# TAILORING ENZYME CATALYSTS BY DIRECTED EVOLUTION

Thesis by

Jeffrey C. Moore

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Submitted May 28, 1996)

**Acknowledgment**

I would like to acknowledge my Dads, my sister and my wife for their supportive discussions, and the memory of Mom when those discussions weren't enough.

## Abstract

Directed evolution provides a nature-inspired methodology for the improvement of macromolecular properties. By imitating processes attributed to natural evolution, we have been able to improve an enzyme's activity against a desired substrate. Four rounds of random mutagenesis and screening have led to a p-nitrobenzyl (pNB) esterase with 16-fold improvement in catalytic turnover, 25-fold improvement in catalytic efficiency, and 30-fold improvement in total activity. An extensive library was screened during the fourth generation of random mutagenesis and screening which generated several clones more active than the third generation parent. These clones were recombined in a fifth generation using two techniques similar to that of recombination. A partial sampling of the resulting population contained individuals demonstrating activity two to four-fold higher than the most active variant from the fourth generation, making the total activity improvements greater than 100-fold. The mutations found to increase pNB esterase's activity in this directed evolution experiment were mapped onto a structural model. None of the effective amino acid substitutions lie in segments of the enzyme predicted to interact directly with the bound substrate. To predict in advance that these substitutions enhance pNB esterase activity would be practically impossible.

Sexual recombination provides a mechanism of information-sharing in nature, and an approximation of this method is available as DNA shuffling. The approximation was examined and compared to the natural method in terms of screening optimization. The natural method of pairwise sexual recombination appears optimal when additivity of mutational effects is assumed. In the absence of this assumption, the DNA shuffling procedure can be used to generate all the combinations of mutations present in a parent pool at the expense of requiring the screening of large numbers of clones. A compromise strategy is suggested which removes much of the additivity assumption and requires screening of substantially reduced numbers of clones.

The correlation between various enzyme properties were examined. Enzyme activities in varying concentrations of organic solvent are highly correlated, while no correlation was observed between enzyme activity and enzyme stability in oxidative environments. For the enzyme evolved, variants with increased stability were far more frequent than those with increased activity.

## Table of Contents

# Introduction

Enzymes have the potential to dominate the chemical catalyst area as a direct result of their high catalytic activities and specificities toward targeted substrates. The excitement over this potential dims somewhat with the realization that many important synthetic conversions have no corresponding enzyme catalyst, and the enzymatic catalysts of those that do often are limited by the reaction conditions required as a result of the environment in which these enzymes evolved. For instance, most enzyme catalysts require limited ranges of temperature, pH and ion concentration as a direct result of the environment the organism producing the enzyme is found. Additionally, the complex regulatory mechanisms established for organism survival often create requirements in the enzymatic catalysts unfriendly to use *in vitro.*

Directed evolution arose as a method to address the discrepancy between the properties we would like macromolecules to have *in vitro* and those they have as a result of being evolved *in vivo.* Directed evolution attempts to answer the question, "Can we artificially create an environment specific to our properties requirements and generate the impetus for the creation of macromolecules containing such properties?" Attempting to answer this question brings to light several new questions, "How do we create these environments?", "How do we evolve activity in harsh environments toxic to the organism producing the macromolecule?", "Does placing an organism in an environment of interest correspond to placing an enzymatic catalyst in the same environment?", and "Does a solution satisfying our requirements exist?" Further, we realize from a practical point of view that we do not have the kind of time that nature utilized in generating the diversity of specific, finely-tuned enzymatic catalysts found today.

Looking to nature allowed us to believe these and other questions could be addressed. The wide variety of different environments, including the harsh environments of high temperature, extreme pH's, and high pressures, in which organisms proliferate, suggested that often our concept of a harsh environment is not shared by a large number of microorganisms. The knowledge that the macromolecules within these organisms operate

under some or all of the external environmental pressures led to the belief that solutions satisfying our requirements do exist. We chose an environment of practical interest to synthetic chemistry that we believed to be "harsh" in an attempt to answer the questions asked above, and in addition, the many new questions that arose. This thesis, through the demonstration of our ability to improve an enzymatic catalyst's performance in non-native environments and on synthetic substrates, addresses these issues. In this way, it expands our awareness of what is possible in the tailoring of properties of enzyme catalysts of interest to the scientific and industrial communities and is at least a beginning to understanding the optimization of enzymatic catalysis.

Examining the mechanisms of natural evolution has also provided some guidance for performing directed evolution. In an effort to speed up the time scale of evolution from that found in nature, genetic diversity for the gene of interest was created *in vitro*. As discussed in more detail in the successive chapters, the methods used to create diversity parallel the sexual and asexual reproductive methods employed in nature. The ability of natural selection to highly attune an organism to its environment gave insight into the care required in the design of screening methods needed to sort through the created diversity.

This work also raises several interesting new issues. Directed evolution can clearly be used to improve enzymatic turnover rates, binding affinities, stabilities, differential activities between related substrates and many more properties. The idea of improvement implies that an initial activity on a substrate of interest exists. Often the initial enzyme is found by screening the natural diversity. Natural selection has clearly developed multitudes of new activities over the course of evolution, and current directed evolution methodology is unlikely to generate dramatically novel activities. Is the creation of new activities beyond the reach of directed evolution? What mechanism(s) does natural evolution use to create these new activities, and could similar methodology be employed in directed evolution experiments? Are there technical limitations that would prevent our use of these methods? At what point do activity requirements on two similar substrates create the need for two

different enzymes? These questions, and many like them, provide interesting opportunities to expand the impact of directed evolution as we look toward the future.

# Chapter 2

# Random Mutagenesis and Screening of pNB Esterase

## Introduction

The discovery of bactericidal activity in cultures of bread mold launched the science of antibiotics and with it a multi-billion dollar industry. Over the course of several years, industry learned how to make these compounds through totally synthetic routes, although in most instances microbial fermentation is still the most cost effective means of production. Often the fermentation products are used as raw materials for synthetic chemistry procedures designed to improve the performance of the antibiotic. These procedures might incorporate chemical groups to increase the potency against certain bacteria, to increase the amount of antibiotic that gets transferred into the body when taken orally, or to improve its stability once it gets there.

During total synthesis or chemical modification of antibiotic, several sites on the antibiotic could be adversely affected by the reagents used to carry out any given reaction step. Organic chemists have developed a series of protecting groups designed to mask the antibiotic sites adversely affected. One commonly used protecting group designed to protect carboxylic acid functionalities in cephalosporin-derived antibiotics is para-nitrobenzyl alcohol (pNB-OH) (U. S. Patent 3,725,359 [1975]) (Brannon *et al.*, 1976). The pNB alcohol is covalently linked to the carboxylic acid through an ester coupling, and this linkage is stable enough to withstand the various reaction conditions required. After chemical synthesis is completed, deprotection is required to the return the cephalosporin-pNB ester to its original and active carboxylic acid form. The chemistry required to deprotect the carboxylic acid uses a catalytic form of zinc in concentrated organic solvent, and on an industrial scale this process generates large amounts solvent and zinc-containing waste material. In 1975, scientists at Eli Lilly & Co. interested in pursuing alternative methods of deprotection under milder conditions, began screening for an esterase capable of performing this deprotection reaction (Brannon *et al.*, 1976).

The enzyme known as para-nitrobenzyl esterase was discovered in 1975 by scientists at Eli Lilly & Co., who screened whole cell preparations of numerous bacterial

and fungal cultures for those possessing catalytic activity toward the hydrolysis of a p-nitrobenzyl protected cephalosporin (Brannon *et al.*, 1976). A *Bacillus subtilis* culture (NRRL B8079) showed the highest catalytic activity toward two cephalosporin-derived pNB-protected substrates of all the cultures tested. Although the reaction yield was high, the partially purified enzyme preparations of "pNB esterase" could not compete with the speed, economy, or the small reaction volumes (due to lack of solubility of substrate in purely aqueous environments) with which the zinc-catalyzed deprotection performed, and the interest in an enzymatic deprotection waned.

The interest is back 15 years later. On an industrial scale the deprotection process generates large amounts of solvent and zinc-containing waste material. Interested in pursuing more cost-effective alternatives, scientists at Eli Lilly & Co. began re-investigating pNB esterase's potential. A chromatographically pure solution of pNB esterase was isolated and its amino acid sequence partially determined. Using this partial sequence, DNA primers were constructed and used to isolate the gene responsible for producing pNB esterase and clone it into *E. coli*, where it was over-expressed (Zock *et al.*, 1994). The pNB esterase was then produced in large quantities.

The substrates have changed over the 15-year period as well. Cephalosporin-derived antibiotics continued to evolve from the first generation cephalexin (one of the two original cephalosporin substrates used to screen for pNB esterase), second generation cefaclor, third generation cefixime, and fourth generation loracarbef. These antibiotics have evolved to be readily absorbed (generation one), more potent (generation two), much more potent (generation three), and finally immensely more stable in the body (generation four) (Cooper, 1992). They all are synthesized using the pNB ester protecting group, (Zock *et al.*, 1994) and in protected form, all are only sparingly soluble in water.

The pNB esterase still suffers from its original problem, a problem common to a large number of enzyme reactions in the performance of synthetic chemistry: the desired substrates are only sparingly soluble in water, and the enzyme's catalytic ability is

drastically reduced by even small quantities of non-aqueous solvents. Directed evolution techniques have fairly recently developed to tackle this type of problem. These techniques rely on being able to screen a large number of slightly different variations of the same enzyme, and then to accumulate those variations over several such screenings. Generating the variations in a random fashion has been labeled random mutagenesis; the whole process is directed evolution. The serine protease subtilisin, for example, has been evolved to be almost five hundred times more active than the naturally occurring enzyme under conditions of 60% dimethylformamide (DMF) (Chen and Arnold, 1991, Chen and Arnold, 1993, and You and Arnold, 1995). This work enhanced subtilisin's ability to degrade peptide products in DMF, a job subtilisin has evolved to do in aqueous environments. Since there is no reason to expect that this *Bacillus* esterase has evolved to perform on the laboratory-generated substrate, there is every reason to hope that this type of directed evolution experiment will generate similar results on the pNB esterase enzyme.

The pNB esterase gene has been subjected to four rounds of sequential random mutagenesis and screening to dramatically increase pNB esterase's catalytic ability toward a particular pNB protected antibiotic. By increasing its specificity toward this pNB containing substrate and by increasing its catalytic ability in mixtures of water and non-aqueous solvent, this enzyme has been evolved into an industrially useful tool for the deprotection of pNB esters in amounts of organic solvents required to solubilize sufficient quantities of non-polar substrates.

**Experimental Design and Strategy**

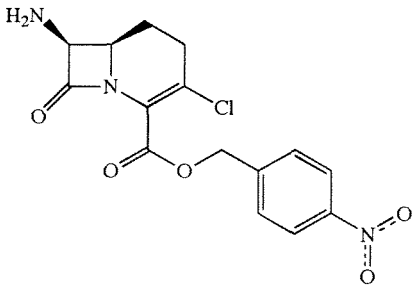
Loracarbef (LCN) is a cephalosporin-derived antibiotic and is marketed in modified form under the trade name Lorabid. The production of loracarbef is different from many traditional antibiotics in that it is synthesized chemically with no microbial fermentation steps. This ensures that the antibiotic is free from any microbially-produced toxins generated during fermentation. The functional antibiotic requires a free carboxylic acid

moiety; the carboxylic acid is protected during the necessary synthetic steps through its coupling to pNB alcohol, forming an ester linkage. The pNB esterase enzyme is expected to catalyze the deprotection of the carboxylic acid toward the end of the chemical synthesis. This reaction is shown in Figure 1a. In addition to protecting the carboxylic acid, the ester-linked pNB group makes the resulting LCN-pNB compound virtually insoluble in water.

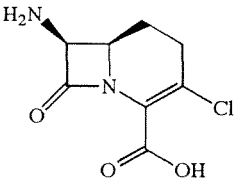
In designing an experiment which attempts to direct the evolution of an enzyme towards activity and specificity on a given substrate, several important parameters require consideration. One such parameter arises from examining the frequency with which enzyme variants with enhanced performance on the desired substrate arise as a result of random mutagenesis. Nature has demonstrated repeatedly that most variations in an enzyme's amino acid sequence either do not alter the enzyme's structure or function or are deleterious. This suggests that a large number of variants need to be examined to find a variant more effective at performing the desired ester hydrolysis in Figure 1a. A rapid procedure is required to screen large numbers of enzyme variants, and visual assays are most often optimal in this regard. The reaction in Figure 1a is problematic for rapid screening of activity because the absorbance spectra of the reactant and the two products are very similar. In addition, the reactant and products do not absorb in the visible region, making the rapid assaying of activity difficult.

Para-nitrophenyl acetate (pNPA) is a general esterase substrate; the enzyme catalyzed reaction is shown in Figure 1b. The pNPA substrate solves the absorbance problem, as the nitrophenol product is yellow while the other reaction components are colorless. The ability of the alcohol oxygen to form resonance structures which participate in conjugation with the phenyl ring gives rise to the yellow color. Lowering the pH below 6.5 severely shifts the resonance structure away from the conjugation and eliminates the yellow color associated with nitrophenol solutions. The ability to form resonance structures also makes nitrophenol an excellent leaving group, as demonstrated by pNPA's gradual hydrolysis in buffer alone. This spontaneous hydrolysis accelerates with

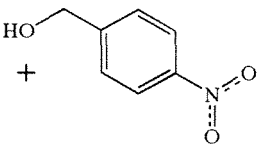
a)



p-Nitrobenzyl loracarbef nucleus (LCN-pNB)

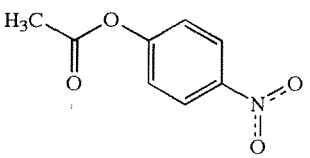


Loracarbef nucleus (LCN)

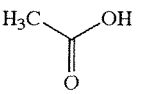


p-Nitrobenzyl alcohol (pNB)

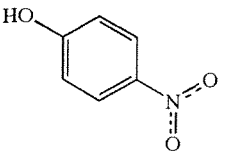
b)



p-Nitrophenyl acetate (pNPA)

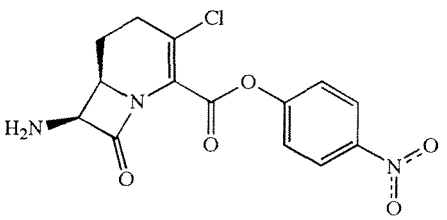


Acetate

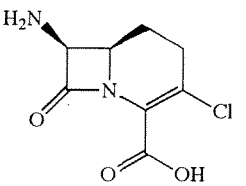


p-Nitrophenol (pNP)

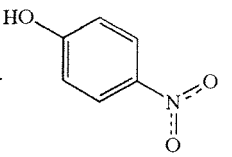
c)



p-Nitrophenyl loracarbef nucleus (LCN-pNP)



Loracarbef nucleus (LCN)



p-Nitrophenol (pNP)

Figure 2.1: Substrates and products of reactions catalyzed by pNB esterase. Figure 2.1a shows the reaction which motivates this study. Figure 2.1b and 2.1c show substrates and reaction products of the two p-nitrophenol substrates used for screening.

increasing pH, and at pH's above 8.5 occurs almost immediately. The pNPA substrate is also membrane permeable. This substrate was hydrolyzed rapidly by whole *E. coli* cells expressing pNB esterase intracellularly. The same cells, but without the expression plasmid, do not catalyze the conversion. This substrate is sterically and chemically very different from the LCN-pNB substrate, and as such does not make an ideal choice for directing the evolution of the esterase toward activity on LCN-pNB. It does, however, allow for optimization of enzyme expression in new bacterial hosts, where the increase in amount of enzyme produced translates into increased activity during screening.

To generate an enzyme with improved activity toward the LCN-pNB substrate, a third substrate was designed to include the chromophore from the pNPA substrate and as much of the loracarbef nucleus as possible. The result was the "hybrid" substrate whose structure and reaction are shown in Figure 1c. This substrate, like pNPA, is membrane permeable, making cell lysis unnecessary.

A second important parameter in directed evolution experiments is the choice of screening conditions. The more the screening conditions differ from the desired reaction conditions, the more likely that variants found to have a positive effect in screening will have no effect in the desired reaction conditions. For example, the native pNB esterase has its pH optimum at pH 8.3, while the pNP leaving group requires that screening be performed at lower pH for chemical stability. This raises the issue of whether the 'evolved' enzyme variants will truly be more active on any of the substrates, or whether a simple shift in the pH optimum will generate the appearance of enhanced activity under the screening conditions. The best enzyme will be used under its optimal conditions for performing the desired catalysis; if the 'evolved' variants do not outperform the wild type enzyme in its optimal conditions, the evolved enzyme variants will not be used.

The screen consists of resuspending individual colonies of bacteria in a small volume of buffer and measuring the turbidity of the bacterial suspension using a spectrophotometer in order to estimate the cell concentration in the buffered solution. A

small volume of this bacterial suspension is added to a buffered solution containing a pNP substrate, and the release of product is measured by following the formation of yellow color. The rate of product appearance is normalized to the cell concentration by the turbidity measurement. This is indicative of enzyme activity per bacterium; those colonies which generate higher activity to turbidity ratios are retested. The variant pNB esterases contained within the best clones are then purified and tested on the screening substrate to determine the extent of improvement and on LCN-pNB to determine whether the improvement applies to the substrate of ultimate interest.

Our strategy for directing the evolution of pNB esterase is first to make a large library of pNB esterase genes, each with a small number of random alterations in the 1500 base pair DNA sequence which codes for the pNB esterase. Ideally these alterations are random both in location and type of substitution. This collection of DNA sequences is then placed into *E. coli*, which translates copies of the DNA sequences into amino acid sequences. Because the DNA sequence has been altered slightly, the amino acid sequence of the enzyme may be altered. The pNP-containing substrates are then used to identify those *E. coli* that are producing an enzyme which appear to outperform the original. The best performer of this group then serves as a template for the next round of mutagenesis and screening. This sequence of events is repeated until the desired goal is achieved.

