from racemic nitroaldol at pH 5.0. The detection limit of the assay is 0.05. Figure was drawn using InkScape

3.3 The active site of PHYL2 is larger than the active site of *Hb*HNL

Molecular modeling of PHYL2 and *Hb*HNL and their active sites suggested that the active site of PHYL2 is larger than that of *Hb*HNL. One possibility is that the improved k_{cat} of PHYL2 with NPE, compared to *Hb*HNL, is caused by creating more space for the larger substrate to fit into the active site.[61,71]

Molecular modeling with a homology model of PHYL2 predicted that the active site of the enzyme is larger than the active site of *Hb*HNL (**Figure 3.3**). However, visualization of the hypothesized active sites suggests that this extra volume is not in the area where the substrate docks, which suggests that it is not near the catalytic residues (**Figure 3.3**). If NPE is docked by aligning the phenyl ring and hydroxyl groups with MNN, the extra volume in the active site does not correspond to the location of the nitro group.[30]

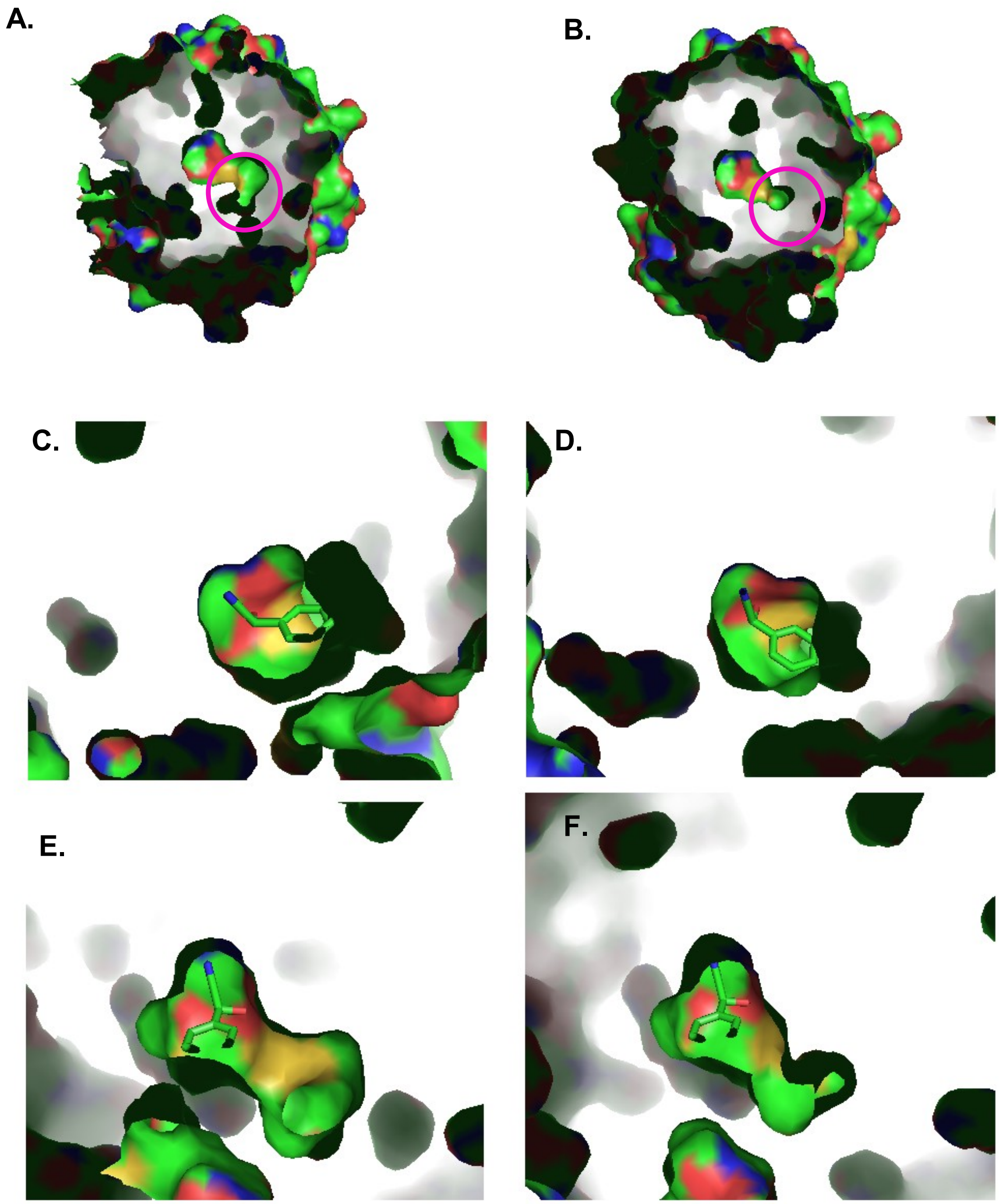


Figure 3.3 Additional space in PHYL2 active site is not near substrate. A. Homology model of PHYL2 based on *Hb*HNL, **B.** *Hb*HNL crystal structure (PDB: 1YB6). The location of the extra space in the active site is indicated by the pink circle. **C.** MNN modeled into PHYL2 as docked in PDB 1YB6. **D.** MNN in *Hb*HNL, similar angle to C. **E.** MNN modeled into PHYL2 shown from a different angle. **F.** MNN in *Hb*HNL active site, same angle as E. Atoms are color-coded: carbon (green), oxygen (red), sulfur (yellow), nitrogen (blue).

One possibility is that the extra space in the active site leads to increased promiscuity of PHYL2 compared to *Hb*HNL because the active site allows for a greater number of substrate orientations and sizes.[36,61] It is possible that because the substrate is more able to move around within the active site, it is free to make more hydrogen bonds and to not forced into a single specific orientation with catalytic residues, which might explain why PHYL2 is less efficient as an HNL but still able to catalyze that reaction as well as the nitroaldol reaction.[47,61,72]

Interestingly, modeling indicates that resurrected ancestral β-lactamases constructed by Risso *et al*. may indeed have a greater active site volume than their modern day counterparts, based on volumes calculated in our lab using the SiteMap program (Schrödinger). These volume calculations, however, are based on static PDB models which do not take into account conformational changes that could affect active site size. Flexibility, which is also increased in reconstructed β-lactamases, may also be involved in the promiscuity of these enzymes (**Figure 3.4**).

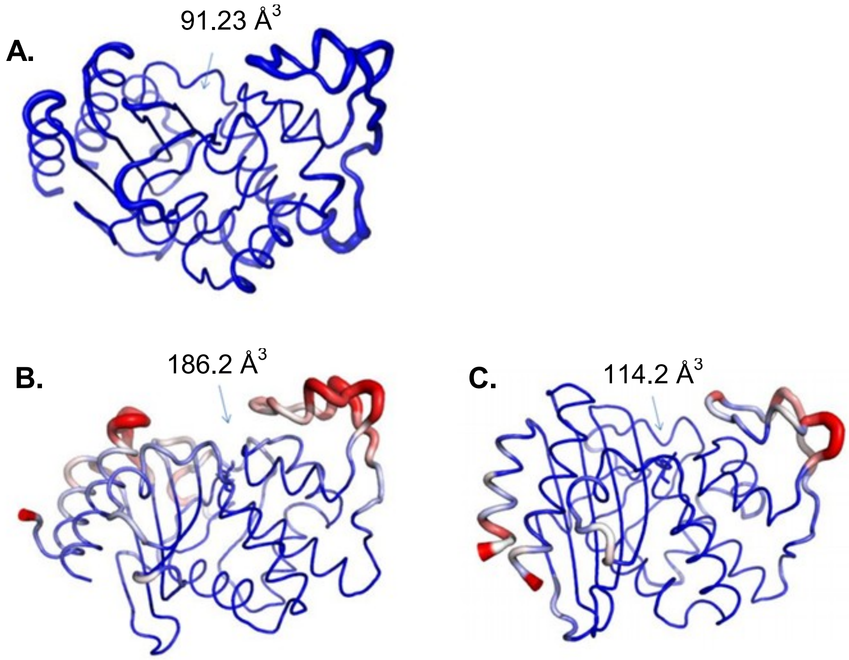


Figure 3.4 Flexibility and active site sizes of reconstructed β-lactamases. B factor measurements are visualized on a scale of $\leq$10 (blue) to $\geq$50 (red). Light blue arrows indicate active site, labeled in cubic angstroms above (Å^3). **A.** TEM-1 β-lactamase (PDB: 1BTL), **B.** Ancestral β-lactamase reconstructed by Risso *et al.* (PDB: 3DZJ), **C.** Ancestral β-lactamase reconstructed by Risso *et al.* (PDB: 4B88).

To further explore what makes PHYL2 different from *Hb*HNL, we decided to swap groups of amino acids from PHYL2 into *Hb*HNL. We decided to swap groups instead of individual residues 1) in order to reduce the possibility of negative interactions among amino acids when a mutation is introduced, and 2) to identify sections of the protein that contribute to the improved NPE activity of PHYL2. We first identified amino acids that were near one another in the primary sequence that were different between *Hb*HNL and PHYL2.

We specifically focused on mutations in the loop regions rather than on more rigid secondary structural elements because loops are generally assumed to be more amenable to substitutions. Flexible loops have a greater range of motion than helices and sheets because they are not restricted by internal hydrogen bonds.[73] Certain substitutions, like proline, would possibly be tolerated in a loop but would likely act as a secondary structure breaker in a helix or sheet.[7,74] Likewise, loop structures allow greater number of torsional angles for side chains, allowing rearrangement when an amino acid of a different charge is introduced.[7,61,73–76]

Areas where substitutions between the two enzymes appear range from near the active site, on the cap domain, and on the opposite side of the enzyme from the active site. Between PHYL2 and *Hb*HNL, there are 49 amino acids that are different (19% of the total sequence). Of these 49 differences, 23 of them are on flexible loops and 26 of them are on secondary structural elements. On the 4 loops closest to the active site (on the interior of the enzyme), there are 10 sites that are different between *Hb*HNL and PHYL2.

We began by swapping groups of amino acids that are located on loops because loops are often more amenable to substitutions than helices and β-sheets.[7] We arbitrarily chose six of these mutants to investigate further: V43I, C94P, K138S/D139N/G140N/K141E, L160N/G162P, Q206E/E208K/L211P, and K229R/E231Q. These mutations are varying distances from the active site of the enzyme. None of these residues appear to be involved with the catalytic mechanism. In total, only one of the substitutions tested (Q206E/E208K/L211P) between *Hb*HNL and PHYL2 is within 10 Å of the substrate, as based on the crystal structure of *Hb*HNL with MMN (PDB: 1YB6). All other substitutions range from 18.2 to 24.9 Å from the substrate.

Although these residues may not directly alter the catalytic mechanism, we hypothesized that they may alter the properties of the catalytic residues, modify the shape of the active site itself, or remodel key structural elements of the enzyme such as the cap domain that might affect the catalytic efficiency of the enzyme.[77,78] Sections of PHLY2 might also be more flexible than *Hb*HNL, which might allow the structure to conformationally adjust to the substrate.[77–79] We also hypothesized that by substituting groups of residues that occur near one another on the enzyme, detrimental interactions will be reduced. Our experiments aimed to 1) test whether swapping groups of neighboring residues improves the expression of protein mutants and 2) assess whether swapping sections of PHYL2 into *Hb*HNL would improve the catalytic activity of *Hb*HNL with NPE.

3.4 Expression

The mutant constructs produced soluble protein, however in much lower levels than wild-type *Hb*HNL and PHYL2 (**Figure 3.5**). This shows that switching sections of amino acids near one another might not be sufficient to efficiently reduce unfavorable residue interactions.

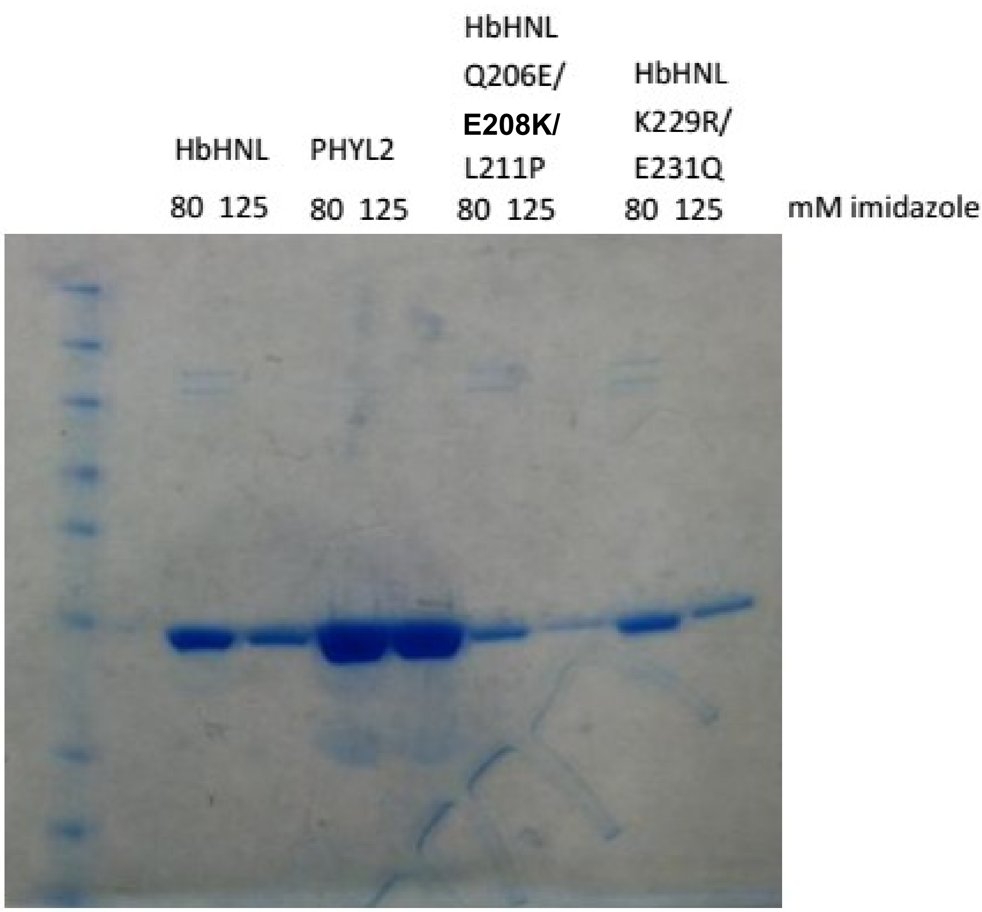


Figure 3.5 Expression of *Hb*HNL-PHYL2 swapped mutations. Expression was lower for *Hb*HNL K229R/E231Q and K138S/D139N/G140N/K141E mutants than for *Hb*HNL. All other *Hb*HNL-PHYL2 mutants showed lowered expression as well (data not shown). Elutions were from 80 mM and 125 mM imidazole

3.5 Mandelontrile (MNN) activity vs. nitrophenylethanol (NPE) activity in *Hb*HNL and PHYL2

All mutants tested showed varying levels of activity with MNN – all lower than wild-type *Hb*HNL but higher than PHYL2. This makes logical sense because PHYL2 shows lower activity with MNN than *Hb*HNL does. Interestingly, even mutants with substitutions positioned far from the active site showed decreases in activity, some more so than mutants with mutations closer to the active site. K138S/D139N/G140N/K141E, for example, reduced MNN activity in HbHNL to 38% of the wild-type activity level even though it is 24.9 Å from the active site (**Figure 3.6**). This suggests that even distal mutations in *Hb*HNL may shift regions of the enzyme that affect the active site.[77] This could also be the result of destabilization of the enzyme, as multiple mutations may destabilize the native folded state of a protein.

Table 3.1 Activities of *Hb*HNL-PHYL2 swapped mutants measured in mU/mg of enzyme as a percent of activity in *Hb*HNL.

Enzyme	% Remaining/MNN	% Remaining/NPE	MNN/ NPE
*Hb*HNL	100%	100%	4.24
Q206E/E208K/L211P	58%	74%	3.33
K229R/E231Q	54%	18%	12.4
C94P	57%	4%	67.5
V43I	42%	359%	0.491
K138S/D139N/G140N/K141E	38%	0%	-
L160N/G162P	66%	9%	30.5

One reason why PHYL2 might show lower catalytic activity than *Hb*HNL with MNN, but higher activity than *Hb*HNL with NPE is that there is mechanistic compromise occurring between the two enzymes. If this is the case, we would expect that *Hb*HNL mutants that showed lower catalytic activity with MNN would show higher catalytic activity with NPE. However, all mutants seemed to show decreased activity with NPE as well except for V43I (20.3 Å from the active site-

Figure 3.6 and **Table 3.1**), which show improved enzyme activity. Further testing and characterization of this mutant is needed to verify that this improved activity is real and to determine why it improves activity. This suggests that the characteristics of PHYL2 are not necessarily additive in terms of activity. If the mutations were additive, we would expect that every residue substituted into *Hb*HNL from PHYL2 would result in an increase in nitroaldol activity and a decrease in hydroxynitrile lyase activity. What we see instead, however, is that most substitutions result in decreases in both activities. It is likely that mutations need to occur in a certain order to yield a functional enzyme.