The number of random substitutions introduced in the 1500 base pair sequence (substitution frequency) is a third important design parameter in directed evolution experiments. If the frequency of substitutions is too high, most of the enzymes produced will be inactive. If the frequency is too low, most of the DNA base substitutions produced will be an exact copy of the original DNA sequence, and the resulting enzymes will not be any different than the original. Because approximately one-third of the altered DNA sequences lead to the same amino acid sequence in a protein, the ideal number of DNA base substitutions is greater than one. At one substitution or less per sequence, much of the DNA produced will produce exact copies of the original protein sequence, and a substantial

portion of the screening effort will be spent searching through copies of the original enzyme. At greater than three substitutions per sequence, on average greater than two amino acid substitutions per enzyme are being produced. The enzyme's activity is a function of all the substitutions contained within; the activity becomes a competition between the rare substitutions which are beneficial and the less rare substitutions which are deleterious (Chen and Arnold, 1991). The ideal number of substitutions is therefore somewhere greater than one and not too much larger than three.

The substitution frequency is calculated as the number of substitutions made in a given sequence divided by the number of possible sites for substitution in the sequence and is usually expressed as a percentage. This means that the substitution frequency required to generate one to three substitutions per gene depends on the sequence length of the DNA coding for the enzyme. For instance, if an enzyme of interest is coded for by 900 DNA base pairs (300 amino acids), a higher substitution frequency is required for a DNA sequence 1500 base pairs long (500 amino acids) in order to maintain the one to three substitutions within this region. PCR conditions which generate substitution rates from 0.25 to 20 substitutions per 1000 base pairs have been characterized (Leung *et al.*, 1989, Eckert and Kunkel, 1991, and Cadwell and Joyce, 1992), and the appropriate conditions can be used to accommodate most gene lengths of interest.

This outlines the process by which we can direct the evolution of pNB esterase's ability to better catalyze the desired reaction. Evolution also implies accumulating improvements in activity over several generations, and this process is repeated multiple times, each time beginning with the best variant from the previous generation. A large library of genes each containing a small number of DNA substitutions are generated using error-prone PCR techniques. This library is placed in *E. coli*, where it is translated from DNA to enzyme. The enzyme library is screened for those enzymes which outperform the original. The best new enzyme then becomes the original as the process is repeated until a desired result is achieved.

## Results

*Homology Studies*

A homology search of the major protein data bases (Protein Information Resource, Swiss Protein, translated GenBank, and Protein Data Bank) revealed that pNB esterase shares significant homology with a number of esterases. Eleven of the most homologous enzymes, representing seven distinct classes of esterases, were chosen for sequence comparisons with pNB esterase. These enzymes, their EC classification, the organism from which they were isolated, and their percent identity and similarity to pNB esterase are listed in Table 1. Acetylcholinesterase is the only enzyme of this group to play an important role in the nervous system by hydrolyzing acetylcholine at diffusion-controlled rates (Hasinoff, 1982). While bacteria may have little need for nervous system responses of the higher vertebrates, most of the remaining enzymes are digestive enzymes. Butyrylcholinesterases, carboxylesterases, thioesterases, lipases and cholesterol esterases are important in metabolizing carbon and energy sources (Cygler *et al.*, 1993). The last enzyme, carbamate hydrolase, was discovered in the same way pNB esterase was discovered: screening the environment for an enzyme with activity on a desired substrate, phenmedipham, an herbicide carbamate (Pohlenz, 1992). Carbamates are structurally similar to esters, containing a nitrogen linkage not present in esters (R-**N**-COO-R' vs. R-C-COO-R'), and are known to inhibit esterases. This degradative activity was discovered in an *Arthrobacter oxidans* strain from soil samples of phenmedipham-treated fields. Whether this is the enzyme's natural function is not known. While all of these enzymes demonstrate activities toward esterase substrates, the known natural substrates for these enzymes vary widely. As shown in Figure 2, the lipases and thioesterases hydrolyze esters of medium to long chain (10 to 18 carbon atoms) fatty acids with large leaving groups (like coenzyme A). At the other extreme, acetyl- and butyryl- cholinesterases hydrolyze very short esters made up of short acids (1 or 2 carbon atoms) and small leaving groups (like choline).

| Enzyme | Code | Species | % Id. | % Sim. | Reference |
|---|---|---|---|---|---|
| Acetylcholinesterase | EC 3.1.1.7. | *Torpedo californica* | 32.5 | 53.7 | Sussman, 1991 |
| | | *Oryctolagus cuniculus* | 36.0 | 58.3 | Jbilo, 1994 |
| Butyrylcholinesterase | EC 3.1.1.8. | *Oryctolagus cuniculus* | 35.0 | 56.7 | Jbilo, 1990 |
| Carboxylesterase | EC 3.1.1.1. | *Oryctolagus cuniculus* | 36.7 | 57.2 | Ozol, 1989 |
| | | *Homo sapiens* | 37.2 | 58.6 | Shibata, 1993 |
| | | *Dictyostelium discoideum* | 34.4 | 55.5 | Bomblie, 1990 |
| Thioesterase | EC 3.1.2.14. | *Anas platyrhynchos* | 38.7 | 58.3 | Hwang, 1993 |
| Triacylglycerol lipase | EC 3.1.1.3. | *Geotrichum candidum* | 30.4 | 48.6 | Schrag, 1993 |
| | | *Candida rugosa* | 29.1 | 49.5 | Lotti, 1993 |
| Cholesterol esterase | EC 3.1.1.13. | *Candida rugosa* | 29.6 | 49.8 | Kaiser, 1994 |
| Carbamate hydrolase | EC 3.1.1.-. | *Arthrobacter oxidans* | 34.0 | 56.8 | Pohlenz, 1992 |

Table 2.1: Comparison of amino acid sequence between pNB esterase and esterases isolated from various organisms. The enzymes were identified using a BLAST search of the PDB, PIR, SWISS-PROT, and translated GenBank databases. Percent identity and percent similarity were determined using the BESTFIT tool in the GCG software package.

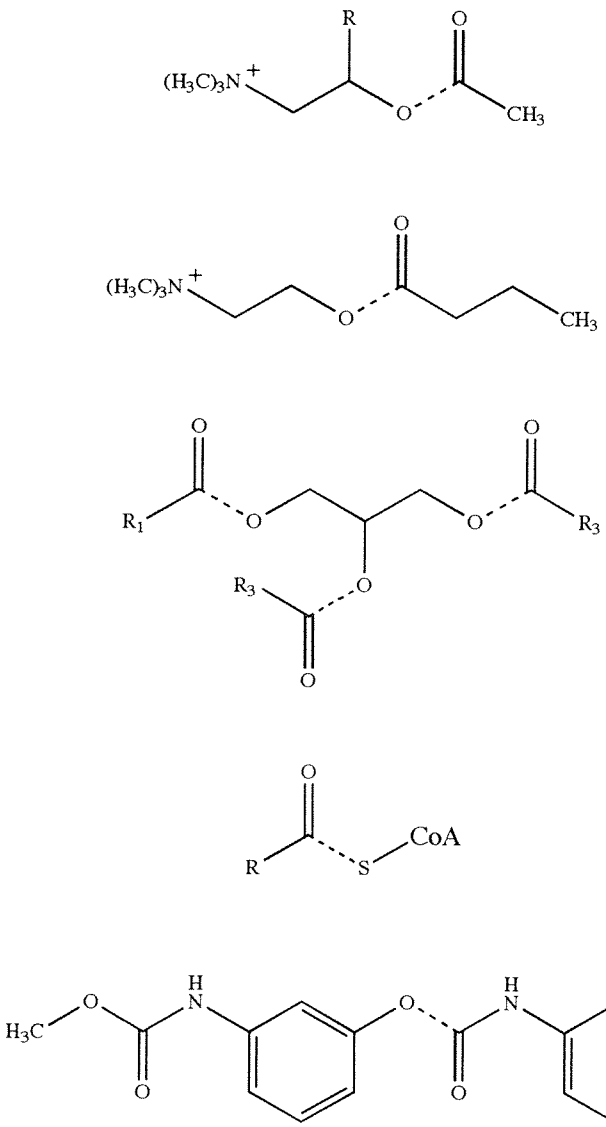



Figure 2.2: Substrates of esterases sharing homology to pNB esterase. The bond cleaved is indicated by a dotted line. The cholinesterase substrates are small in size relative to the triacylglycerol hydrolases (lipases) and the thioesterases. The compound phenmedipham is a man-made substrate of intermediate size.

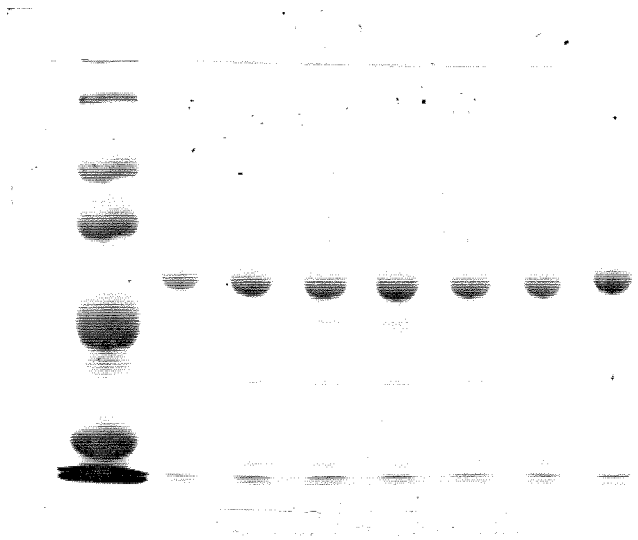
Some of the esterases in this group of homologous enzymes are noted for the feature of substrate inhibition at high substrate concentration (Taylor and Radic, 1994). In particular, substrate inhibition has been a well-noted feature of acetylcholinesterase analysis. While the mechanism of inhibition is not clear, people have chosen to model the inhibition using the premise that the substrate can bind at two locations within the enzyme, and do so with different binding constants (Radic *et al.*, 1993). Butyrylcholinesterase does not share this inhibition, and this fact is often used to distinguish the two cholinesterases. Studies have determined some of the residues responsible for this behavior by altering acetylcholinesterase residues to the appropriate butyrylcholinesterase residues and examining the inhibitory behavior (Harel *et al.*, 1992). This leaves open the question of whether pNB esterase is substrate inhibited, but the fact that the inhibition characteristics can be altered by substitution suggests that enzyme variants can be screened for the absence of inhibition, if desired.

*Purification*

The original purification scheme included a pH precipitation, an ammonium sulfate fractionation and three chromatographic steps (Chen *et al.*, 1995 and Zock *et al.*, 1994). The three chromatographic steps were reduced to two by replacing the dye affinity column and an ion exchange column with a single (IDA - $Cu^{2+}$) metal affinity column (IMAC). The wild type pNB esterase open reading frame contains 12 histidines, which are the amino acid residues generally responsible for retention on a metal affinity column (Arnold, 1991). Although the surface accessibility of these histidines is unknown, the elution of pNB esterase at 4-5 mM imidazole in an imidazole gradient is consistent with one or two histidine interactions with the chromatographic support (Todd *et al.*, 1994).

After the enzyme samples were exchanged into Tris buffer, pH 7.0, the enzyme concentrations were measured. Two SDS-PAGE gels were run to compare purity of the protein samples. The first of those gels is shown in Figure 3a. The first lane contains

a)



b)

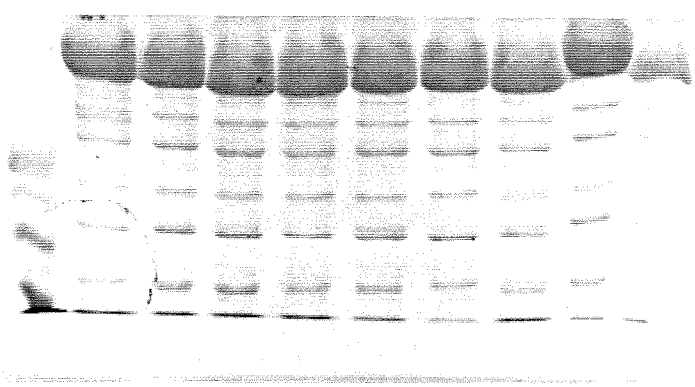


Figure 2.3a: SDS-PAGE gel containing from left to right: molecular weight marker, wt purified using the original protocol, and wt and five variants purified using the new purification scheme. 2.3b demonstrates a similar gel overloaded to visualize contaminants. Again the lanes from left to right are molecular weight marker, wt purified using the original protocol, and wt and seven variants purified using the new purification scheme.

molecular weight markers, the second contains pNB esterase purified using the original method, and the remaining lanes show the several pNB esterase preparations purified using the modified procedure. Figure 3b shows the results of overloading a similar SDS-PAGE gel for the purpose of comparing contaminants. As evidenced by these two gels, the purity of pNB is not compromised by replacing the two chromatographic steps with one IMAC column. From these gels the purity is estimated to be at least 95%. In addition to removing a chromatographic step, this replacement also conveniently removed the need for dialysis between columns, as the high salt content after the first ion exchange column does not affect the performance of the metal affinity column. Only after the metal affinity column was dialysis performed.

*Activities of Related Enzymes*

Six enzymes chosen based on their homology with pNB esterase or their known activity on a broad range of ester substrates were tested for the ability to hydrolyze the loracarbef nucleus-pNB substrate. This was performed because, although an extensive library of bacteria and fungi were screened to find pNB esterase, the search was by no means exhaustive (Brannon *et al.*, 1976). The possibility exists that a related enzyme might also have substantial activity on this substrate and provide information that could be used to accelerate a search for a highly active pNB esterase. This possibility is further enhanced by the fact that many of the enzymes showing the highest homology with pNB esterase are mammalian enzymes, which were not tested in the original screen. The tested enzymes were the highly similar acetyl and butyrylcholinesterases, lipase, and cholesterol esterase, as well as the broadly specific subtilisin A and elastase enzymes. None of the enzymes demonstrated measurable activity on the LCN-pNB substrate.

*Random Mutagenesis of pNB Esterase by Error-Prone PCR*

The pNB esterase gene is flanked by the restriction site Xba I 51 base pairs prior to the start of the open reading frame and by the restriction site Bam HI 313 base pairs after the stop codon of the open reading frame (Zock *et al.* 1994), as shown in Figure 4. Small, single-stranded DNA primers were synthesized to complement regions 25 base pairs upstream of the Xba I site (forward primer) and 143 base pairs downstream of the Bam HI site (reverse primer). The locations of these primers were chosen because the DNA between the two primers is the region that will be altered and amplified during the mutagenic polymerase chain reaction (error-prone PCR). The error-prone PCR conditions used were based on the requirements that the substitution frequency be between one and three substitutions per thousand bases (1.5 to 4.5 substitutions per gene) (Leung *et al.*, 1989). Changes in any part of the open reading frame resulting in enhanced activity are useful, and as a result the whole open reading frame was given the opportunity to be altered by the mutagenesis. Additionally, once the DNA is amplified and mutagenized, it must be inserted into a circular DNA plasmid. By cutting the amplified DNA with the restriction enzymes Xba I and Bam HI, the ends of this insert are properly prepared to ligate to the plasmid. Finally the primers are located far enough outside of the restriction sites that the small pieces of DNA liberated when the insert is cut by Xba I and Bam HI are visible by standard gel electrophoresis techniques. This ensures that the cutting step has occurred properly, should the ligation perform poorly.

*Screening and Analysis of Variants*

An initial round of error prone PCR was performed to produce substitutions within the pNB esterase gene. The resulting DNA product was cloned into the expression vector and expressed in *E. coli.* Of the resulting colonies, 1000 were screened for esterase activity on the pNPA substrate in 20% DMF in 96 well plates. Of the 1000 original colonies, 33 were rescreened as potential positive variants. The three colonies

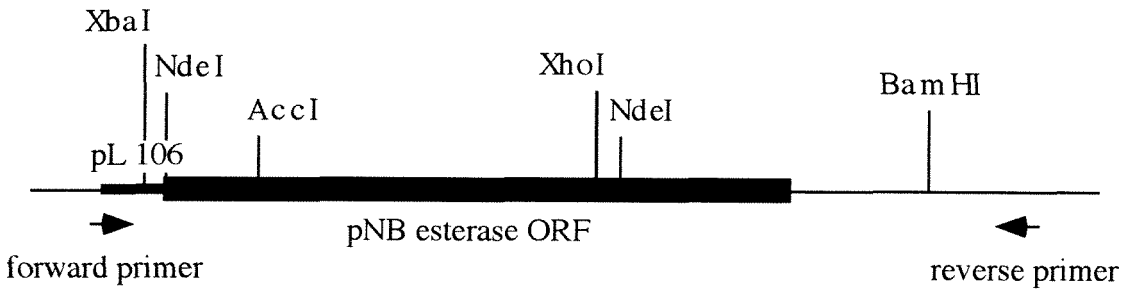



Figure 2.4: Linear map of the pNB esterase gene. The thick line indicates pNB esterase open reading frame, the medium line is the pL106 promoter region, and the thin line indicates the continuation of plasmid DNA. Restriction sites are marked as in the plasmid map. The location of the forward and reverse primers used in the error-prone PCR mutagenesis are indicated by small arrows below the linear map.

with the highest activity to cell density (turbidity) ratio were grown, along with the wild type pNB esterase, in 1 L cultures and the pNB esterase variants partially purified using the precipitation, ammonium sulfate fractionation, and a single DE-52 ion exchange column (Chen *et al.*, 1995). These partially purified enzymes were then assayed along with wild type on the pNPA and LCN-pNB substrates. All showed higher total activity than wild type on the pNPA substrate, and only one, 1-1H9, showed significant total activity increase over wild type on the actual LCN-pNB substrate. (1H9 indicates the variant designation; the initial 1- indicates round or generation number. This should be read "variant 1H9 of generation 1." All variant designations will follow this format.) This variant was therefore used as the parent for the second round of mutagenesis.

The second generation of the evolution process began with a mutagenic PCR reaction on the gene isolated from variant 1-1H9. In 96 well plates, 2800 colonies were screened this time using the LCN-pNP substrate and 15% (v/v) DMF. From these, 65 colonies were rescreened as potential positive variants, and again the best three were grown in 1 L cultures along with the wild type and the 1-1H9 parent. The pNB esterase variants (2-13F3, 2-19E10, and 2-23E1, 1-1H9 and wild type) from these colonies were purified and assayed on all three ester substrates (pNPA, LCN-pNP, and LCN-pNB), with the results shown in Figure 5a, b and c. While the second round variants had lost some of their ability to hydrolyze pNPA, all three exhibited increased activity on LCN-pNP. For two second round variants, 2-13F3 and 2-19E10, the increase in activity also applied to the desired p-nitrobenzyl substrate, LCN-pNB. However, 2-23E1, the variant showing the most activity on LCN-pNP, did not show marked improvement on LCN-pNB. Because 2-19E10 showed slightly better performance characteristics, it was used as the parent for the third round of mutagenesis.

Screening of the third round of mutagenesis involved examination of 1500 colonies using the LCN-pNP substrate and 20% DMF. Forty were rescreened as potential positive variants. The three best (3-7D5, 3-9E10, and 3-10C4) were then grown in 500 mL

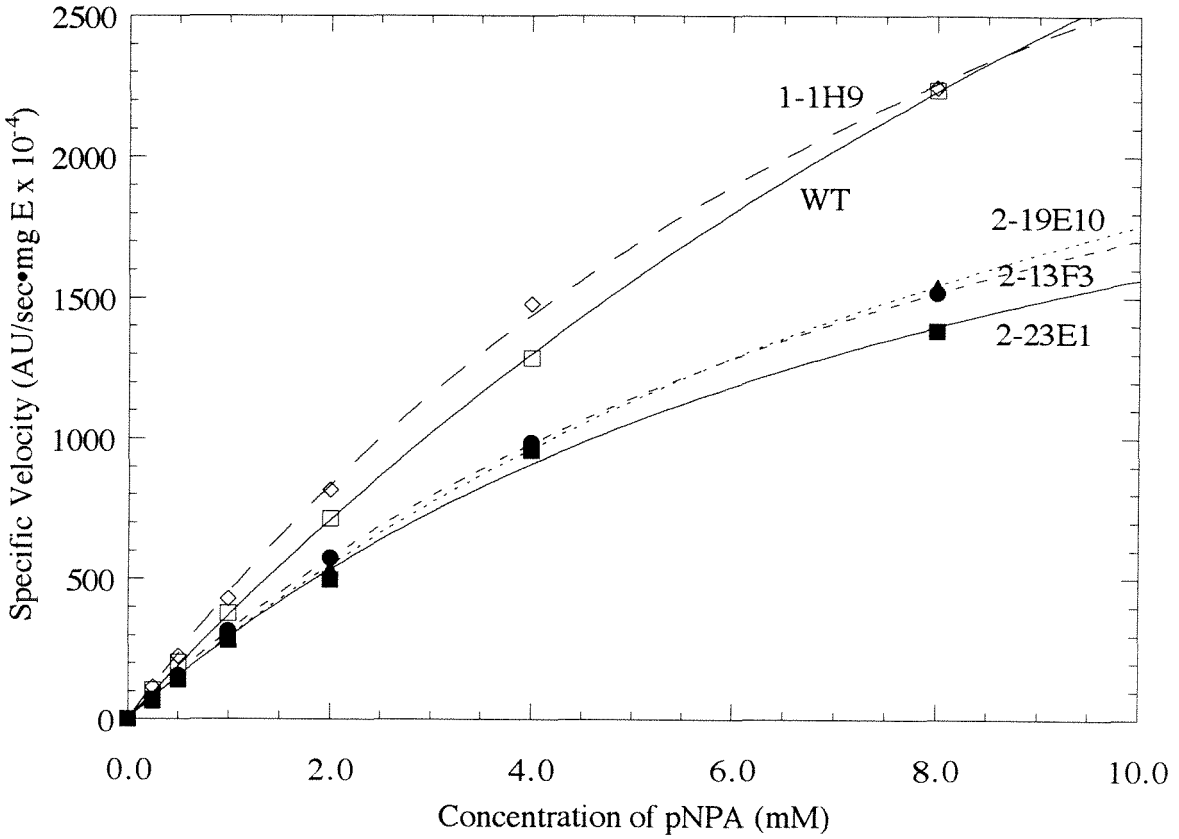



Figure 2.5a: Variant and wild type pNB reaction kinetics on p-nitrophenyl acetate (pNPA) in 15% dimethylformamide (DMF). Enzymes were added to a 30 °C reaction solution consisting of 0.1 M Tris-HCl pH 7.0, 15% DMF, and varying concentrations of pNPA.

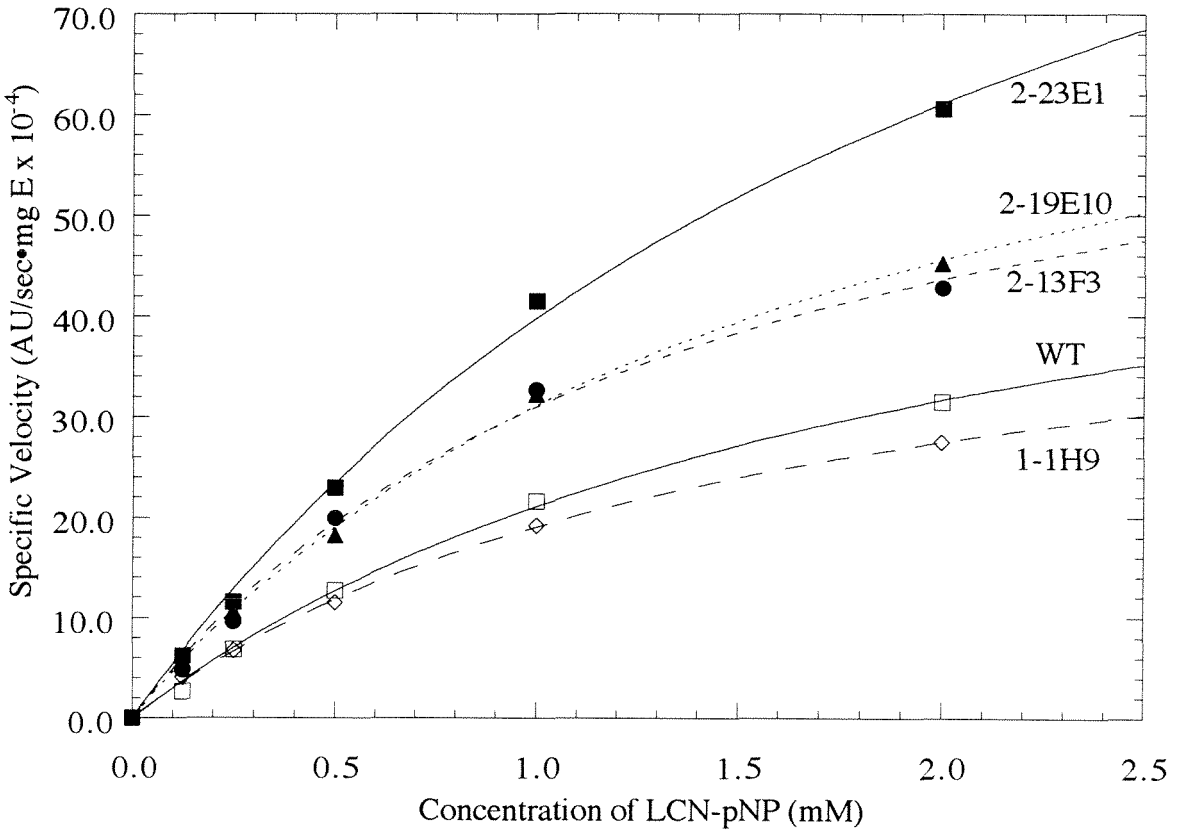



Figure 2.5b: Variant and wild type pNB esterase reaction kinetics on p-nitrophenyl loracarbef nucleus (LCN-pNP) in 15% dimethylformamide (DMF). Enzymes were added to a 30 °C reaction solution consisting of 0.1 M Tris-HCl pH 7.0, 15% DMF, and varying concentrations of LCN-pNP.

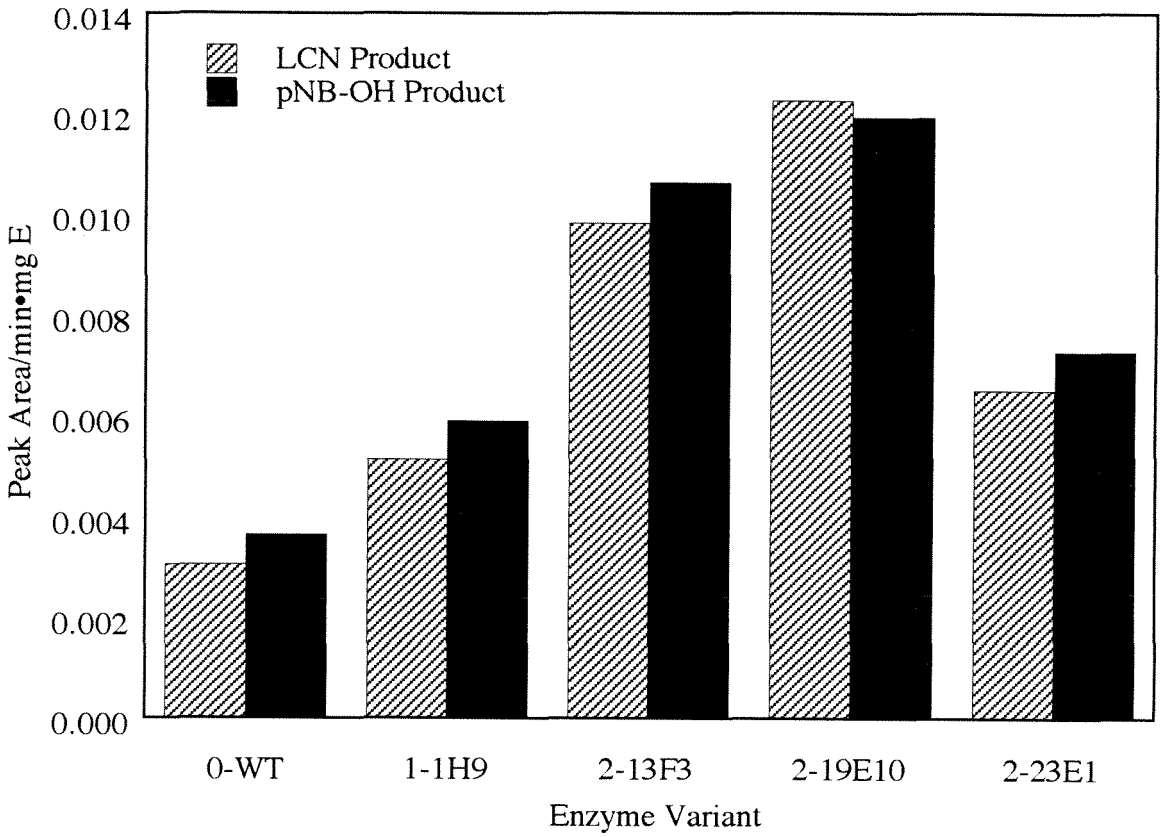



Figure 2.5c: Product formation by variant and wild type pNB esterase reaction rate on 1.0 mM p-nitrobenzyl loracarbef nucleus (LCN-pNB) in 15% dimethylformamide (DMF). Enzyme was added to a 30 °C reaction solution consisting of 0.1 mM Tris-HCl pH 7.0, 15% DMF, and 1.0 mM LCN-pNB.

cultures, and the enzymes were purified. Of these three showing best activity on LCN-pNP, only one, 3-10C4, showed increased activity on the actual substrate, LCN-pNB, as demonstrated in Figure 6. 3-10C4 shows a 40% improvement over 2-19E10 in 2.5% DMF and a 50% improvement in 15% DMF.

The fourth round of mutagenesis and screening examined 7400 colonies using LCN-pNP substrate and 20% DMF. Of these, 250 were rescreened as potential positives. Of the 250, 64 positives consisting of those either most active in 5% DMF, most active in 20% DMF, or the best ratio of activities in 20% to 5% DMF were additionally screened along with wild type, 1-1H9, 2-19E10, and 3-10C4 on the actual substrate, LCN-pNB. The screening results on both LCN-pNP and LCN-pNB were normalized to the activity of the parent 3-10C4 and are shown in Figure 7. Of the 64 colonies chosen, 5 show activity increases of 50% or more over 3-10C4, and 16 show increases of greater than 20% over 3-10C4. The best five variants were determined from Figure 7 based on the ability to hydrolyze the desired substrate LCN-pNB only. The best variant with over a 2-fold improvement on 3-10C4 was 4-54B9. The remaining four variants all demonstrated approximately 60-65% improvement over 3-10C4; these variants were labeled 4-38B9, 4-43E7, 4-53D5 and 4-73B4.

Fifth generation variants were generated by combining the fourth round variants using restriction and ligation techniques. The purified plasmids from the five variants from the fourth generation were each restricted by Xba I, Bam HI, and Xho I and purified, as described in the Materials and Methods. The DNA fragments were mixed with the DNA fragments from 4-54B9, the variant which appeared to outperform all others from the fourth generation, in pairwise fashion (e.g., one tube contained the fragments from 4-38B9 and 4-54B9, another tube contained fragments from 4-43E7 and 4-54B9, etc.). These mixtures of DNA fragments were each ligated simultaneously with the expression plasmid, transformed as before, and assayed on LCN-pNP substrate in 20% DMF. Of the four sets of ligations performed, only one, the mixture ligating 4-54B9 with 4-38B9, resulted in an

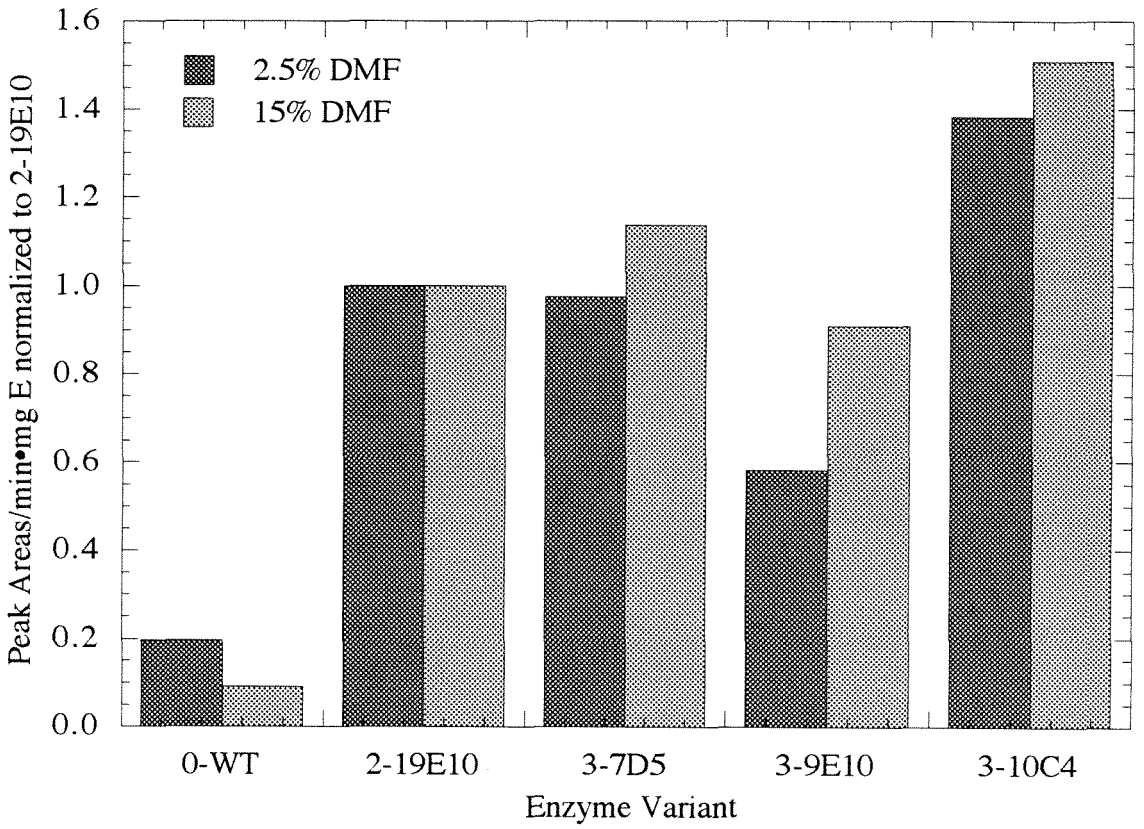



Figure 2.6: Reaction rates of variant and wild type pNB esterases on 0.25 mM p-nitro-benzyl loracarbef nucleus (LCN-pNB) in 2.5 and 15% dimethylformamide (DMF). The disappearance of substrate from a 30 °C solution consisting of 0.1 M Tris-HCl pH 7.0, DMF and 0.25 mM LCN-pNB and one of the enzymes listed was measured by HPLC. The resulting peak areas were normalized to the parent of generation three, 2-19E10.