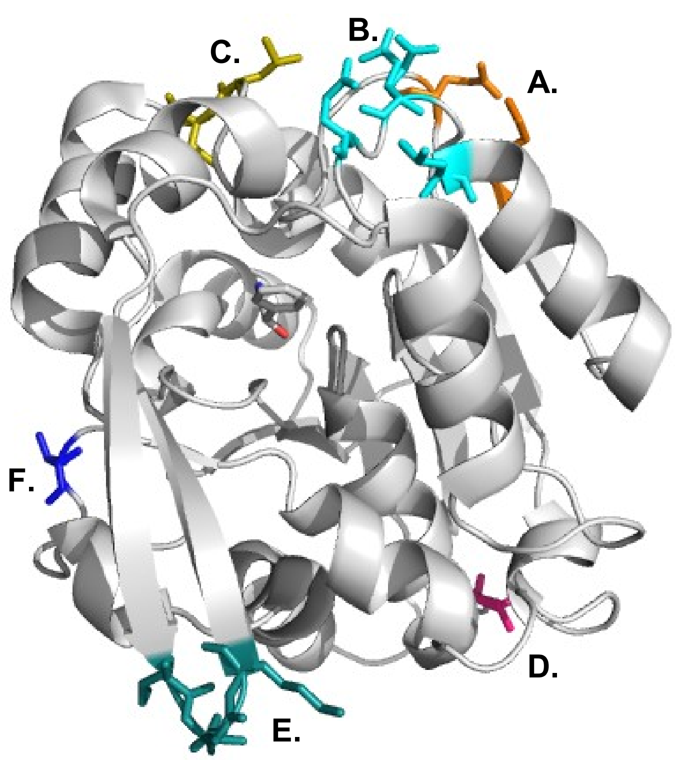


Figure 3.6 Sites mutated in *Hb*HNL to corresponding residues in PHYL2. Mutated sites are highlighted in Orange **A.** (K229R/E231Q – 21.9 Å from substrate), Cyan **B.** (Q206E/E208K/L211P- 9.8 Å), Yellow **C.** (L160N/G162P – 18.2 Å), Magenta **D.** (C94P – 26.7 Å), Turquoise **E.** (K138S/D139N/G140N/ K141E – 24.9 Å), and Blue **F.** (V43I – 20.3 Å).

3.6 Discussion

Currently, only homology models of PHYL2 and docking simulations of *Hb*HNL and PHYL2 with NPE are available. Likewise, it is difficult for molecular modeling software to predict dynamic conformational changes that occur during catalysis.[75] It is possible that structural elements of PHYL2 are dynamic and change conformation in response to different substrates more so than in *Hb*HNL.[8,44,61,76,80] Already, within the α/β hydrolase family, lipases have been demonstrated to have an open and closed cap conformation.[8,44,80] It might be possible that other members of this family have flexible elements that are involved in catalysis – especially in the highly variable cap domain region [8,44] Cap domains differ not only in sequence, but in secondary structure and length among proteins, which can contribute to the stability and flexibility of the domain structures.

Though the predicted secondary structure, including loop length, of PHYL2 is similar to *Hb*HNL, even primary sequence substitutions can alter flexibility. For example, Hiblot *et al.* showed that a single mutation in one loop increased the catalytic promiscuity of *Sso*Pox lactonase in correlation with an increase in the B-factor of the loop on which the residue position is located.[61] We are are working on determining conditions for the crystallization of PHYL2. The crystal structure of this enzyme will allow us to determine the flexibility of the PHYL2 active site. As shown by the ancestral reconstructions designed by Risso *et al*. on ancestral β-lactamases, structural flexibility may be an important characteristic of enzyme promiscuity.[42] This trend has also been noted previously in Cytochrome P450s – the more flexible the flexible the enzyme, the more promiscuous.[81] Flexible regions of an enzyme may contribute the catalytic step of a reaction in ways such as adopting multiple conformations or by bringing key residues into closer proximity of the substrate during catalysis.[77]

Proteins are networks of interactions, so substituting a single residue can have huge consequences on stability and other properties that are not directly related to catalysis.[77,78] Approaches like Combinatorial Active Site Test (CASTing) aim to ameliorate some of the problems caused by substituting residues.[33,82–84] CASTing substitutes two interacting residues simultaneously and aims to replace not just residues but residue interactions. This engineering approach has been successfully applied to tweak the activities and enantioselectivities of enzymes.[82–84] The difficulty with CASTing is predicting important interactions that maintain the balance of the protein network.

On the opposite end of the protein engineering spectrum from single and double site mutagenesis, entire protein domains from multi-domain proteins can be swapped to produce chimeric proteins.[85,86] Chimeric proteins are produced by fusing large sections of proteins to one another (e.g. the binding region of one protein to the catalytic region of another protein) to get an engineered protein with a novel function. In α/β hydrolases, an analogous method would be to swap the cap domains among proteins with this fold to determine whether activity can be changed. While the α/β architecture of α/β hydrolases is conserved, the structure and size of the cap domain can vary widely among groups within the superfamily.[8,44]

Our findings show that swapping groups of residues near one another in the primary sequence neither provides stability nor are these mutations additive (**Table 3.1**). While swapping PHYL2 residues into *Hb*HNL decreased MNN activity, these changes also decreased NPE activity. Interestingly, however, some distal mutations appeared to play an important role in the catalysis of MNN. This fits with the findings of Agarwal *et al*. that surface residues contribute the catalytic mechanisms through interconnected residue networks.[77,78,87] While the observation of interdependence of amino acid residues in a protein has been noted by others, it has also been shown that in very promiscuous enzymes – enzymes showing higher levels of "plasticity" – mutations are better tolerated.[88,89] Yoshikuni *et al*. used a mathematical algorithm that assumed additivity among mutations to design more selective sesequeterpine synthases from the highly promiscuous ɣ-humulene synthase.[88,90] It is possible that more promiscuous enzymes are more amenable to substitutions due to increased flexibility. It will be interesting to see if PHYL2 would better tolerate groups of mutations from *Hb*HNL and to see how this relates to structural differences between the two enzymes.

This experiment also suggests the importance of understanding the order of evolution of residues. While the groups of residues coexist to produce a stable, folded enzyme in PHYL2, when substituted into *Hb*HNL, they did not. One possibility is that mutations on and near the cap domain affect the overall structure of the enzyme. Although K138S/D139N/G140N/K141E is 24.9 Å from the active site, it was by far the most deleterious substitution made (**Table 3.1** and **Figure 3.6**). This group of residues is located on one of the loops of the cap domain – changing these residues may adjust the position of the cap domain to disrupt the overall structure of the enzyme. While these mutations (S,N,N, and E) reduce both the nitroaldol and hydroxynitrile lyase activities in *Hb*HNL, PHYL2 is able to function well with these residues. This suggests that other changes need to occur in *Hb*HNL before it is able to function like PHYL2. Determining evolutionary paths that might have been taken by enzymes may provide a key to understanding tolerated residues within a protein network and the order in which these must occur to maintain a stable, functioning enzyme, possibly because the evolutionary trajectory accounts for intermediary stabilizing mutations. The fact that the reconstructed ancestral enzymes in this work as well as other works, both express and are, for the most part, catalytically active suggest that following a logical evolutionary trajectory yields functional results.[41,42]

3.7 Future directions

Future directions for work on PHYL2 include improving the NPE activity of this enzyme to industrially useful levels. Work by Alissa Rauwerdink showed that substitution of a number of conserved residues had dramatic effects on the ratio of MNN activity to NPE activity. Positions E79 and H14 were the best mutants for lowering this ratio. When E79 was mutated to His, we saw a decrease in both

MNN and NPE activities, however, the activity ratio was lowered to 2, indicating that this residue site may be important for activity. In addition to these sites, a number of sites have been predicted through *in silico* and *in vitro* methods to be important for NPE catalysis. We plan to use saturation mutagenesis target these sites rationally as well as random mutagenesis to improve the activity of PHYL2 with NPE. We have also tried several in vivo screening methods, such as using nitro-methane as a nitrogen source for *E. coli*, detecting the formation of aldehydes using the colorimetric Schiff's reagent, and by using substrates with fluorescent R groups.[50,51,91] Many antibiotics contain nitro groups as the active functional group.[92,93] We searched the literature for potential antibiotics that could be cleaved by PHYL2 and *Hb*HNL through the Henry reaction mechanism, but have not found a good candidate yet. It will be interesting to see if these methods are more effective at evolving PHYL2 than they are for evolving HbHNL for improved activity with NPE.

3.8 Materials and Methods

Homology Models. Because crystal structures of ancestral enzymes were not available, we generated homology models of enzymes using SWISS-MODEL using *Hb*HNL (PDB: 1YB6) as a template.[18,94–96] SWISS-MODEL generates a three dimensional structure of a protein based on the input primary peptide sequence. The homology sequences were downloaded as PDB files.

Active Site Modeling. Active sites and surfaces were visualized using the PyMOL Molecular Graphics System.[97] The PDB file of PHYL2, generated from SWISS-MODEL, was aligned with the PDB file of *Hb*HNL (PDB: 1YB6). The active site was located by showing the MNN ligand in *Hb*HNL. Active site size between PDB models was compared by rendering the superimposed models as surfaces of different colors.

Active Site Size Calculation. To calculate the volume of active sites, we used the SiteMap software (Schrödinger).[98] SiteMap identifies the active site of a protein by selecting the ligand or by identifying locations on or near the surface of the receptor (the protein) that would be suitable for ligand binding. Sitemap then generates hydrophobic, hydrophilic, donor, acceptor, and metal binding regions using a virtual probe that simulates a water molecule in terms of size and electrostatic qualities. Volume was calculated using the “Manage Surfaces” command which provides a volume calculation. The algorithm determines the volume of the active site by drawing lines radially from points near the active site location. Any points for which less than 60% of lines do not contact the protein within 8 Å are excluded.

Site Directed Mutagenesis. Site directed mutagenesis was carried out as described in **Chapter 2**.