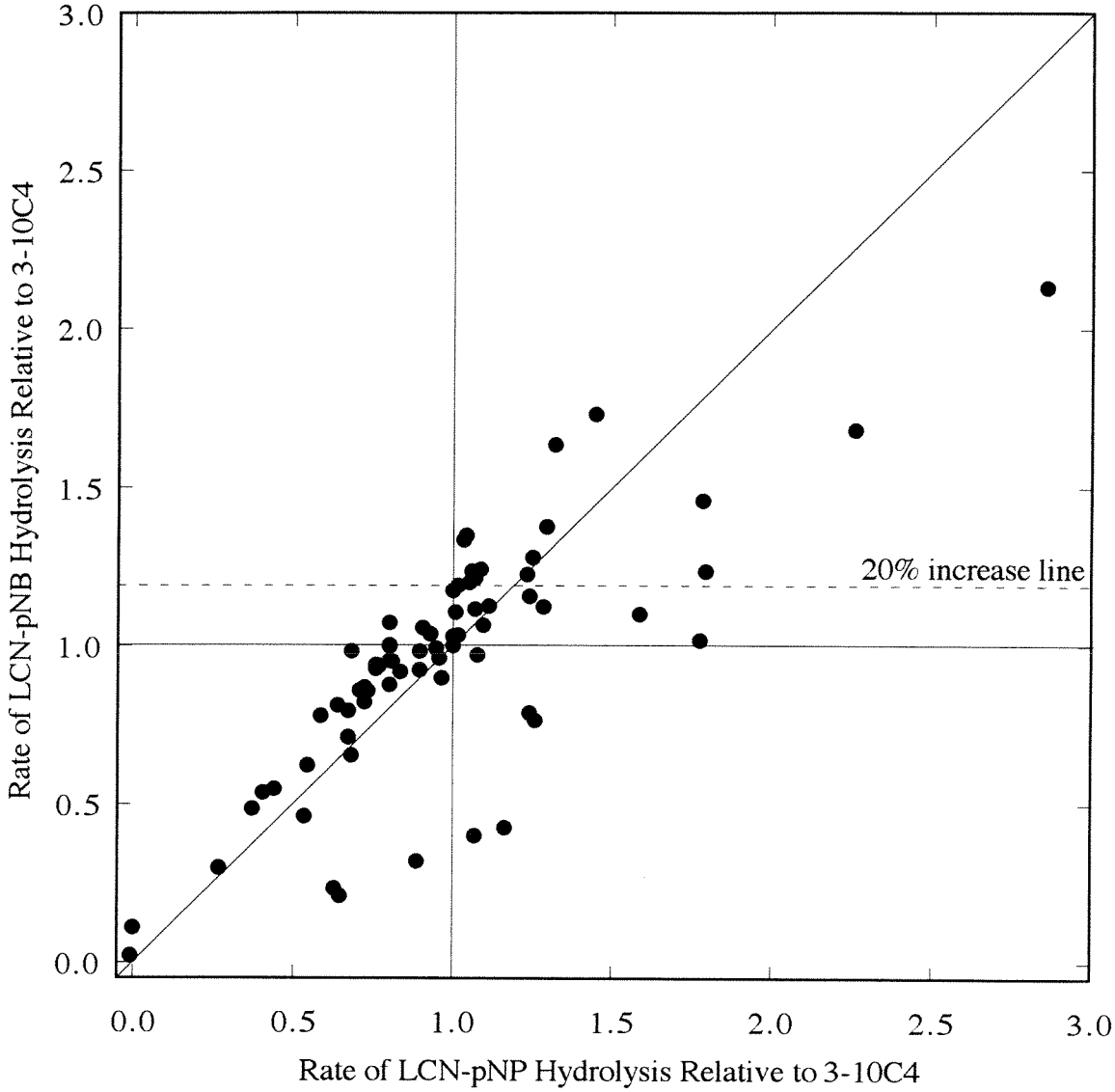



Figure 2.7: Hydrolysis rates of fourth generation potential variants between the screening and actual substrates. The rates are normalized to the third generation variant 3-10C4. After an eight-hour induction period, whole cell screening assays were performed at 25 °C in a 0.1 mM Tris-HCl pH 7.0, 15% dimethylformamide reaction solution containing 0.8 mM of either p-nitrophenyl loracarbef nucleus (LCN-pNP) or p-nitrobenzyl loracarbef nucleus (LCN-pNB).

enhancement in activity over 4-54B9. The results from screening colonies expressing this mixture of recombined genes are shown in Figure 8. According to the screening data, this combination of mutations (labeled 5-1A12) displays approximately twice the activity of 4-54B9. This demonstrates that positive mutations can be combined for additional beneficial effects. Certainly, further beneficial combinations of mutations could be found by combining mutations using random recombination methods or by site-directed mutagenesis, once the DNA sequences were determined. Figure 9 shows the lineage of all the enzyme variants noted during the screening process.

*Kinetic Characterization of Evolved pNB Esterases*

Seven pNB esterases consisting of 0-WT, 1-1H9, 2-19E10, 3-10C4, 4-38B9, 4-43E7, and 4-54B9 were grown in one liter cultures and purified. Figure 10 shows the specific reaction rate as a function of pNPA substrate concentration for this series of evolved variants from the four generations of mutagenesis and screening. Only 1-1H9 was chosen based on its performance on this substrate, and chromatographic purification suggests that the majority of improvement in activity demonstrated by this variant is due to an approximate four-fold increase in amount of enzyme produced. The wild type enzyme outperforms this variant in purely aqueous environments (Figure 10a). The screening was done in the presence of DMF, however, and in 15 and 30% DMF (Figures 10b and 10c respectively), 1-1H9 has higher specific activity towards pNPA than wild type. Similar trends are seen in the remaining variants assayed on this substrate. Wild type is the most active enzyme in the absence of DMF but is only average among the variants in 15% DMF. In 30% DMF, wild type drops still further. Variants 4-38B9 and 4-43E7, presumably by virtue of having been screened in DMF for four rounds of mutagenesis, become the best performers.

Figure 11 shows the results of similar kinetic analyses performed using the LCN-pNP substrate with which three out of the four rounds of screening were carried out. In

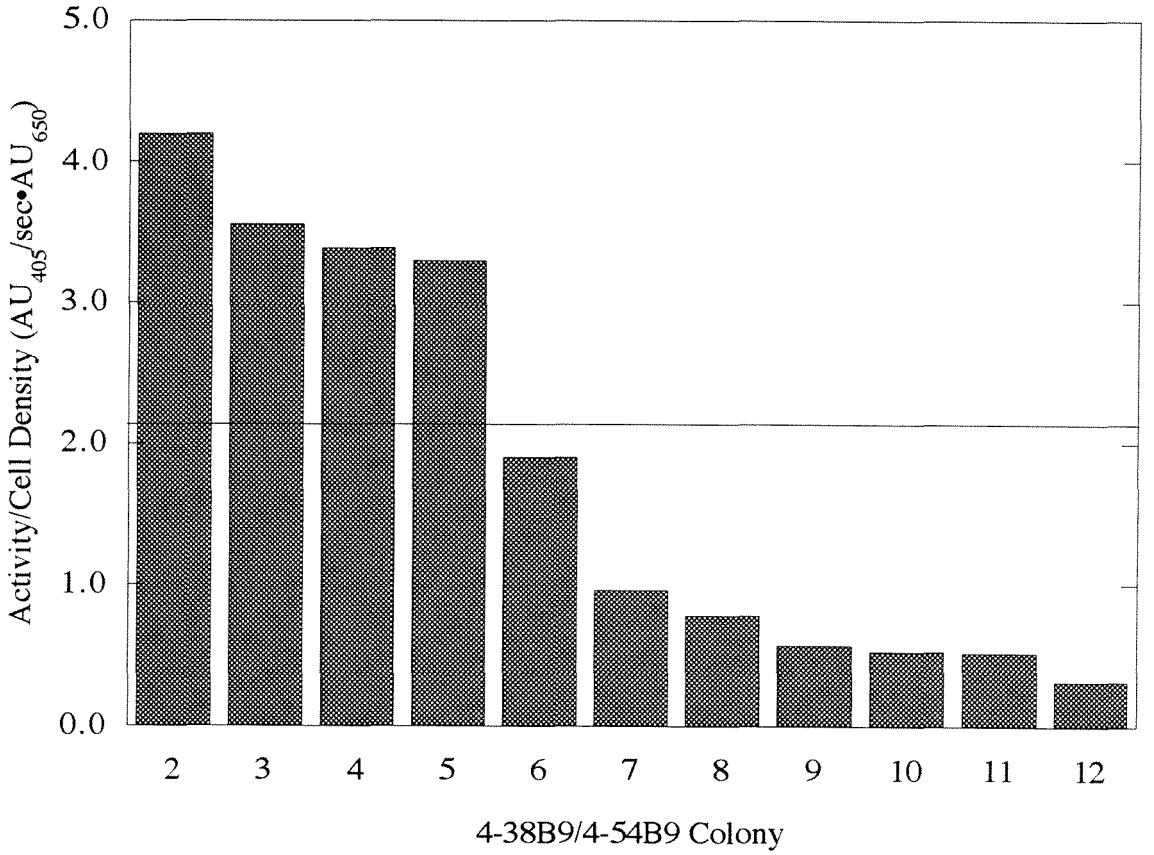



Figure 2.8: Screening activity of ligation mixture 4-38B9 and 4-54B9. The horizontal line indicates the activity of the most active fourth round variant, 4-54B9. Colony 12 was saved for further study.

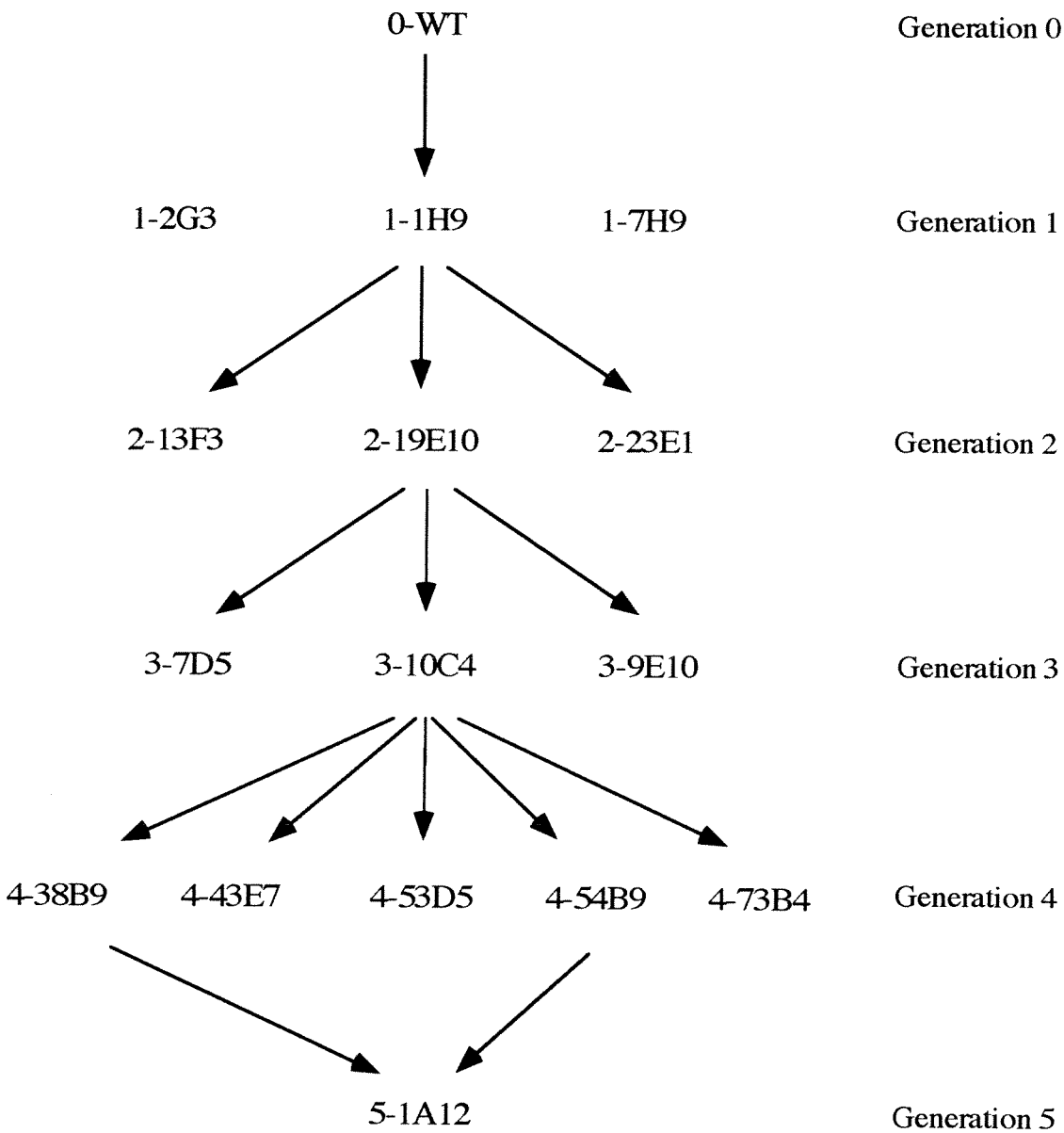



Figure 2.9: Tree showing lineage of pNB esterase variants.

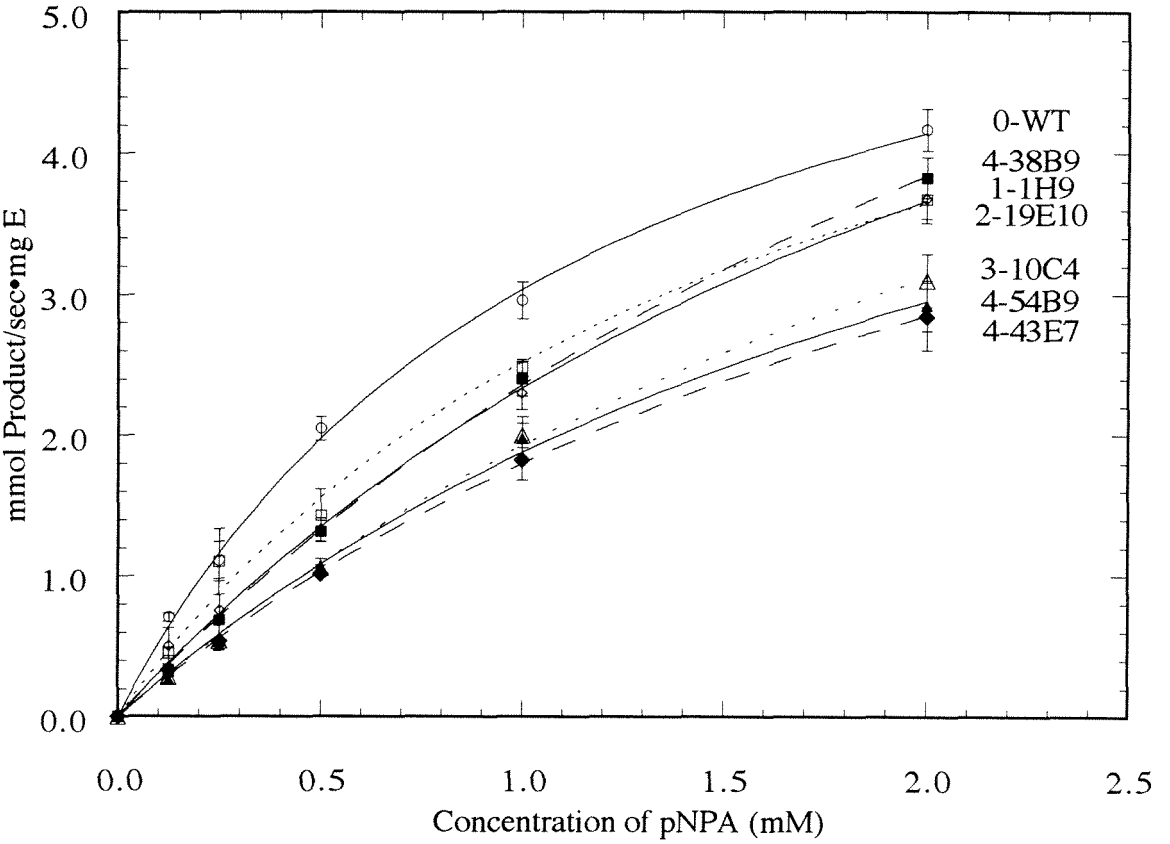



Figure 2.10a: Variant and wild type pNB esterase reaction kinetics on p-nitrophenyl acetate (pNPA) in aqueous buffer (0% dimethylformamide). Enzymes were added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0 and varying concentrations of pNPA.

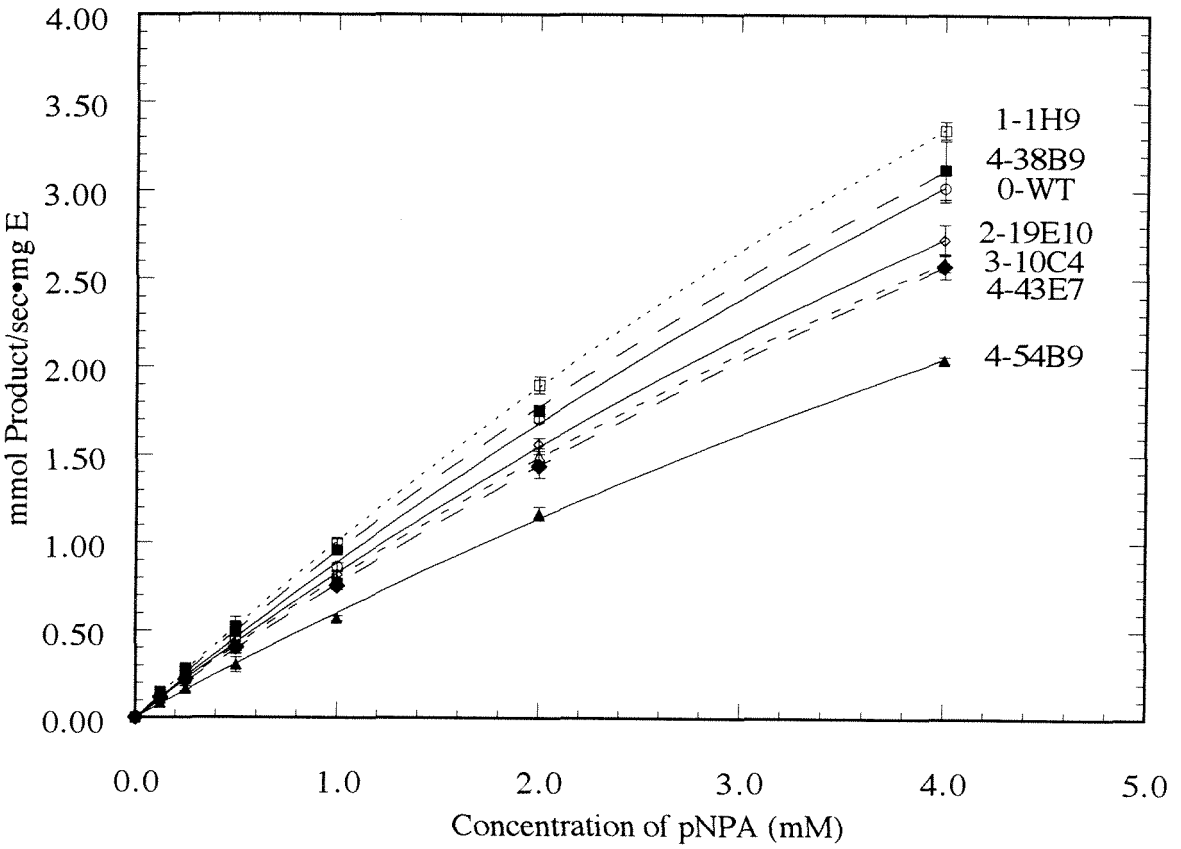



Figure 2.10b: Variant and wild type pNB esterase reaction kinetics on p-nitrophenyl acetate (pNPA) in 15% dimethylformamide (DMF). Enzymes were added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 15% DMF, and varying concentrations of pNPA.

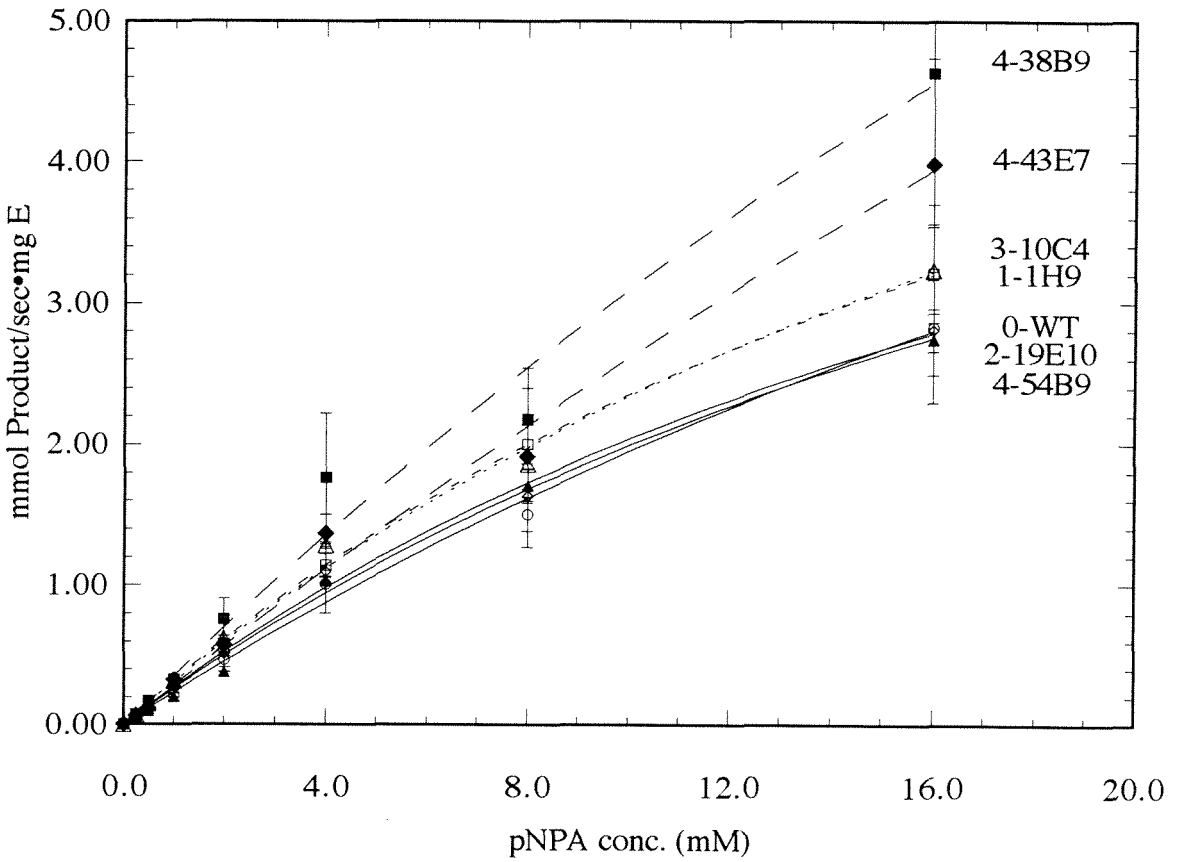



Figure 2.10c: Variant and wild type pNB esterase reaction kinetics on p-nitrophenyl acetate (pNPA) in 30% dimethylformamide (DMF). Enzymes were added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 30% DMF, and varying concentrations of pNPA.

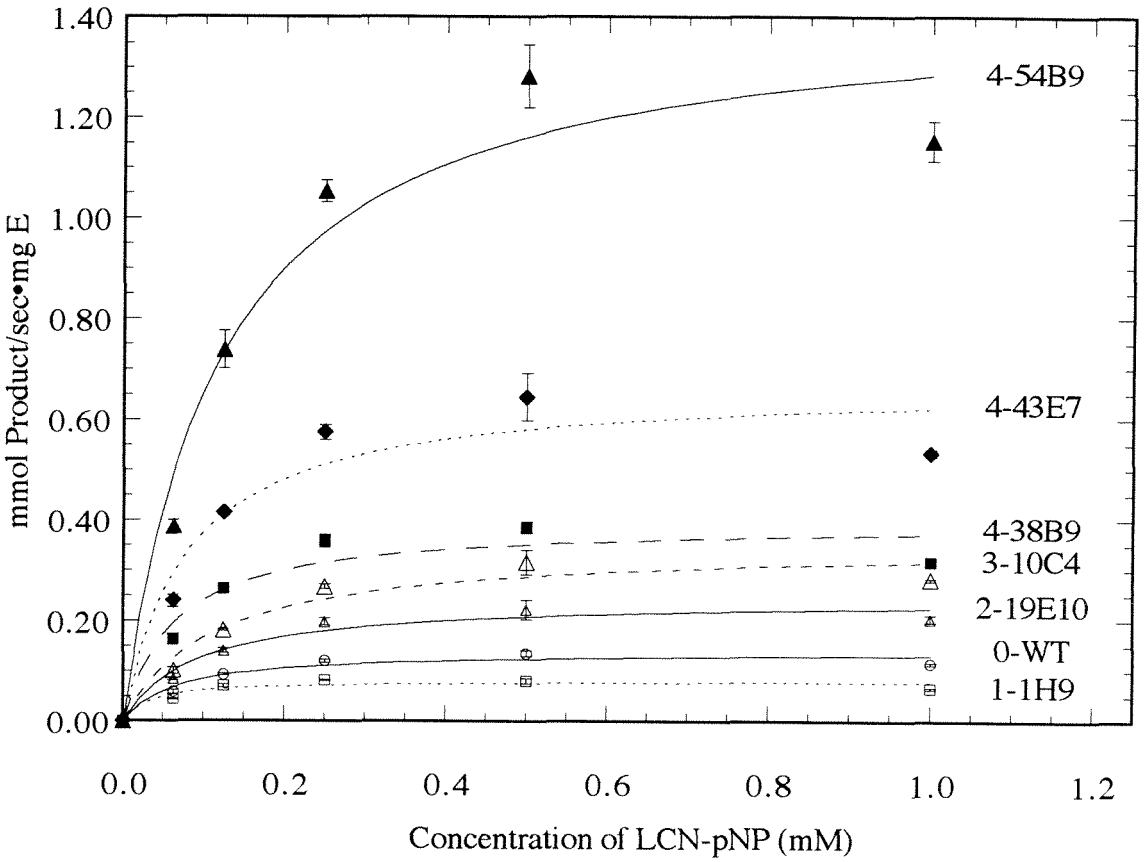



Figure 2.11a: Variant and wild type pNB esterase reaction kinetics on p-nitrophenyl loracarbef nucleus (LCN-pNP) in 1% dimethylformamide (DMF). Enzymes listed were added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 1% DMF, and varying concentrations of LCN-pNP.

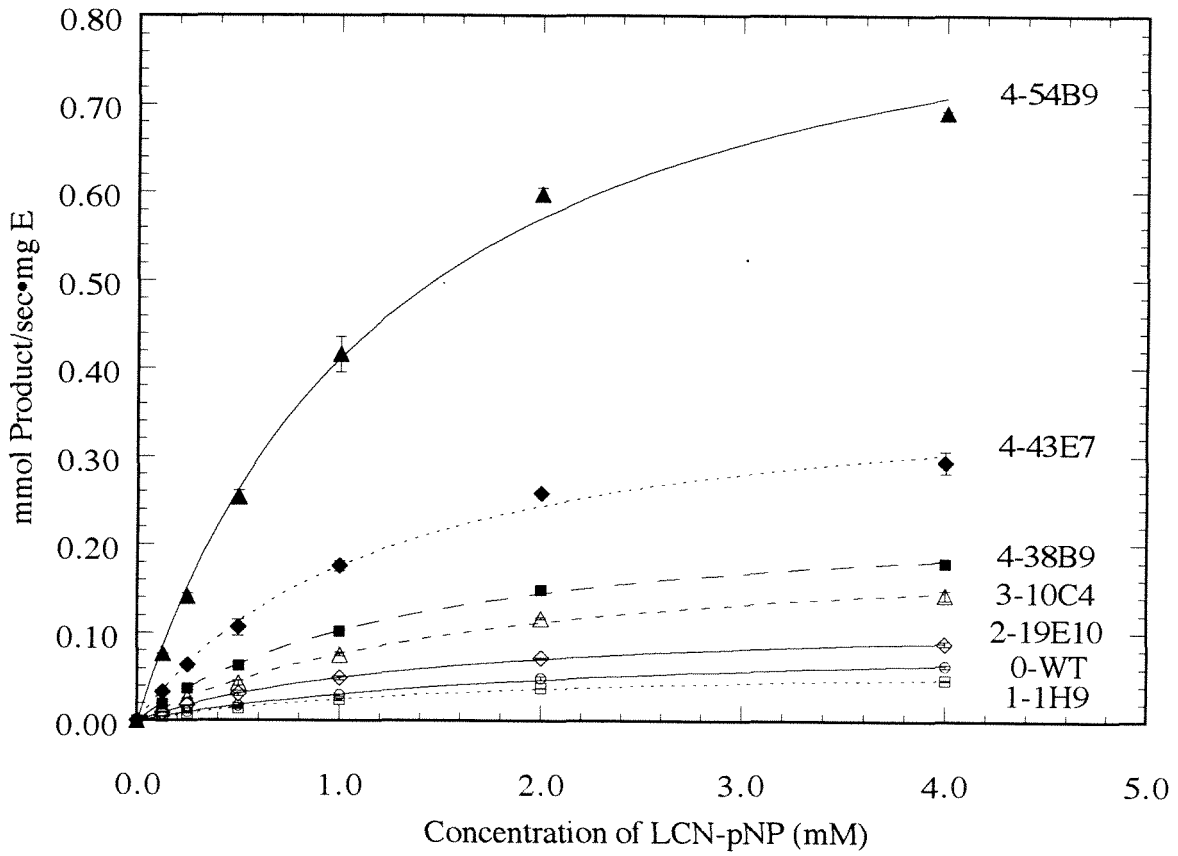



Figure 2.11b: Variant and wild type pNB esterase reaction kinetics on p-nitrophenyl loracarbef nucleus (LCN-pNP) in 15% dimethylformamide (DMF). Enzyme were added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 15% DMF, and varying concentrations of LCN-pNP.

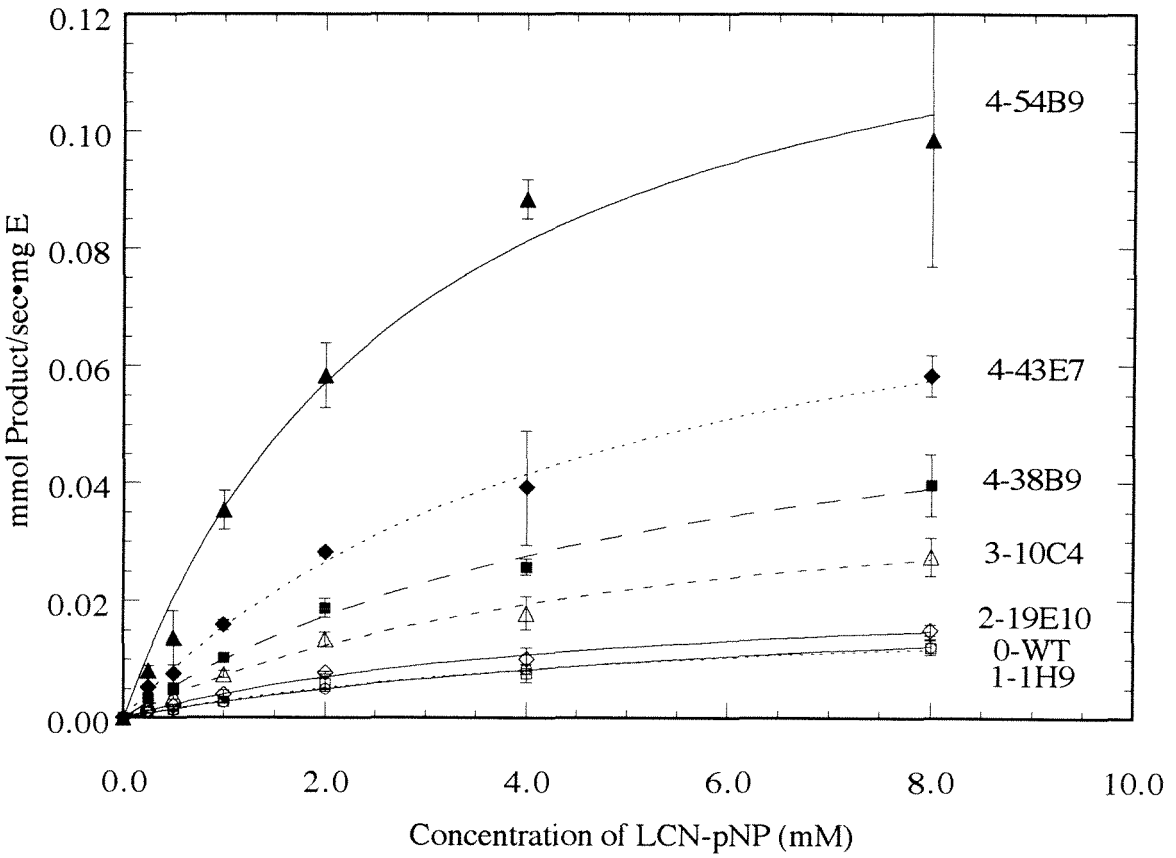



Figure 2.11c: Variant and wild type pNB esterase reaction kinetics on p-nitrophenyl loracarbef nucleus (LCN-pNP) in 30% dimethylformamide (DMF). Enzymes were added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 30% DMF, and varying concentrations of LCN-pNP.

these plots the results of directed evolution can clearly be seen. In 1, 15, and 30% DMF concentrations (Figures 11a, 11b, and 11c), the two variants least active on this substrate are the wild type enzyme and 1-1H9. This is not surprising given that neither of these enzymes had been exposed to this substrate previously. Additionally, as the concentration of DMF increases, the activity of 1-1H9 increases with respect to wild type, so that in 30% DMF (Figure 11c) these enzymes display identical kinetics. The next enzyme from the bottom is 2-19E10 from the second generation of mutagenesis and screening. This variant increases the maximum reaction rate by a factor of two over wild type and three over its parent 1-1H9 at low percentages of DMF. The activity increase is sensitive to the presence of DMF, decreasing to a smaller improvement in 30% DMF. This trend continues with the 3-10C4 variant, which is 50% faster at producing product than its 2-19E10 parent in 1% DMF. 3-10C4 does not lose activity as rapidly as its parent in DMF and is 100% faster than 2-19E10 in 30% DMF. 3-10C4 is the parent of the remaining variants, all of which show enhanced activity. With a 20% increase in activity in 1% DMF, 4-38B9 shows the least amount of improvement. This increase in activity is improved to a 50% increase by increasing concentrations of DMF to 30% DMF. 4-43E7 shows a constant two-fold increase in activity across all DMF ranges, and 4-54B9 is the most active of all the variants with a constant four-fold increase in activity over its parent. It is approximately 16 times more active than wild type pNB esterase. By comparing the scales of the axes, 4-54B9 retains the same activity in 30% DMF as the wild type enzyme in 1% DMF.

In these graphs we also begin to see what may be substrate inhibition, a feature noted in acetylcholinesterases. Note, for instance, the highest concentration of LCN-pNP in Figure 11a. For all the enzymes measured, the specific rate of hydrolysis for the highest concentration of substrate (1 mM) drops significantly in the absence of DMF. This feature seems diminished in the presence of DMF, as shown by Figures 11b and 11c. One possible explanation for why the presence of DMF diminishes the observed inhibition is that DMF seems to increase substrate $K_M$ values considerably (Chen and Arnold, 1993).

This effect might not only affect binding of the substrate for catalytic purposes, but for inhibition purposes as well by shifting inhibition to higher substrate concentrations.

Figure 12 shows the kinetic data obtained on the substrate of ultimate interest, LCN-pNB. All the same trends are observed on the LCN-pNB substrate that were seen on the screening substrate (i.e., the fourth generation variants are more active than the third generation variant, which is more active than the second generation variant, etc.) with only minor exceptions. The first exception is that 1-1H9 no longer lags wild type in specific activity. The second is that many of the variants, and especially those in the fourth generation, exhibit slightly lower increases in activity over wild type. For example, 4-54B9 is now approximately 14 times more active than wild type on LCN-pNB, versus 16 times wild type on LCN-pNP. Additionally, DMF affects the LCN-pNB hydrolysis reaction more strongly than it does the hydrolysis of LCN-pNP. On LCN-pNP, 15% DMF reduces the maximal activity by a factor of two in the two best variants in 1% DMF, while 15% DMF affects the LCN-pNB hydrolysis by reducing the maximal activity by a factor of three over the 1% DMF activity. Similar data in 30% DMF are not reported because the change in absorbance was monitored at 289 nm, and at this wavelength the increased DMF absorbs strongly enough to make accurate measurements difficult. This, combined with the high substrate concentrations required in 30% DMF, made accurate measurements impossible.