Primers. All primers were synthesized by Integrated DNA Technologies (Coralville, IA). The following primers were used for the loop swapping mutants in the *Hb*HNL gene: V43I (Forward: 5' GAC CTT GCA GCA AGC GGC ATT GAC CCA AGG CAA ATT G 3', Reverse: 5' CAA TTT GCC TTG GGT CAA TGC CGC TTG CTG CAA GGT C 3'), C94P (Forward: 5' CAA TTG CTG CTG ATA AAT ACC CGG AAA AGA TTG CAG CTG CTG 3', Reverse: 5' CAG CAG CTG CAA TCT TTT CCG GGT ATT TAT CAG CAG CAA TTG 3'), K138S/D139N/G140N/K141E (Forward: 5' GAC ACC ACG TAT TTT ACG TAC ACT AGC AAT AAT GAA GAG ATA ACT GGA TTG AAA CTG GGC 3', Reverse: 5' GCC CAG TTT CAA TCC AGT TAT CTC TTC ATT ATT GCT AGT GTA CGT AAA ATA CGT GGT GTC 3'), L160N/G162N (Forward: 5' CTT CTG AGG GAA AAT TTA TAT ACC AAT TGC CCG CCT GAG GAA TAT GAA CTG GCG AAG 3', Reverse: 5' CTT CGC CAG TTC ATA TTC CTC AGG CGG GCA ATT GGT ATA TAA ATT TTC CCT CAG AAG 3'), Q206E/E208K/L211P (Forward: 5' GTG TGG ACC GAC GAA GAT AAA ATT TTT CCG CCT GAA TTT CAA CTC TGG C 3', Reverse: 5' GCC AGA GTT GAA ATT CAG GCG GAA AAA TTT TAT CTT CGT CGG TCC ACA C 3'), K229R/E231Q (Forward: 5' CTA TAA ACC AGA CAA GGT TTA TCG TGT GCA GGG TGG AGA TCA TAA ATT GCA GC 3', Reverse: 5' GCT GCA ATT TAT GAT CTC CAC CCT GCA CAC GAT AAA CCT TGT CTG GTT TAT AG 3').

Protein expression and purification. Protein expression and purification was carried out as described in **Chapter 2**.

Hydroxynitrile lyase activity. HNL activity was measured as described in **Chapter 2**.

Nitroaldolase (Henry reaction) activity. We monitored for the release of benzaldehyde (ε_{280}=1352 $M^{-1}cm^{-1}$) from 2-nitro-1-phenylethanol at 280 nm and pH 5.0. The reaction was performed in microtiter plate with a total volume of 100 μL or 200 μL, which contained 2-nitro-1-phenylethanol (2 mM, 4 μL of 100 mM 2-nitro-1-phenylethanol in citric acid/acetonitrile buffer), citrate buffer (50 mM, pH 5.0) and varying concentrations of enzyme in BES buffer (5 mM, pH 7.2). All the reactions were performed triplicate. One unit of enzyme activity was defined as the amount of protein required for the release of 1 μmol of benzaldehyde per minute under the assay conditions.

Chapter 4. Screening ancestral enzymes for aldolase activity

4.1 The aldol reaction

The Henry reaction, catalyzed by HNLs, is similar in many ways to the pharmaceutically and industrially relevant aldol reaction.[99] The aldol reaction is one of the most industrially useful carbon-carbon bond forming reactions. Aldols serve as intermediates for a wide variety of products including perfumes, plastics, solvents, and pharmaceuticals. A few products of aldol reactions include widely used drugs such as Lipitor (a cholesterol-lowering drug), tetracycline (a chemotherapeutic drug), and multiple other pharmaceutical drugs.[100]

In terms of organic chemistry, the synthesis of aldols often uses amines which interact with the electrophile (the carbonyl group) to form a Schiff's base which stabilizes the transition state enolate.[99] In other cases, metallic catalysts can be used to stabilize the negative charge on the oxygen when the enolate forms. Many glycolytic enzymes work on phosphorylated substrates, such as fructose bisphosphate aldolase. Aldolases have been adapted to catalyze aldol reactions for pharmaceutical purposes. Deoxyribose-5-phosphate aldolase (DERA), for example, has been engineered to produce a double aldol side chain of cholesterol lowering drugs Lipitor (Pfizer) and Crestor (AstraZeneca).[100] Likewise, aldolases have been successfully computationally designed *de novo* following known aldolase motifs.[50,101]

4.2 Similarities between the aldol and Henry reactions

The Henry reaction and aldol reaction are similar in terms of the size and shape of reactants and final products. In both reactions, an aldehyde serves as the electrophile and the reaction of the aldehyde with a nitroalkane or a ketone results in a product containing an alcohol group. Because *Hb*HNL and ancestral enzyme PHYL2 catalyze the nitroaldol reaction, it seemed reasonable that these enzymes might be able to catalyze the aldol reaction as well. In both reactions, one of the reactants is an aldehyde and the resulting product contains a hydroxyl group.[69,99]

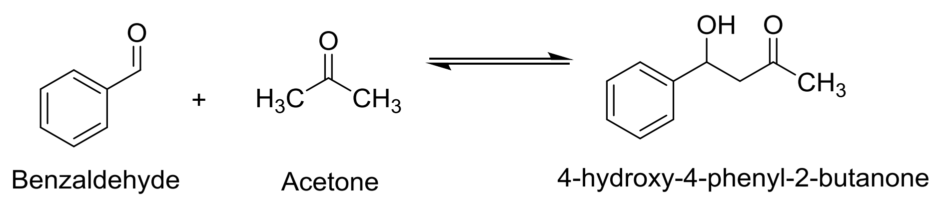
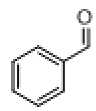
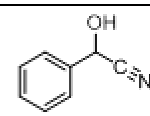
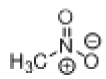
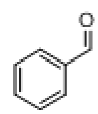
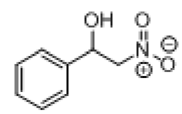
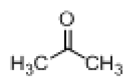
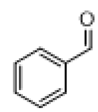
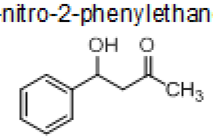


Figure 4.1 Aldol reaction of benzaldehyde and acetone. Acetone (nucleophile) attacks benzaldehyde to produce 4-hydroxy-4-phenyl-2-butanone.

The main difference between the two reactions is that the while the other reactant for the Henry reaction is a nitroalkane and the resulting nitroaldol product contains a nitro group, the corresponding reactant of the aldol reaction is a ketone, *e.g.* acetone, and the resulting molecule contains a carbonyl group. Nitroalkanes and ketoalkanes differ greatly in their reactivity. This is partly due to the stability of the activated nitroalkane (nitromethane) compared with activated ketoalkanes (acetone). The pKa of acetone (p$K_a \approx 19.2$) is 9 pK_a units greater than that of nitromethane (p$K_a \approx$ 10.2), meaning that nitromethane is ~10^9 times more acidic than acetone.[102,103] While the size and shape of acetone are similar to those of nitromethane, this difference in pKa significantly alters the chemistry of the reaction.

Table 4.1 Chemical characteristics of products and starting products of HNL, Henry, and Aldol reactions

Reaction	Nucleophile	pKa	Electrophile	Product
HNL	HC≡N hydrogen cyanide	9.21	 benzaldehyde	 Mandelonitrile
Henry	 nitromethane	10.2	 benzaldehyde	 1-nitro-2-phenylethanol
Aldol	 acetone	19.2	 benzaldehyde	 4-hydroxy-4-phenyl-2-butanone

4.3 Aldolase activity of modern day HNLs

While the less activated nucleophile of the aldol reaction differs substantially from the Henry reaction nucleophile, we hypothesized that because of the similarity in size and shape, it might be possible for HNLs to catalyze this reaction at low levels. We chose to measure the reverse direction of the reaction, the *retro*-aldol reaction, so that we could detect the production of one of the products, benzaldehyde, at 280 nm.

We hypothesized that this reaction would be catalyzed by the formation of a Schiff base with a Lysine, a common catalytic mechanism of Class I aldolases (in contrast to Class II aldolases that utilize a metal cofactor instead of a Schiff base to stabilize the transition state of the reaction). Within the active site of *Hb*HNL, there is a lysine at position 236 that could potentially form the catalytic Schiff base. We hypothesized that Ser80 may serve to deprotonate the hydroxyl group of 4-hydroxy-4-phenyl-2-butanone as it does in the HNL reaction, catalyzing the cleavage of the compound into benzaldehyde and acetone.

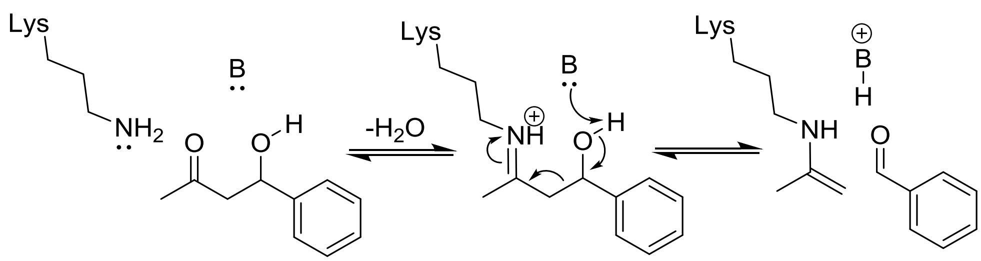


Figure 4.2 Class I *retro*-aldolase mechanism. The retro-aldol mechanism (aldol reaction in the reverse direction) uses a lysine to stabilize the Schiff base and a base to remove the hydrogen from the hydroxyl group.