Figures 10, 11, and 12 were used to calculate the $k_{cat}$, $K_M$ and $k_{cat}/K_M$ values reported in Table 2. In the case of the pNPA substrate, only $k_{cat}/K_M$ is reported because the solubility of the pNPA substrate did not permit the high substrate concentrations required to accurately determine the individual $k_{cat}$ and $K_M$ parameters.

Examining the catalytic efficiency ($k_{cat}/K_M$) of the pNB esterase variants on pNPA, the wild type enzyme demonstrates the best performance in the absence of DMF (Table 2 and shown in Figure 13a). With increasing DMF, the 'evolved' enzyme variants compare more favorably with wild type, and at 30% DMF all the variants are more efficient than

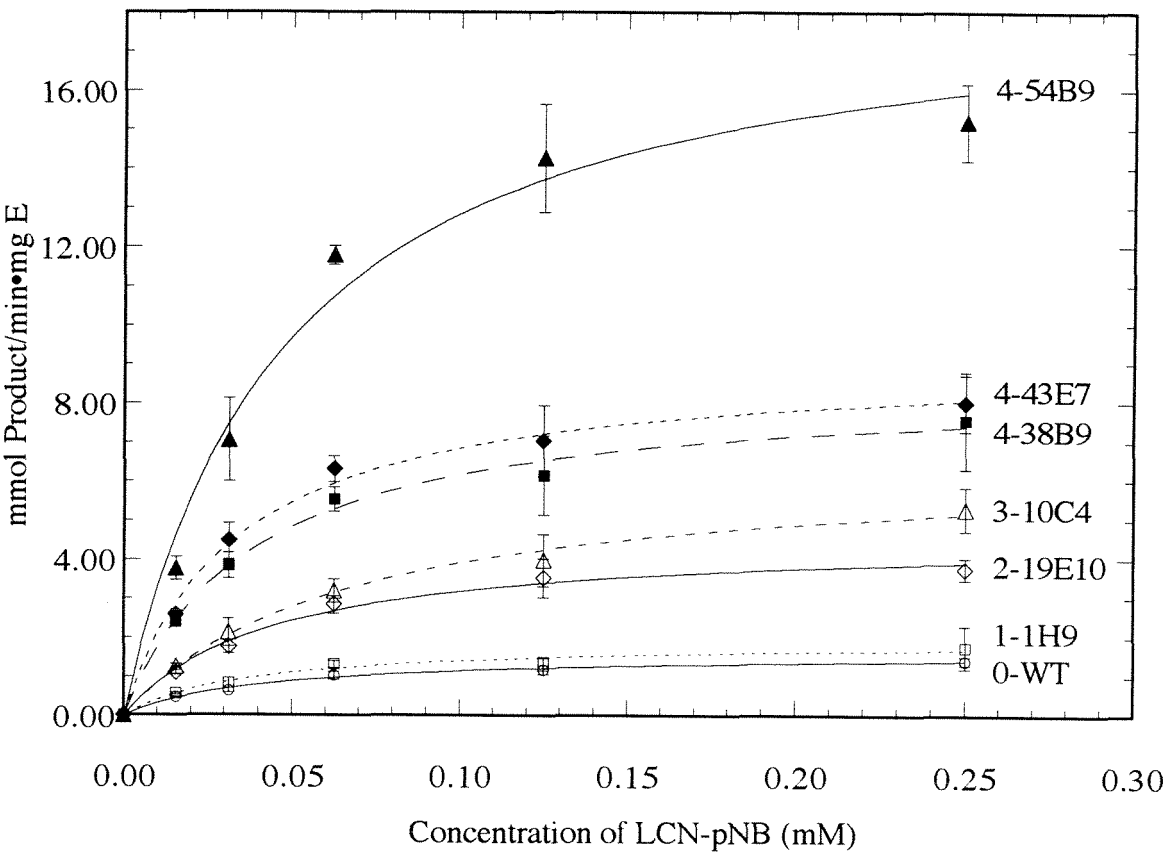



Figure 2.12a: Variant and wild type pNB esterase reaction kinetics on p-nitrobenzyl loracarbef nucleus (LCN-pNB) in 1% dimethylformamide (DMF). Enzymes were added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 1% DMF, and varying concentrations of LCN-pNB.

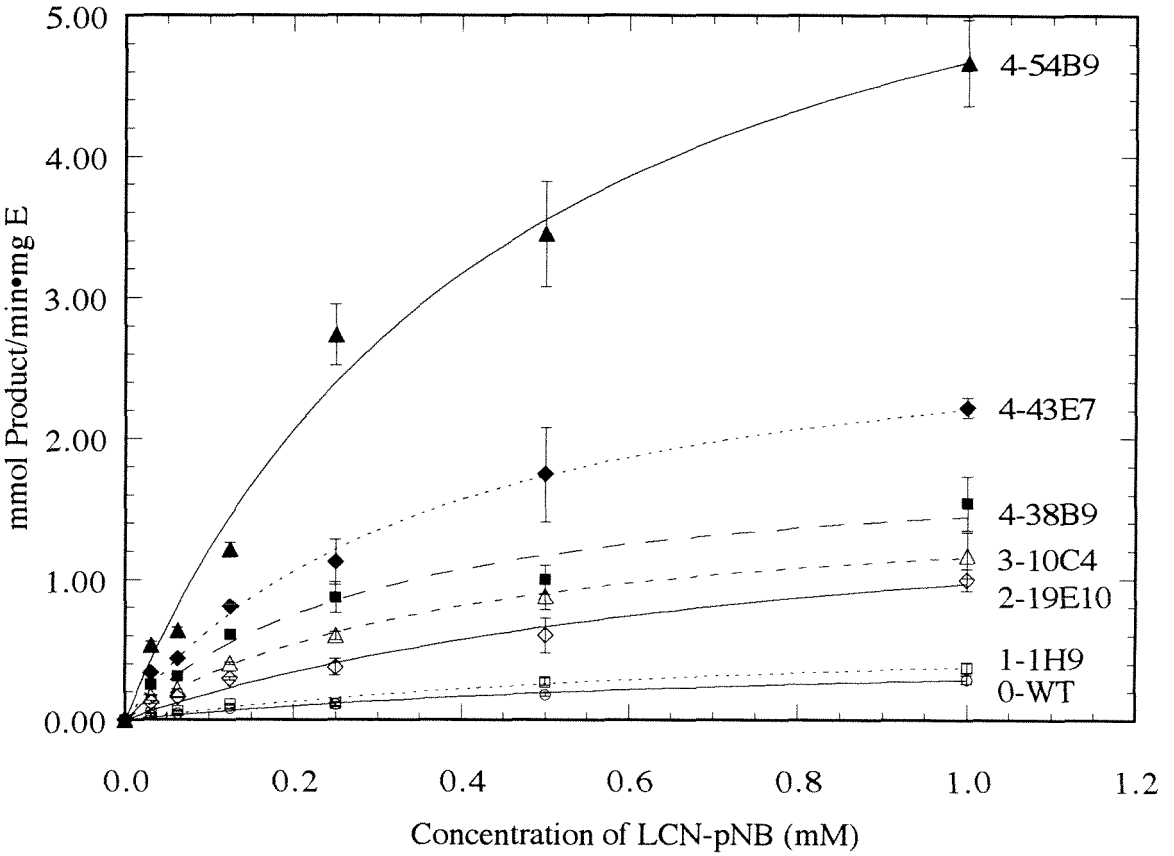



Figure 2.12b: Variant and wild type pNB esterase reaction kinetics on p-nitrobenzyl loracarbef nucleus (LCN-pNB) in 15% dimethylformamide (DMF). Enzyme was added to a 30 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 15% DMF, and varying concentrations of LCN-pNB.

p-nitrophenyl acetate

| Variant | 0% DMF $k_{cat}/K_M$ ($s^{-1} M^{-1}$) | 15% DMF $k_{cat}/K_M$ ($s^{-1} M^{-1}$) | 30% DMF $k_{cat}/K_M$ ($s^{-1} M^{-1}$) |
|---|---|---|---|
| 0-WT | 5700 | 940 | 240 |
| 1-1H9 | 4100 | 1100 | 320 |
| 2-19E10 | 3200 | 890 | 280 |
| 3-10C4 | 2500 | 860 | 330 |
| 4-38B9 | 3000 | 1000 | 360 |
| 4-43E7 | 2400 | 810 | 290 |
| 4-54B9 | 2600 | 640 | 270 |

loracarbef nucleus p-nitrophenyl ester

| | 1% DMF | | | 15% DMF | | | 30% DMF | | |
|---|---|---|---|---|---|---|---|---|---|
| Variant | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ |
| 0-WT | 0.14 | 0.07 | 2100 | 0.10 | 2.36 | 43 | 0.024 | 7.55 | 3 |
| 1-1H9 | 0.08 | 0.03 | 2500 | 0.07 | 1.84 | 38 | 0.021 | 6.05 | 3 |
| 2-19E10 | 0.25 | 0.09 | 2700 | 0.12 | 1.50 | 82 | 0.023 | 4.67 | 5 |
| 3-10C4 | 0.35 | 0.11 | 3200 | 0.21 | 1.77 | 118 | 0.044 | 5.18 | 9 |
| 4-38B9 | 0.39 | 0.06 | 6200 | 0.25 | 1.41 | 174 | 0.067 | 5.68 | 12 |
| 4-43E7 | 0.67 | 0.08 | 8500 | 0.40 | 1.27 | 313 | 0.094 | 5.09 | 18 |
| 4-54B9 | 1.44 | 0.12 | 12000 | 0.93 | 1.27 | 735 | 0.141 | 2.92 | 48 |

loracarbef nucleus p-nitrobenzyl ester

| | 1% DMF | | | 15% DMF | | |
|---|---|---|---|---|---|---|
| Variant | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ | $k_{cat}$ ($s^{-1}$) | $K_M$ (mM) | $k_{cat}/K_M$ |
| 0-WT | 0.027 | 0.04 | 640 | 0.009 | 0.88 | 10 |
| 1-1H9 | 0.032 | 0.04 | 850 | 0.012 | 0.85 | 14 |
| 2-19E10 | 0.077 | 0.04 | 1700 | 0.030 | 0.87 | 35 |
| 3-10C4 | 0.11 | 0.07 | 1600 | 0.027 | 0.40 | 68 |
| 4-38B9 | 0.14 | 0.04 | 3700 | 0.031 | 0.31 | 100 |
| 4-43E7 | 0.15 | 0.03 | 4500 | 0.050 | 0.37 | 135 |
| 4-54B9 | 0.32 | 0.05 | 6600 | 0.11 | 0.46 | 246 |

Table 2.2: Kinetic parameters $k_{cat}$, $K_m$ and $k_{cat}/K_m$ for pNB esterase variants, as derived from the curve fits of kinetic data. All $k_{cat}/K_M$ data are in units of $s^{-1} M^{-1}$. All data were collected in 0.1 M PIPES, pH 7.0 at 30 °C.

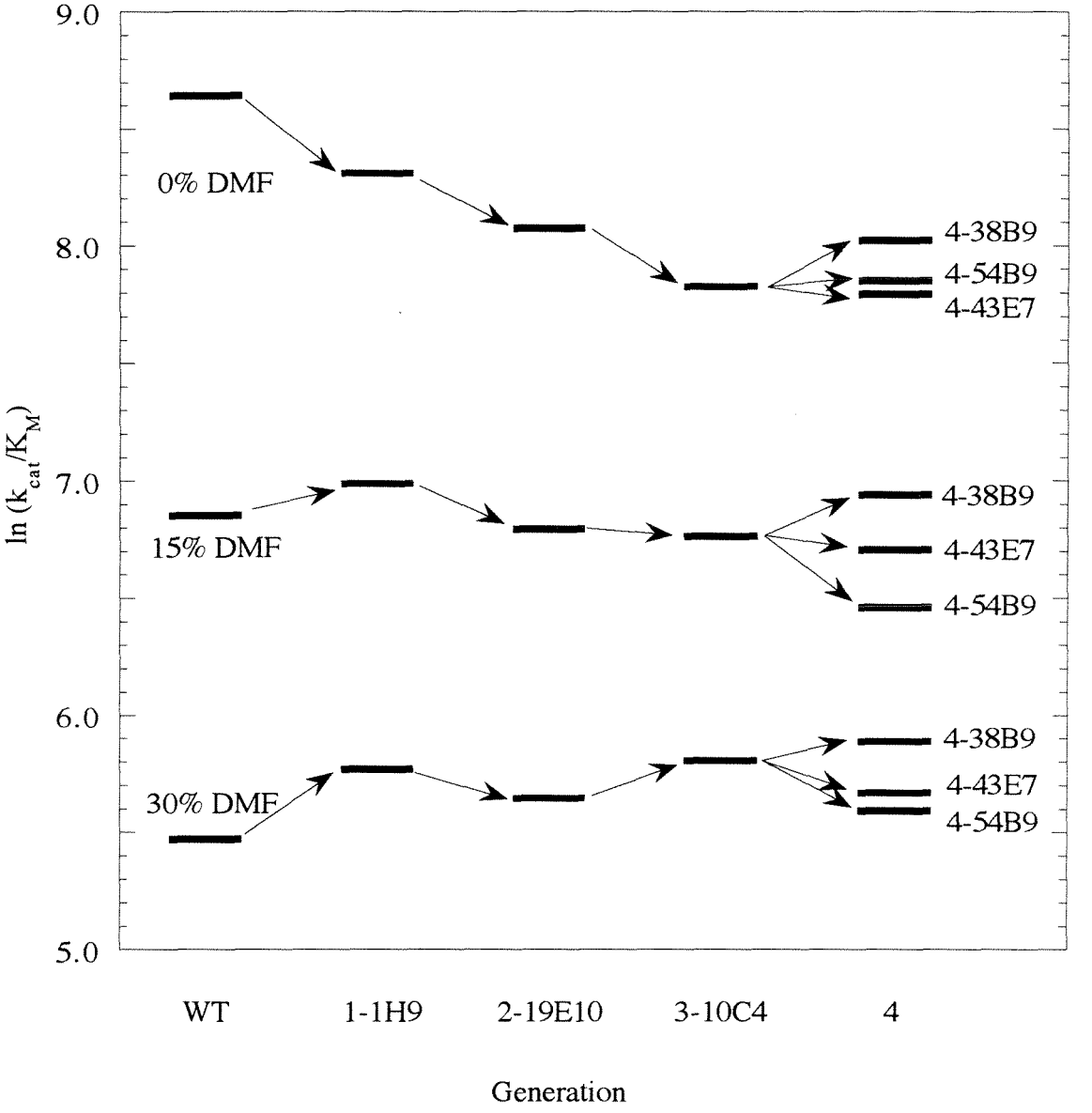



Figure 2.13a: Evolutionary progression of catalytic efficiencies for the evolved pNB esterases on pNPA. The pNPA substrate was used to screen the first generation. Subsequent generations were screened on LCN-pNP.

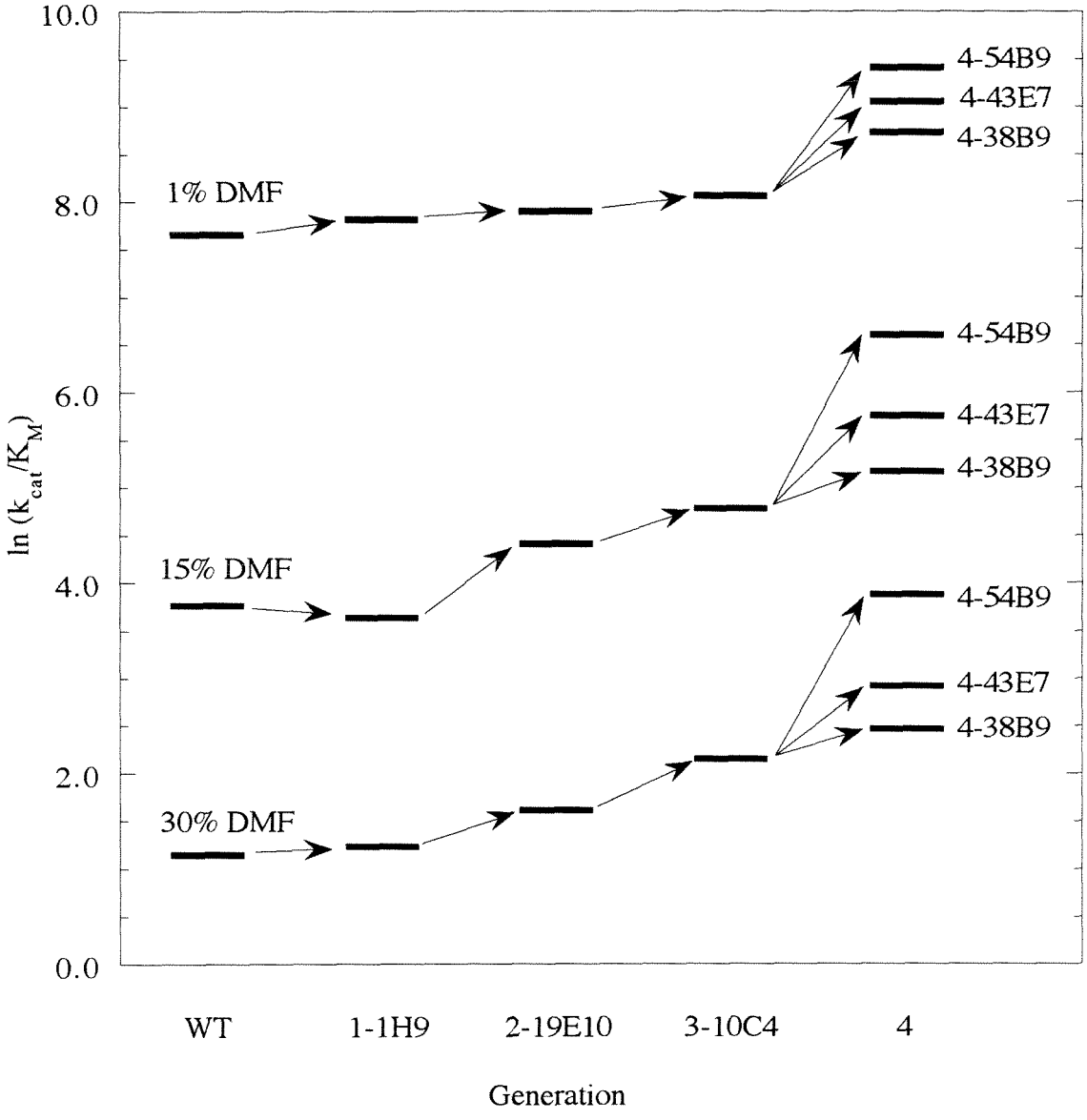



Figure 2.13b: Evolutionary progression of catalytic efficiencies for the evolved pNB esterases on LCN-pNP. The pNPA substrate was used to screen the first generation. Subsequent generations were screened on LCN-pNP.