The reaction of 250 uM 4-hydroxy-4-phenyl-2-butanone with *Hb*HNL resulted in no measurable activity when monitoring benzaldehyde concentration at 280 nm by UV/vis. Increasing enzyme concentration or substrate concentration did not result in increased activity, indicating that this was not below the level of detection (0.3 mU/mg).

In addition to *Hb*HNL, we also screened *At*HNL for *retro*-aldol activity. *At*HNL, unlike *Hb*HNL, contains a methionine residue in place of Lys236.

However, *At*HNL and *Hb*HNL contain a second lysine residue in a similar area of the active site at position 229 that could possibly catalyze this reaction. Similar to *Hb*HNL, *At*HNL showed no measurable activity with 4-hydroxy-4-phenyl-2-butanone.

4.4 Aldolase activity of ancestral enzymes

Because the framework of the HNL active site appears ideal for the catalysis of the aldol reaction, we next decided to screen reconstructed HNLs for activity with 4-hydroxy-4-phenyl-2-butanone. Among the four original ancestral reconstructions, only PHYL4 appeared to show very low activity for this reaction when measured by UV/vis at 280 nm. While PHYL4 did not show good activity with the HNL nor the nitroaldol reactions, it seemed possible that this was the result of a mechanistic tradeoff due to structural differences, meaning that in order for the enzyme to more efficiently catalyze a certain reaction, it might lose the potential to catalyze other reactions.[61,104] However, it was later shown that this absorbance was due to enzyme precipitation. Furthermore, this activity was unable to be reproduced by the same method and could not be replicated by HPLC. PHYL2, PHYL6, and PHYL8 also did not show activity with this substrate.

4.5 Detection of Schiff base formation in PHYL4

To see if Lys236 in PHYL4 could possibly form a Schiff base with 4-hydroxy-4-phenyl-2-butanone, we used the PROPKA algorithm (Schrödinger). In order for a lysine to form an imine bond with a substrate, the lysine side chain needs to be deprotonated to allow the lone pair of the nitrogen to attack the carbonyl group of the substrate. The average pK_a of lysine is 10.79, meaning that at pH 10.79, fifty percent of lysines in the population will be protonated.

The average pK_a of lysine is much higher than the pH of most environments in which enzymes naturally function.[101] This means that lysine would be protonated if their pK_a were that value. However, in most aldolases, the pK_a is generally lowered to a more neutral pK_a of between 6 or 7, meaning that at neutral pH, the amino group of the lysine side chain protonated half of the time. This perturbed pK_a of lysines in Class I aldolases is the result of the surrounding amino acids, such as the presence of nearby positive charges or non-polar micro-environments.[50,51]

Molecular modeling using the crystal structures of *Hb*HNL and a homology model of PHYL4 suggested that in *Hb*HNL, Lys236 has a pKa of approximately 9.9, close to the pKa of free lysine. The lysine at the corresponding position in PHYL4 has a much lower pKa – approximately 7, indicating that this residue would be half deprotonated at neutral pH. We compared primary sequences of PHYL4 and *Hb*HNL to identify substitutions that may result in the lowered pK_a of lysine in the ancestor. We discovered that Glu79 in *Hb*HNL was replaced by His79 in the ancestor. In modeling simulations, when Glu79 was replaced with histidine in *Hb*HNL, the calculated pK_a dropped from 9.9 to approximately 7. This also occurred when glutamate was replaced with lysine and arginine. Likewise, when His79 in PHYL4 was replaced with glutamate, the pK_a was raised to approximately 10.

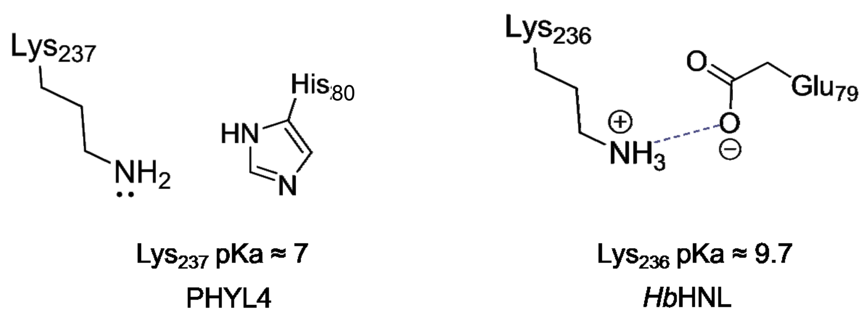


Figure 4.3 Active site lysines of PHYL4 and *Hb*HNL. In PHYL4, the lysine is adjacent to a positively charged histidine, whereas in HbHNL, a negatively charged glutamate forms a stabilizing hydrogen bond with the positive charge of the lysine, driving up the pK_a.

Glutamate is an acidic amino acid with a pK_a of approximately 4.07, meaning that this amino acid is mostly negatively charged at neutral pH.[105] In contrast, the histidine side chain is considered basic and has a pK_a of 6.04 in solution.[105] We therefore concluded that the higher calculated pK_a of the Lys236 in *Hb*HNL relative to PHYL4 was because of the stabilizing hydrogen bonds that could form between the protonated side chain amine and the negatively charged glutamate.

While we did not find evidence that the enzymes could completely catalyze the aldol reaction, it is not clear whether the reaction might be partially catalyzed. This would mean that the transition state would be partially formed within the active site, even if the final steps of the reaction that result in full catalysis are unable to happen.

Although this reaction is similar to the nitroaldol reaction, it is possible that the aldol is able to dock within the active site but is unable to be completely catalyzed due to the difference in the reactivity between acetone and nitromethane. Further investigation into whether this enzyme can form a Schiff base intermediate is necessary for determining whether HNLs can catalyze the *retro*-aldol.

UV/Vis measurements detected an absorbance peak that may correspond to the formation of an imine when PHYL4 is reacted with pyridoxal phosphate (PLP). PLP, also known as vitamin B_6, is a cofactor that binds via a Schiff base with lysines. PLP can only react with lysines that are deprotonated. The unreacted form of PLP has a λ_{max} of 390 nm. The reacted form, on the other hand, has a λ_{max} that is slightly shifted upward to 410 nm. Therefore, if PHYL4 reacted with PLP shows an absorbance peak at 410 nm, compared to the PLP blank, which absorbs at 390 nm, it would suggest that the lysine in PHYL4 is capable of forming a Schiff base.

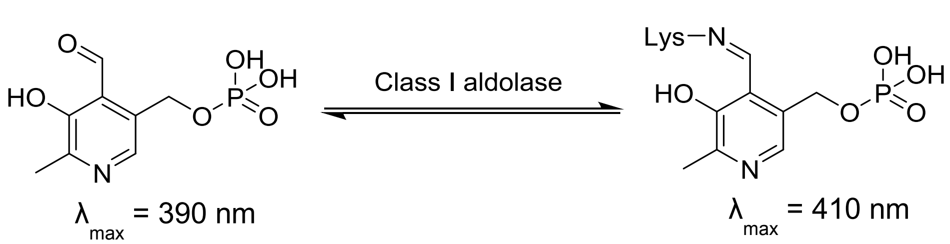


Figure 4.4 Pyridoxal phosphate (PLP) interaction with lysine. PLP shows a shift in λmax from 390 nm to 410 nm when it forms a Schiff base with lysine.

We found that PLP alone absorbed at 390 nm, whereas protein absorbed at 280 nm. When PHYL4 was reacted with PLP, there appears to be an absorbance shift in the spectrum from 390 towards 400 nm, though it does not appear to completely reach 410 nm. If this shift comes from a reaction of Lys237 with PLP, the absorbance less than 410 nm may be due to not all of the PLP reacting with the enzyme.

To see if the lysine reacting is the active site lysine, Lys237, and not surface residues or other lysines within the enzyme, we made mutant PHYL4-K237M. While methionine is similar in size and shape to lysine, it lacks an amine group and therefore cannot form an imine. Interestingly, when PHYL4-K237M was reacted with PLP, the absorbance spectrum more closely resembles that of unreacted PLP (λ_{max} = 390 nm) (**Figure 4.5**). This suggests that while the *retro*-aldol reaction cannot be fully catalyzed by PHYL4, the key step, imine formation, may be able to occur. Other steps in the reaction, such as the release of products from the enzyme, may be hindering the efficacy of this reaction.