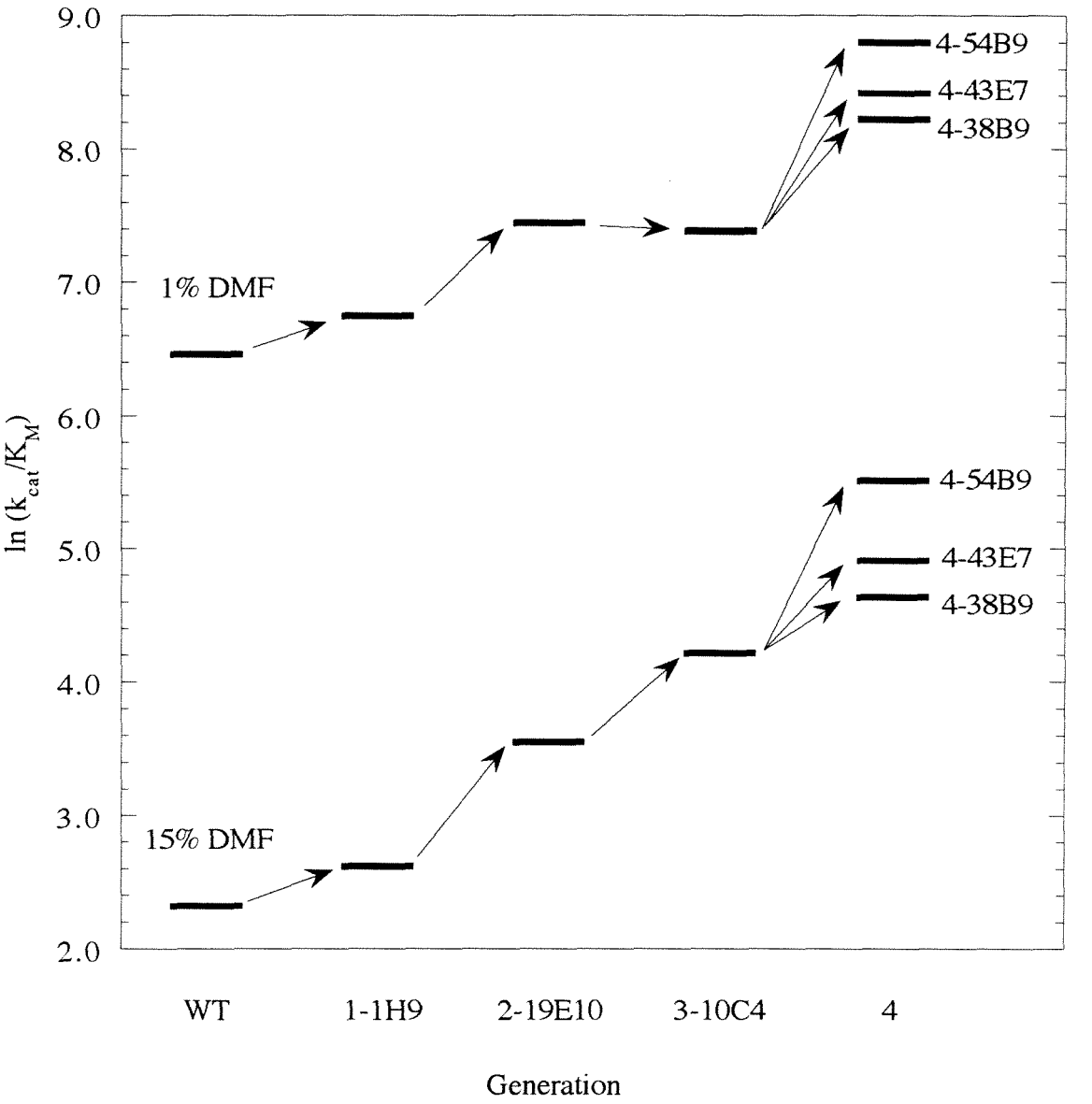
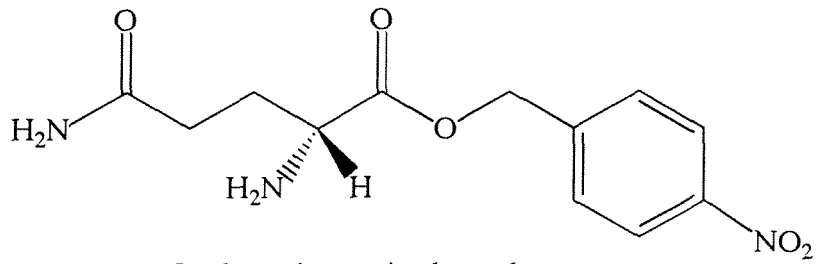



Figure 2.13c: Evolutionary progression of catalytic efficiencies for the evolved pNB esterases on LCN-pNB. The pNPA substrate was used to screen the first generation. Subsequent generations were screened on LCN-pNP.

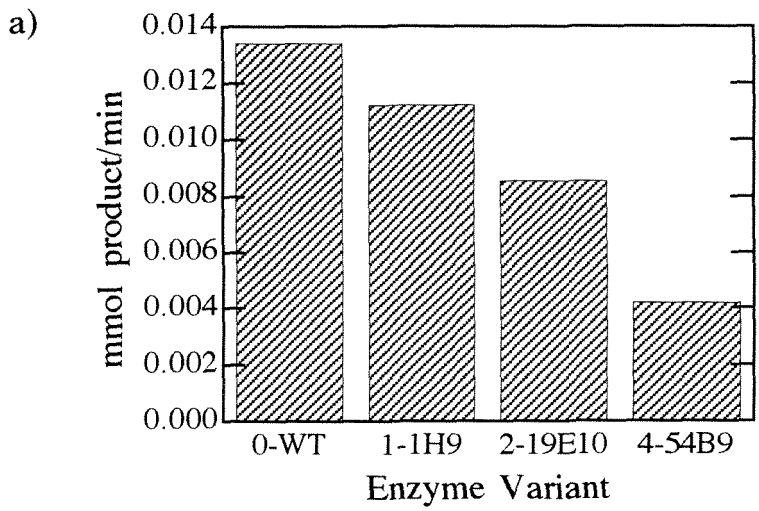
wild type. On LCN-pNP the wild-type enzyme begins at the bottom of the catalytic efficiency list and maintains that position throughout (Table 2 and Figure 13b). Note the turnover number $k_{cat}$ for wild type in the absence of DMF is 0.14 $s^{-1}$ is identical to the value of $k_{cat}$ for 4-54B9 in 30% DMF. In 30% DMF, 4-54B9 has the ability to hydrolyze LCN-pNP as fast as wild type does in water. The results for the LCN-pNB substrate are similar to those for the LCN-pNP substrate. The wild-type enzyme is again at the bottom of catalytic efficiency list (Table 2 and Figure 13c). While the catalytic efficiency of 4-54B9 is 10 times that of wild-type in the absence of DMF, it is 24 times that in 15% DMF.

*Enzyme Performance on Other Substrates*

Two other p-nitrobenzyl ester substrates, L-glutamine p-nitrobenzyl ester and p-nitrobenzyl benzoate, were tested to determine the extent to which newly constructed enzymes had increased activity as general pNB esterases. Each substrate was reacted with four purified pNB esterase variants (0-WT, 1-1H9, 2-19E10, and 4-54B9), and the reaction products were monitored using HPLC, as described in Materials and Methods. The resulting peak areas are shown for the different substrate/enzyme combinations in Figures 14 and 15. L-glutamine pNB is soluble in aqueous buffer. Its enzyme-catalyzed deprotection in the absence of DMF is shown in Figure 14a. All the enzyme variants catalyze this reaction, with wild type being the most active. As the similarity to wild type decreases from generation to generation, the activity of the enzyme variant decreases. The results are altered significantly when 20% DMF is added (Figure 14b). Under these conditions, wild type pNB esterase demonstrates activity toward removing the pNB-protecting group from L-glutamine p-nitrobenzyl ester; the first generation variant 1-1H9 outperforms wild type by 20%; and the second generation variant 2-19E10 outperforms the wild type enzyme by 60%. The fourth generation variant 4-54B9 apparently introduces a substitution which disrupts the earlier generations' enhancements of activity: this enzyme has lost the ability to catalyze this reaction better than wild type. The study of p-nitrobenzyl




L-glutamine p-nitrobenzyl ester



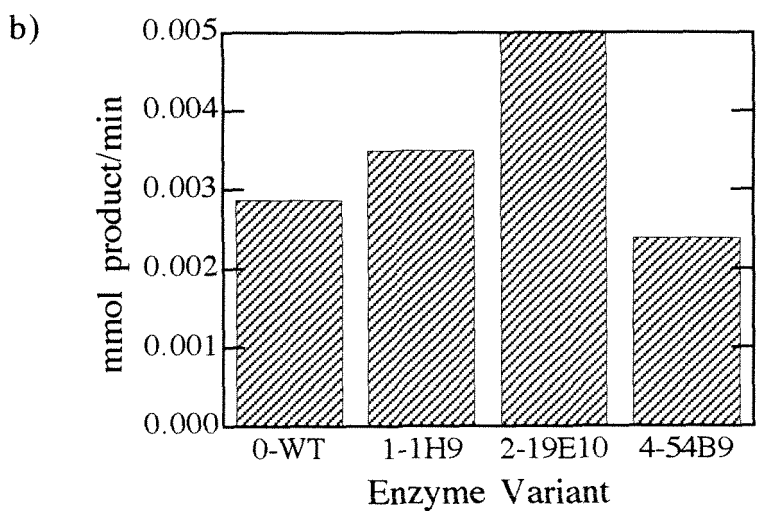
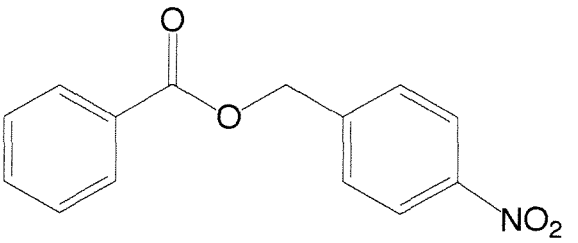




Figure 2.14a: Variant and wild type pNB esterase reaction rates on 1.0 mM L-glutamine p-nitrobenzyl ester in 1% dimethylformamide (DMF). Enzyme were added to a 25 °C reaction solution consisting of 0.1 M PIPES pH 7.0, 1% DMF and measured using HPLC. 2.14b: The same assay performed in 20% DMF.




p-nitrobenzyl benzoate

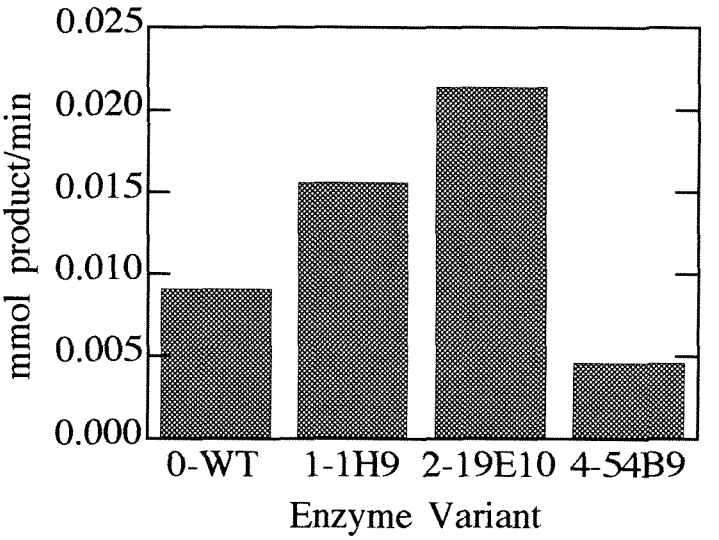



Figure 2.15: Variant and wild type pNB esterase reaction rates on 1.0 mM p-nitrobenzyl benzoate in 20% dimethylformamide (DMF). The reaction was performed in 0.1 M PIPES pH 7.0, 20% DMF and 1.0 mM substrate and monitored using HPLC.

benzoate in Figure 15 shows more dramatically the same trends demonstrated in Figure 14b. Because p-nitrobenzyl benzoate is not soluble in aqueous buffer, assays on involving this substrate were performed only in 20% DMF. In this environment 1-1H9 is 60% better than wild type and 2-19E10 is 2.5 times better than wild type. Again 4-54B9 has lost the ability to outperform previous generation variants, including the wild type pNB esterase.

*pH Studies*

The optimal activity of wild type pNB esterase occurs at pH 8.3 (Chen *et al.*, 1995). The screening, however, was carried out at pH 7.0. To see if the pH optima of any of the variant enzymes had been altered as a result of directed evolution, activity was measured as a function of pH. The activities were normalized to the maximum activity for each enzyme variant, and the resulting profiles are shown in Figure 16. While the pH optima of the enzyme variants are not significantly different from the wild type pNB esterase, the pH profiles of some of the enzyme variants have been systematically broadened through successive generations of evolution to increase the percentage of maximal velocity seen at lower pH values. At pH 7.5, the percent of maximal activity progresses from 60% in wild type to 85% in four generations. In cases where pH near an enzyme's optimum creates problems with substrate or product chemistry (such as stability or solubility), this suggests that enzyme activity can be improved at pH values away from the optimum. Substitutions which specifically alter the pH optimum appear to be much more rare than those which affect the pH activity profile.

*Sequence Analysis*

Figure 17 presents the DNA sequence alignment of all the variants sequenced during this study. The variants are listed in order by generation; the sequences start with DNA base one (A of the first codon ATG); and the DNA bases conserved between all members of this pNB esterase family are boxed. Where a mutation has occurred, the

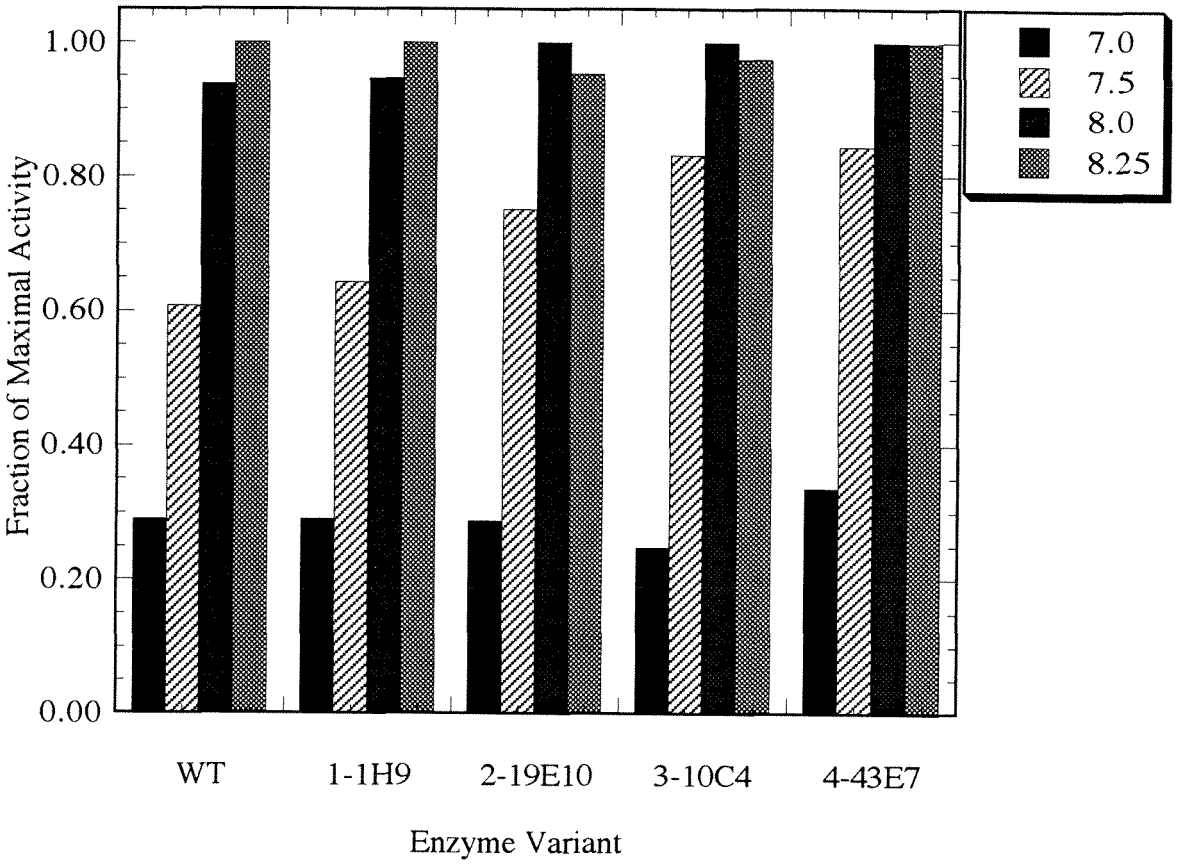



Figure 2.16: pH-activity profiles from pH 7.0 to 8.25 are displayed for a variant from each generation. The activity values for each variant are normalized to its maximal value. The enzyme samples were added to a 30 °C reaction solution consisting of 1% dimethyl-formamide (DMF), 0.5 mM p-nitrophenyl loracarbef nucleus, and 0.1 M PIPES pH 7.0 to 8.25.