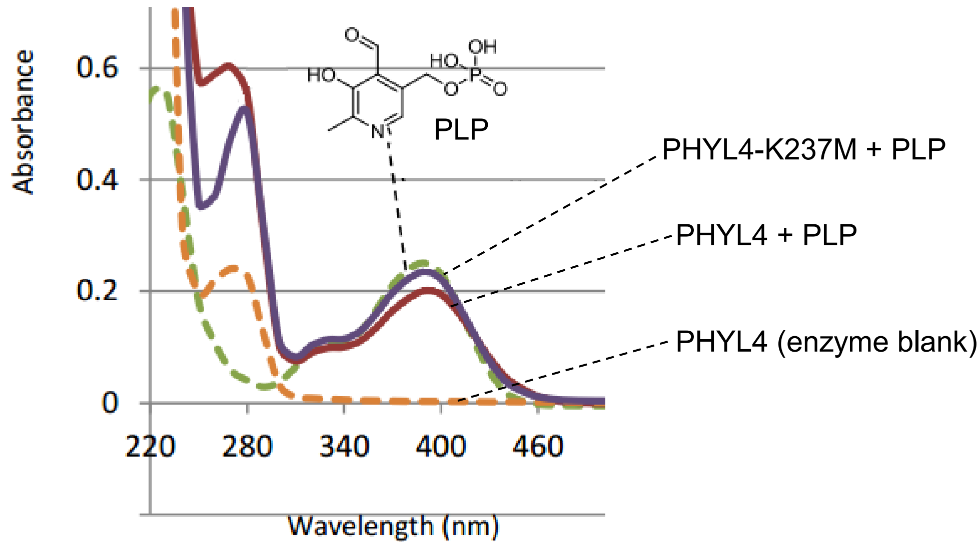


Figure 4.5 Detection of Schiff base in PHYL4 using PLP. PLP Blank (dashed green line), PHYL4 enzyme blank (dashed orange line), PHYL4 and PLP (burgundy), PHYL4-K237M and PLP (purple).

Further experiments will need to be conducted to verify if a Schiff base forms in PHYL4 at Lys237 with an aldol substrate. Imine bonds are transient; one possible way to verify the presence of the Schiff base would be to reduce the imine bond to a covalent single bond. To do this, we would use sodium cyanoborohydride ($NaCNBH_3$), a mild reductant that should only reduce imines. After reduction, the proteins could be analyzed by mass spectrometry. The residues bound to the substrate could be identified by the mass shift in the peptide caused by the substrate covalently bound to the side chain. The exact location of the residue can be identified by comparing the data with mass spectrometry data from unreacted PHYL4 to identify which peptide showed a shift. Further optimization of the $NaCNBH_3$ treatment of enzymes will need to be completed before carrying out this experiment.[106]

4.6 Discussion

Although modern day and ancestral HNLs have been shown to catalyze the nitroaldol reaction, the *retro*-aldol requires a more basic catalyst. The difficulty in stabilizing the enolate intermediate of the aldol reaction explains the difficulty in developing an enzymatic catalyst for this reaction. Naturally occurring Class I aldolases often work on phosphorylated substrates. Phosphorylated molecules are more reactive and less stable than non-phosphorylated molecules, reducing the energy needed for catalysis.[107] This instability is sometimes caused by charge-charge repulsion caused by the negatively charged phosphate group. A significant amount of modification would need to be carried out on most naturally occurring aldolases.

4.7 Materials and Methods

Synthesis of 4-hydroxy-4-phenyl-2-butanone. 10 mmol benzaldehyde (0.95 ml), 200 mmol acetone (15.76 ml), and 15 ml water were mixed by stirring at room temperature. 240 ul (30 mol% of amount of benzaldehyde) of amine (pyrrolidine) was added to the stirring mixture and let the reaction proceed 5 minutes. 1 N HCL was added until solution was neutral or slightly acidic. After quenching, the reaction was extracted with 100 ml chloroform. The reaction washed with 50 ml 5% sodium bicarbonate to neutralize the solution. The product was dried by adding Na_2SO_4 to remove water. Chloroform was evaporated and the product was stored at 4 degrees for one day. The compound was purified over silica gel and eluted in 85% Hexane/15% ethyl acetate.[108] The product was a pale yellow, transparent oil.

Product and purity was verified by ^{1}HNMR and TLC. R_f = 0.12 (ethyl acetate: hexane; 1:4). ^{1}H NMR ($CDCl_3$): (400 MHz, CDCl3) δ 7.27-7.38 (5H, m), 3.32 (1H, s), 5.17 (1H, m), 2.87 (2H, m), 2.21(3H, s).

Screening for *retro*-aldol activity. 4-hydroxy-4-phenyl-2-butanone was dissolved in DMSO to a final concentration of 25 mM. This was further diluted in citric acid buffer to 1 mM. A final concentration of 25 mM 4-hydroxy-4-phenyl-2-butanone was present in the final volume of 100 ul phosphate buffered saline (200 mM NaCl, 50 mM phosphate, pH 7.0). UV/Vis screening to measure release of benzaldehyde was carried out as in **Chapter 3**.

Enzyme Purification. Enzymes were purified as described in **Chapter 3**.

p*K*$_a$ Calculations. pK_a calculations were carried out using the PROPKA algorithm in the Maestro software (Schrödinger).

Homology Models. Because crystal structures of ancestral enzymes were not available before the writing of this thesis, we generated homology models of enzymes using SWISS-MODEL.[18,94–96] SWISS-MODEL generates a three dimensional structure of a protein based on the input primary peptide sequence. The homology sequences were downloaded as PDB files.

Capturing Schiff Base with Pyridoxal Phosphate. A 1 mL solution with a final concentration of 26 μM of enzyme and 125 μM in PBS (50 mM phosphate, 100 mM NaCl) was reacted for 45 minutes in the dark. Spectral measurements were taken by transferring 100 uL of the reaction mixture into a 96-well plate and reading the spectra from 200 nm to 600 nm (10 nm increments) on a Spectra Max Plus 384 machine.

Chapter 5: Conclusion

As discussed in **Chapter 2**, we reconstructed four ancestral enzymes at evolutionary nodes between the divergence of esterase and HNLs. All four of the original reconstructions based on the maximum parsimony algorithm resulted in soluble protein. While PHYL4 did not show good activity for any reaction for which it was screened, other reconstructions at this node using different algorithms (Poisson, maximum likelihood, as well as a different reconstruction based on maximum parsimony) have shown function with some substrates.

Ancestral HNL PHYL2 catalyzes the nitroaldol reaction with better efficiency than the *Hb*HNL. In **Chapter 3**, we investigated characteristics of PHYL2 that make it a good catalyst for this reaction. We initially thought that this might be due to the greater size of the active site in PHYL2 when compared to HbHNL. However, molecular modeling calculations predict that the active sites are not larger in PHYL2 than in *Hb*HNL.

In **Chapter 4**, we screened the ancestral enzymes for aldol activity, based on our findings with the nitroaldol reaction. While we initially thought PHYL4 showed some activity as an aldolase, this proved to be caused by enzyme precipitation. Interestingly, molecular dynamics studies do predict that PHYL4 would potentially form a Schiff base, a key part of *retro*-aldol activity, based on the pK_a of its active site lysine.

While it remains inconclusive if ancestral α/β hydrolases are overall more promiscuous than their modern day counterparts, our research can conclude that ancestral reconstructions based on modern day enzymes results in functional enzymes. We found that swapping residues between ancestral enzymes and modern day enzymes does not result in added activity and often resulted in poor enzyme expression and lowered efficiency. This suggests that not all substitutions between enzymes are additive – other substitutions are likely necessary for stabilizing the whole structure of the enzyme.

Further characterization of the ancestral enzymes at nodes PHYL6 and PHYL8, as well as the additional reconstructions at nodes PHYL2 and PHYL4 using the various algorithms still needs to take place. We have already seen that enzymes reconstructed by different algorithms result in different activities. These various reconstruction methods may provide information on the order in which mutations may occur to yield functional enzymes. From an evolutionary perspective, these reconstructions may also predict different evolutionary paths taken by the α/β hydrolases.

Our initial goal with this research was to determine whether or not ancestral enzymes can serve as a better starting point for directed evolution than modern day enzymes. It is possible that if ancestral enzymes are promiscuous because of enhanced flexibility, they will be more amenable to substitutions. Crystal structures of PHYL2 and other ancestral enzymes will provide information on structural characteristics of these enzymes.

Future plans for research include the directed evolution of PHYL2 and *Hb*HNL for improvement with the nitroaldol reaction. It will be interesting to see if selecting for improvement of the nitroaldol activity will increase or decrease flexibility and whether there is a trade-off between enhanced activity with one reaction and promiscuity.[61,104]

References

1. Beck, E. C. The Love Canal Tragedy. *EPA J.* (1979).

2. Lewis, C. a & Wolfenden, R. Uroporphyrinogen decarboxylation as a benchmark for the catalytic proficiency of enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 17328–17333 (2008).

3. Michaelis, L. & Menten, M. L. Die Kinetik der Invertinwirkung (The Kinetics of Invertase Action). (1913).

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

5. Whitford, P. C., Onuchic, J. N. & Wolynes, P. G. Energy landscape along an enzymatic reaction trajectory: hinges or cracks? *Hum. Front. Sci. Progr. J.* **2,** 61–64 (2008).

6. Wagner, I. New naturally occurring amino acids. *Angew. Chemie Int. Ed.* **22,** 816–828 (1983).

7. Beck, D., Alonso, D. O. V, Inoyama, D. & Daggett, V. The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 12259–12264 (2008).

8. Nardini, M. & Dijkstra, B. W. α/β Hydrolase fold enzymes : the family keeps growing. *Curr. Opin. Struct. Biol.* **9,** 732–737 (1999).

9. Cooper, J., Walshaw, J. & Mills, A. Tertiary protein structure and folds. *Princ. Protein Struct. , Comp. Protein Model. , Vis.* (1999). at <http://swissmodel.expasy.org/course/>

10. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29,** (1990).

11. Nicholls, A., Sharp, K. A. & Hönig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11,** 281–296 (1991).

12. Spolar, R. S., Ha, J. H. & Record, M. T. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 8382–8385 (1989).

13. Gummadi, S. N. What is the role of thermodynamics on protein stability? *Biotechnol. Bioprocess Eng.* **8,** 9–18 (2003).

14. Szila, A., Kardos, J., Osvath, S., Barna, L. & Zavodsky, P. *Neural Protein Metabolism and Function*. *Neural Protein Metab. Funct.* 3–41 (2007).

15. Baldwin, R. L. Energetics of protein folding. *J. Mol. Biol.* **371,** 283–301 (2007).

16. Pratt, L. R. & Pohorille, A. Hydrophobic effects and modeling of biophysical aqueous solution interfaces. *Chem. Rev.* **102,** 2671–2692 (2002).

17. Lukin, J., Kontaxis, G., Simplaceanu, V., Yuan, Y., Bax, A. & Ho, C. Quaternary structure of hemoglobin in solution. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 517–520 (2003).

18. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L. & Schwede, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 1–7 (2014).

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

20. Bjarnholt, N. & Møller, B. L. Hydroxynitrile glucosides. *Phytochemistry* **69,** 1947–1961 (2008).

21. Lauble, H., Miehlich, B., Förster, S., Wajant, H. & Effenberger, F. Crystal structure of hydroxynitrile lyase from *Sorghum bicolor* in complex with the inhibitor benzoic acid: a novel cyanogenic enzyme. *Biochemistry* **41,** 12043–12050 (2002).

22. Bjarnholt, N., Rook, F., Motawia, M. S., Cornett, C., Jørgensen, C., Olsen, C. E., Jaroszewski, J. W., Bak, S. & Møller, B. L. Diversification of an ancient theme: hydroxynitrile glucosides. *Phytochemistry* **69,** 1507–1516 (2008).

23. Tripathi, D., Jiang, Y.-L. & Kumar, D. SABP2, a methyl salicylate esterase is required for the systemic acquired resistance induced by acibenzolar-*S*-methyl in plants. *FEBS Lett.* **584,** 3458–3463 (2010).

24. Kumar, D. & Klessig, D. F. High-affinity salicylic acid-binding protein 2 is required for plant innate immunity and has salicylic acid-stimulated lipase activity. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 16101–16106 (2003).

25. Vlot, a C., Dempsey, D. A. & Klessig, D. F. Salicylic acid, a multifaceted hormone to combat disease. *Annu. Rev. Phytopathol.* **47,** 177–206 (2009).

26. Chen, Z., Zheng, Z., Huang, J., Lai, Z. & Fan, B. Biosynthesis of salicylic acid in plants. *Plant Signal. Behav.* **4,** 493–496 (2009).

27. Nakashita, H., Yoshioka, K., Yasuda, M., Nitta, T., Arai, Y., Yoshida, S. & Yamaguchi, I. Probenazole induces systemic acquired resistance in tobacco through salicylic acid accumulation. *Physiol. Mol. Plant Pathol.* **61,** 197–203 (2002).

28. Gruber, K., Gartler, G., Krammer, B., Schwab, H. & Kratky, C. 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. *J. Biol. Chem.* **279,** 20501–20510 (2004).

29. Gruber-Khadjawi, M., Purkarthofer, T., Skranc, W. & Griengl, H. Hydroxynitrile lyase-catalyzed enzymatic nitroaldol (Henry) reaction. *Adv. Synth. Catal.* **349,** 1445–1450 (2007).

30. Purkarthofer, T, Gruber, K., Gruber-Khadjawi, M., Waich, K., Skranc, W., Mink, D., Griengl, H. A biocatalytic Henry reaction--the hydroxynitrile lyase from *Hevea brasiliensis* also catalyzes nitroaldol reactions. *Angew. Chem. Int. Ed. Engl.* **45,** 3454–3456 (2006).

31. Andexer, J. N., Staunig, N., Eggert, T., Kratky, C., Pohl, M. & Gruber, K.. Hydroxynitrile lyases with α/β hydrolase fold: two enzymes with almost identical 3D structures but opposite enantioselectivities and different reaction mechanisms. *Chembiochem* **13,** 1932–1939 (2012).

32. Padhi, S. K., Fujii, R., Legatt, G. A., Fossum, S. L., Berchtold, R. & Kazlauskas, R. J. Switching from and esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions. *Chem. Biol.* **17,** 863–871 (2010).

33. Steiner, K. & Schwab, H. Recent advances in rational approaches for enzyme engineering. *Comput. Struct. Biotechnol. J.* **2,** e201209010 (2012).

34. Andexer, J., von Langermann, J., Mell, A., Bocola, M., Kragl, U., Eggert, T. & Pohl, M. An R-selective hydroxynitrile lyase from *Arabidopsis thaliana* with an α/β hydrolase fold. *Angew. Chem. Int. Ed. Engl.* **46,** 8679–8681 (2007).

35. Guterl, J.-K., Sehl, T., von Langermann, J., Frindi-Wosch, I., Rosenkranz, T., Fitter, J., Gruber, K., Kragl, U., Eggert, T. & Pohl, M. Uneven twins: comparison of two enantiocomplementary hydroxynitrile lyases with α/β hydrolase fold. *J. Biotechnol.* **141,** 166–173 (2009).

36. Gartler, G., Kratky, C. & Gruber, K. Structural determinants of the enantioselectivity of the hydroxynitrile lyase from *Hevea brasiliensis*. *J. Biotechnol.* **129,** 87–97 (2007).

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

38. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30,** 409–425 (1976).

39. Ohno, S. *Evolution by Gene Duplication*. (Springer-Verlag, 1970).

40. Bergthorsson, U., Andersson, D. I. & Roth, J. R. Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 17004–17009 (2007).

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

42. Risso, V., Gavira, J., Mejia-Carmona, D. F., Gaucher, E. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J. Am. Chem. Soc.* **135,** 2899–2902 (2013).

43. Fuhshuku, K.-I. & Asano, Y. Synthesis of (R)-β-nitro alcohols catalyzed by R-selective hydroxynitrile lyase from Arabidopsis thaliana in the aqueous-organic biphasic system. *J. Biotechnol.* **153,** 153–159 (2011).

44. Ollis, D. L. *et al.* The α/β hydrolase fold. *Protein Eng.* **5,** 197–211 (1992).

45. Nedrud, D., Lin, H., Lopez, G., Padhi, S., Legatt, G. & Kazlauskas, R. Uncovering divergent evolution of α/β hydrolases: a surprising residue

substitution needed to convert *Hevea brasiliensis* hydroxynitrile lyase into an esterase. *Chem. Sci.* (2014) **DOI:** 10.1039/C4SC01544D.

46. Rehm, S., Trodler, P. & Pleiss, J. Solvent-induced lid opening in lipases: a molecular dynamics study. *Protein Sci.* **19,** 2122–30 (2010).

47. Langermann, J. von, Nedrud, D. M. & Kazlauskas, R. J. Improving the catalytic efficiency of hydroxynitrile lyase from Hevea brasiliensis toward aromatic hydroxynitriles by copying active site residues from an esterase that accepts aromatic esters. *ChemBioChem.* **15,** 1931-38 (2014).

48. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno, B. Large shifts in pK a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci.* **108,** 5260–5265 (2011).

49. Harris, T. K. & Turner, G. J. Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life* **53,** 85–98 (2002).

50. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Zanghellini, A., Gallaher, J. L., Betker, J. L., Barbas, C. F., Hilvert, D., Houk, K. N. & Stoddard, B. L.. De novo computational design of retro-aldol enzymes. *Science.* **319,** 1387–1391 (2008).

51. Giger, L. Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N. & Hilvert, D. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9,** 494–498 (2013).

52. Will, K. Principles of Phylogenetics. 1–4 (2012).

53. Huson, D. in *Algorithms Bioinforma. I* 175–182 (2007).

54. Rosset, S. Efficient inference on known phylogenetic trees using Poisson regression. *Bioinformatics* **23,** e142–e147 (2007).

55. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19,** 596–604 (2009).

56. Selmar, D., Lieberei, R., Biehl, B. & Conn, E. α-Hydroxynitrile lyase in *Hevea brasiliensis* and its significance for rapid cyanogenesis. *Physiol. Plant.* **75,** 97–101 (1989).

57. Yan, G., Cheng, S., Zhao, G., Wu, S. & Liu, Y. A single residual replacement improves the folding and stability of recombinant cassava hydroxynitrile lyase in E. coli. *Biotechnol. Lett.* **25,** 1041–1047 (2013).

58. Purkarthofer, T., Skranc, W., Schuster, C. & Griengl, H. Potential and capabilities of hydroxynitrile lyases as biocatalysts in the chemical industry. *Appl. Microbiol. Biotechnol.* **76,** 309–320 (2007).

59. Tattersall, D. B., Bak, S., Jones, P. R., Olsen, C. E., Nielsen, J. K., Hansen, M. L., Høj, P. B., Møller, B. L. Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science.* **293,** 1826–8 (2001).

60. Copley, S. D. Moonlighting is mainstream: paradigm adjustment required. *Bioessays* **34,** 578–588 (2012).

61. Hiblot, J., Gotthard, G., Elias, M. & Chabriere, E. Differential active site loop conformations mediate promiscuous activities in the lactonase SsoPox. *PLoS One* **8,** e75272 (2013).

62. Loening, A. Site Directed Mutagenesis Protocol. *Andreas Loening's Web Page* 1–2 (2005).

63. Zheng, L., Baumann, U. & Reymond, J.-L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32,** e115 (2004).

64. Bernhardt, P., Hult, K. & Kazlauskas, R. J. Molecular basis of perhydrolase activity in serine hydrolases. *Angew. Chem. Int. Ed. Engl.* **44,** 2742–6 (2005).

65. Osborn III, D. J. Asymmetric Henry and Aza-Henry Reactions Drugs and Sugars. in (2004). at <https://www2.chemistry.msu.edu/courses/CEM958/OrganicSeminars.asp>

66. Borah, J. C., Gogoi, S., Boruwa, J., Kalita, B. & Barua, N. C. A highly efficient synthesis of the C-13 side-chain of taxol using Shibasaki's asymmetric Henry reaction. *Tetrahedron Lett.* **45,** 3689–3691 (2004).

67. Luzzio, F. A. The Henry reaction : recent examples. *Tetrahedron* **57,** 915–945 (2001).

68. Otero, J. M., Estévez, J. C., Sussman, F., Villaverde, M. C. & Estévez, R. J. Total synthesis of (5*S*,6*S*)-6-amino-2,8-dimethylnonan-5-ol and (5*S*,6*S*)-6-amino-7-cyclohexyl-2-methylheptan-5-ol. *ARKIVOC* **2007,** 380–388 (2007).

69. Alvarez-Casao, Y., Marques-Lopez, E. & Herrera, R. P. Organocatalytic enantioselective Henry reactions. *Symmetry (Basel).* **3,** 220–245 (2011).

70. Xu, F., Wang, J., Liu, B., Wu, Q. & Lin, X. Enzymatic synthesis of optical pure β-nitroalcohols by combining d-aminoacylase-catalyzed nitroaldol reaction and immobilized lipase PS-catalyzed kinetic resolution. *Green Chem.* **13,** 2359 (2011).

71. Ekroos, M. & Sjögren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 13682–7 (2006).

72. Chakraborty, S. & Rao, B. J. A measure of the promiscuity of proteins and characteristics of residues in the vicinity of the catalytic site that regulate promiscuity. *PLoS One* **7,** e32011 (2012).

73. Hwang, H., Vreven, T., Whitfield, T. W., Wiehe, K. & Weng, Z. A machine learning approach for the prediction of protein surface loop flexibility. *Proteins* **79,** 2467–74 (2011).

74. Li, S. C., Goto, N. K., Williams, K. a & Deber, C. M. α-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 6676–81 (1996).

75. Cortés, J., Siméon, T., Remaud-Siméon, M. & Tran, V. Geometric algorithms for the conformational analysis of long protein loops. *J. Comput. Chem.* **25,** 956–67 (2004).

76. Gu, J., Gribskov, M. & Bourne, P. E. Wiggle-predicting functionally flexible regions from primary sequence. *PLoS Comput. Biol.* **2,** e90 (2006).

77. Ramanathan, A. & Agarwal, P. K. Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. *PLoS Biol.* **9,** e1001193 (2011).

78. Agarwal, P. K. Role of protein dynamics in reaction rate enhancement by enzymes. *J. Am. Chem. Soc.* **127,** 15248–56 (2005).

79. Münz, M., Hein, J. & Biggin, P. C. The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. *PLoS Comput. Biol.* **8,** e1002749 (2012).

80. Schrag, J. D., Li, Y., Cygler, M., Lang, D., Burgdorf, T., Hecht, H. J., Schmid, R., Schomburg, D., Rydel, T. J., Oliver, J. D., Strickland, L. C.,

Dunaway, C. M., Larson, S. B., Day, J. & McPherson, A. The open conformation of a *Pseudomonas* lipase. *Structure* **5,** 187–202 (1997).

81. Skopalik, J., Anzenbacher, P. & Otyepka, M. Flexibility of human cytochromes P450 : molecular dynamics reveals differences between CYPs 3A4 , 2C9 , and 2A6 , which correlate with their substrate preferences. *J. Phys. Chem. B* **112,** 8165–8173 (2008).

82. Nair, N. U., Denard, C. A. & Zhao, H. Engineering of enzymes for selective catalysis. *Curr. Org. Chem.* 1870–1882 (2010).

83. Bornscheuer, U. & Kazlauskas, R. J. Unit 26.7 Survey of protein engineering strategies. *Curr. Protoc. protein Sci.* **26,** 26.7.1–26.7.14 (2011).

84. Sandström, A. G., Wikmark, Y., Engström, K., Nyhlén, J. & Bäckvall, J.-E. Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 78–83 (2012).

85. Grandori, R., Struck, K., Giovanielli, K. & Carey, J. A three-step PCR protocol for construction of chimeric proteins. *Protein Eng.* **10,** 1099–100 (1997).

86. Diez, M. D. A., Ebrecht, A. C., Martínez, L. I., Aleanzi, M. C., Guerrero, S. A., Ballícora, M. A. & Iglesias, A. A. A Chimeric UDP-glucose pyrophosphorylase produced by protein engineering exhibits sensitivity to allosteric regulators. *Int. J. Mol. Sci.* **14,** 9703–21 (2013).

87. Agarwal, P. K., Geist, A. & Gorin, A. Protein Dynamics and Enzymatic Catalysis : Investigating the peptidyl -prolyl cis-trans isomerization activity of cyclophilin A. *Biochemistry* **43,** 10605–10618 (2004).

88. Styczynski, M. P., Fischer, C. R. & Stephanopoulos, G. N. The intelligent design of evolution. *Mol. Syst. Biol.* **2,** 2006.0020 (2006).

89. Zaccolo, M. & Gherardi, E. The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 β-lactamase. *J. Mol. Biol.* **285,** 775–83 (1999).

90. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. Designed divergent evolution of enzyme function. *Nature* **440,** 1078–82 (2006).

91. Yikmis, M., Arenskötter, M., Rose, K., Lange, N., Wernsmann, H., Wiefel, L. & Steinbüchel, A. Secretion and transcriptional regulation of the latex-clearing protein, Lcp, by the rubber-degrading bacterium *Streptomyces* sp. strain K30. *Appl. Environ. Microbiol.* **74,** 5373–82 (2008).

92. Leiros, H.-K. S., Kozielski-Stuhrmann, S., Kapp, U., Terradot, L., Leonard, G. & McSweeney, S. M. Structural basis of 5-nitroimidazole antibiotic resistance: the crystal structure of NimA from *Deinococcus radiodurans*. *J. Biol. Chem.* **279,** 55840–9 (2004).

93. Ohba, K., Nakayama, H., Furihata, K., Shimazu, A., Endo, T. & Seto, H. Nitropeptin, a new dipeptide antibiotic possessing a nitro group. *J. Antibiot.* **XL,** 709–713 (1986).

94. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 Suppl 1,** S162–73 (2009).

95. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.* **37,** D387–92 (2009).

96. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).

97. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

98. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **49,** 377–89 (2009).

99. Reich, H. J. The Aldol Reaction and Condensation. (2013). at <http://www.chem.wisc.edu/areas/reich/chem547/1-carbonyl{12}.htm>

100. Greenberg, W., Varvak, A., Hanson, S. R., Wong, K., Huang, H., Chen, P. & Burk, M. J. Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 5788–5793 (2004).

101. Giger, L., Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N. & Hilvert, D. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9,** 494–8 (2013).

102. Coetze, J. F. & Chang, T.-H. Recommended methods for the purification of solvents and tests for impurities: Nitromethane. *Int. Union Pure Applied Chem.* **58,** 1541–1545 (1986).

103. Reich, H. J. Bordwell pKa Table (Acidity in DMSO). at <http://www.chem.wisc.edu/areas/reich/pkatable/index.htm>

104. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4,** e1000002 (2008).

105. Parril, A. Amino Acid Structures. at <http://www.cem.msu.edu/~cem252/sp97/ch24/ch24aa.html>

106. Zandvoort, E., Baas, B.-J., Quax, W. J. & Poelarends, G. J. Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: enamine formation and aldolase activity. *Chembiochem* **12,** 602–9 (2011).

107. Muller, M. Energy, ATP, and Enzymes. (2004). at <http://www.uic.edu/classes/bios/bios100/lecturesf04am/lect04.htm>

108. Mahajan, D. & Chimni, S. S. Pyrrolidine catalyzed regioselective and diastereoselective direct aldol reaction in water. *Indian J. Chem.* **46,** 1355–1358 (2007).