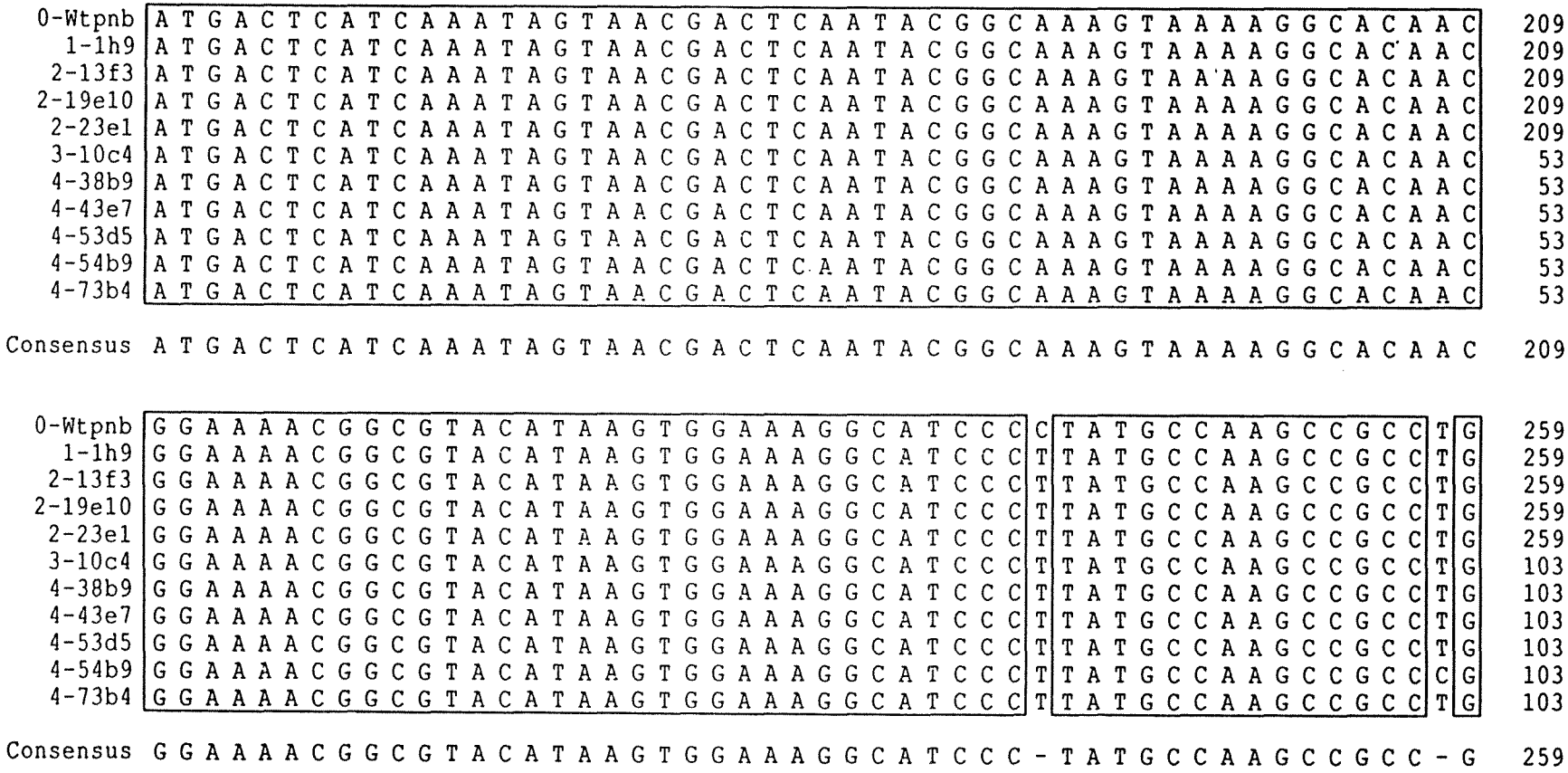
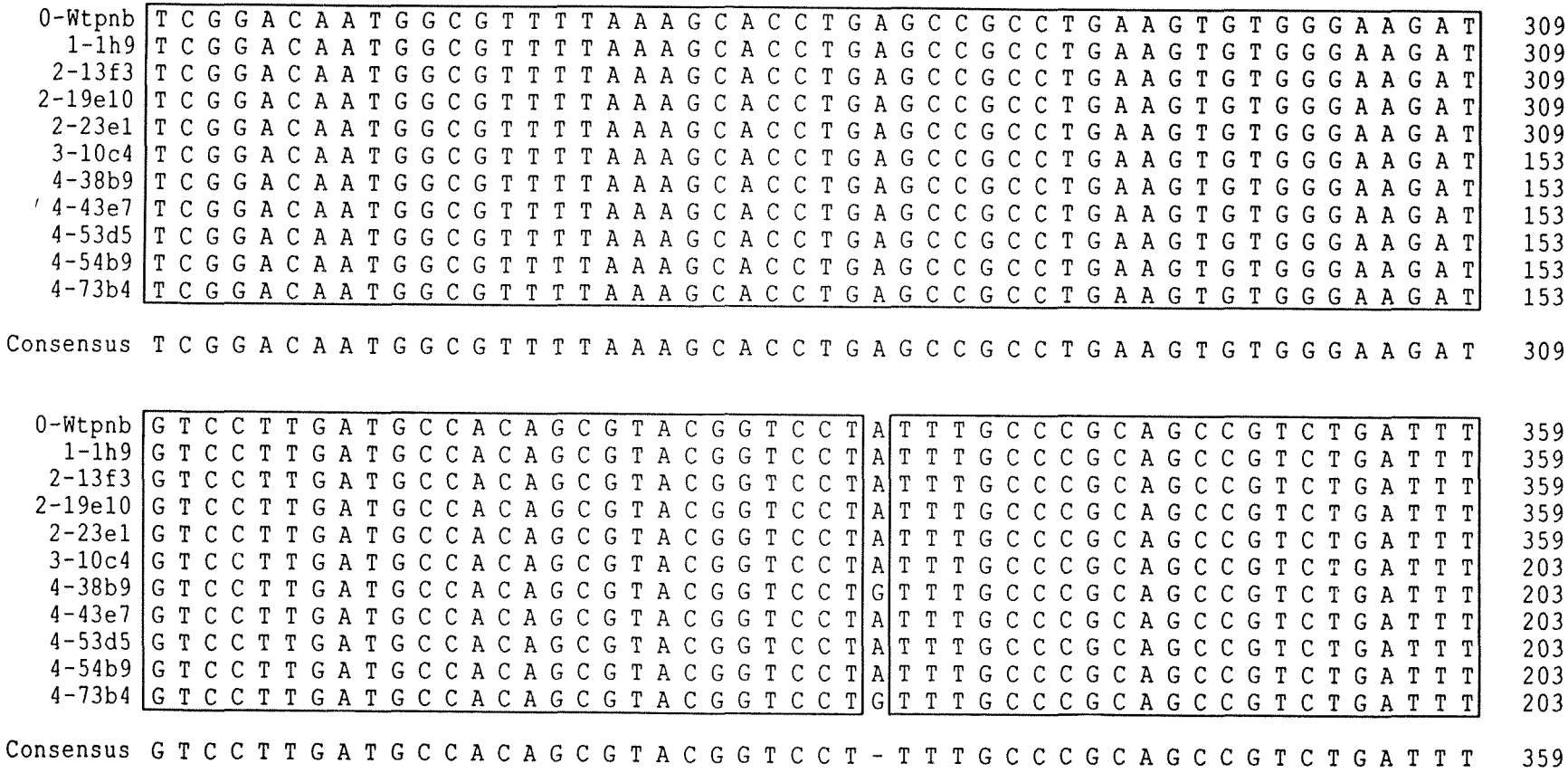
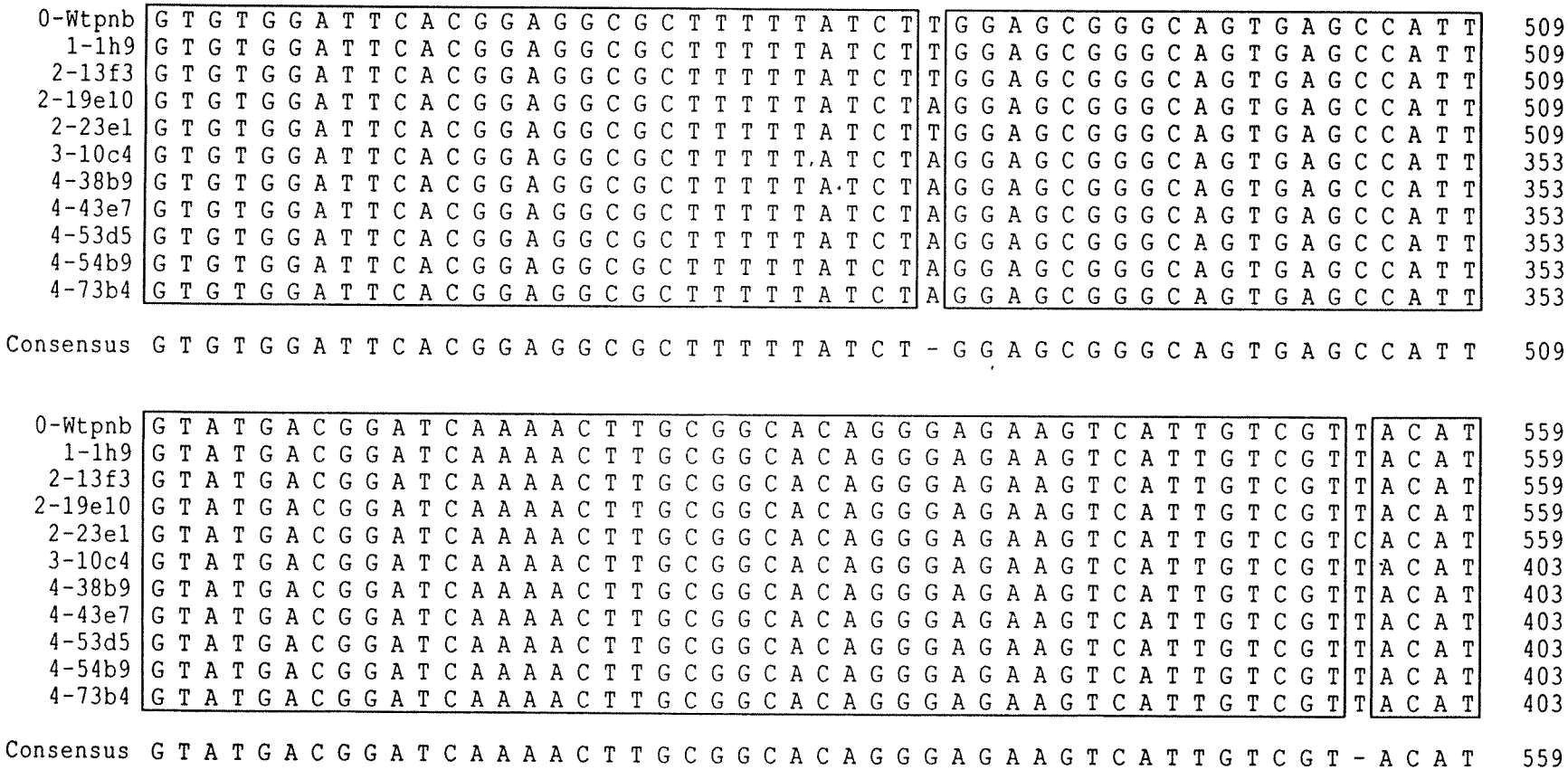



Figure 2.17: DNA sequence alignment of all enzyme variants sequenced in this study. The variants are listed from top to bottom by generation. Boxed regions indicate DNA sequence regions where all variants are identical. The columns of DNA bases not boxed are those where at least one mutation has occurred. These locations are further indicated by dashes in the consensus sequence. This figure was generated using PILEUP and PRETTYPLOT from the GCG software package.

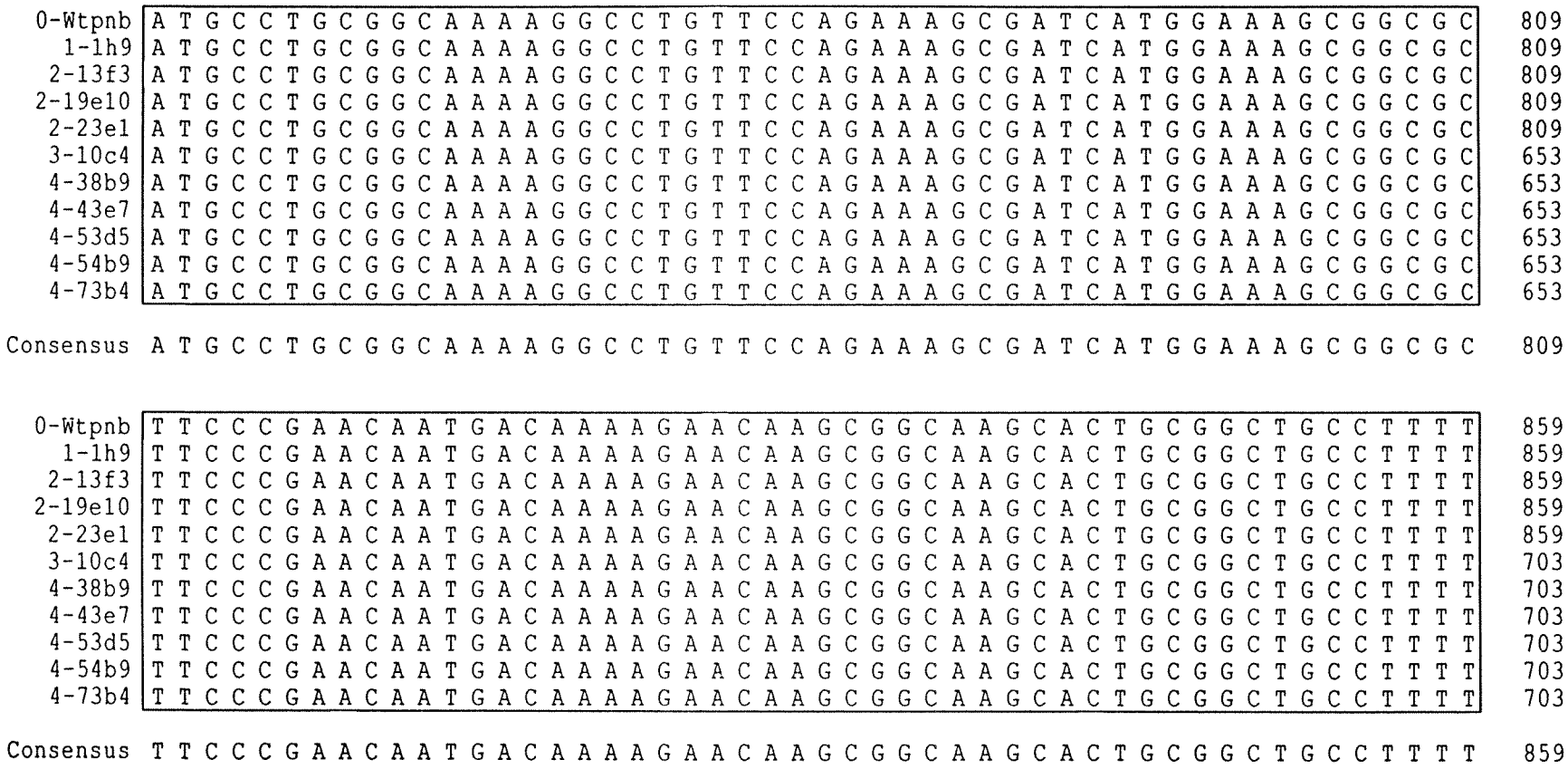

0-Wtpnb T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 309
1-1h9 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 309
2-13f3 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 309
2-19e10 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 309
2-23e1 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 309
3-10c4 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 153
4-38b9 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 153
4-43e7 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 153
4-53d5 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 153
4-54b9 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 153
4-73b4 T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 153
Consensus T C G G A C A A T G G C G T T T T A A A G C A C C T G A G C C G C C T G A A G T G T G G G A A G A T 309
0-Wtpnb G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 359
1-1h9 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 359
2-13f3 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 359
2-19e10 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 359
2-23e1 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 359
3-10c4 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 203
4-38b9 G T C C T T G A T G C C A C A G C G T A C G G T C C T G T T T G C C C G C A G C C G T C T G A T T T 203
4-43e7 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 203
4-53d5 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 203
4-54b9 G T C C T T G A T G C C A C A G C G T A C G G T C C T A T T T G C C C G C A G C C G T C T G A T T T 203
4-73b4 G T C C T T G A T G C C A C A G C G T A C G G T C C T G T T T G C C C G C A G C C G T C T G A T T T 203
Consensus G T C C T T G A T G C C A C A G C G T A C G G T C C T - T T T G C C C G C A G C C G T C T G A T T T 359

```
0-Wtpnb   G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   409
  1-1h9   G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   409
  2-13f3  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   409
 2-19e10  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   409
  2-23e1  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   409
  3-10c4  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   253
  4-38b9  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   253
  4-43e7  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   253
  4-53d5  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   253
  4-54b9  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   253
  4-73b4  G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   253

Consensus G C T C T C A C T G T C G T A T A C A G A G C T G C C C C G C C A G T C C G A G G A T T G C T T G T   409


0-Wtpnb   A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A T C T T C C T G T C A T G   459
  1-1h9   A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A T C T T C C T G T C A T G   459
  2-13f3  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A G T C T T C C T G T C A T G   459
 2-19e10  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A T C T T C C T G T C A T G   459
  2-23e1  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A T C T T C C T G T C A T G   459
  3-10c4  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A C C T T C C T G T C A T G   303
  4-38b9  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A C C T T C C T G T C A T G   303
  4-43e7  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A G G T C A A A A C C T T C C T G T C A T G   303
  4-53d5  A C G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A C C T T C C T G T C A T G   303
  4-54b9  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A C C T T C C T G T C A T G   303
  4-73b4  A T G T C A A T G T A T T T G C G C C T G A C A C T C C A A G T C A A A A C C T T C C T G T C A T G   303

Consensus A - G T C A A T G T A T T T G C G C C T G A C A C T C C A - G T C A A A - - C T T C C T G T C A T G   459
```

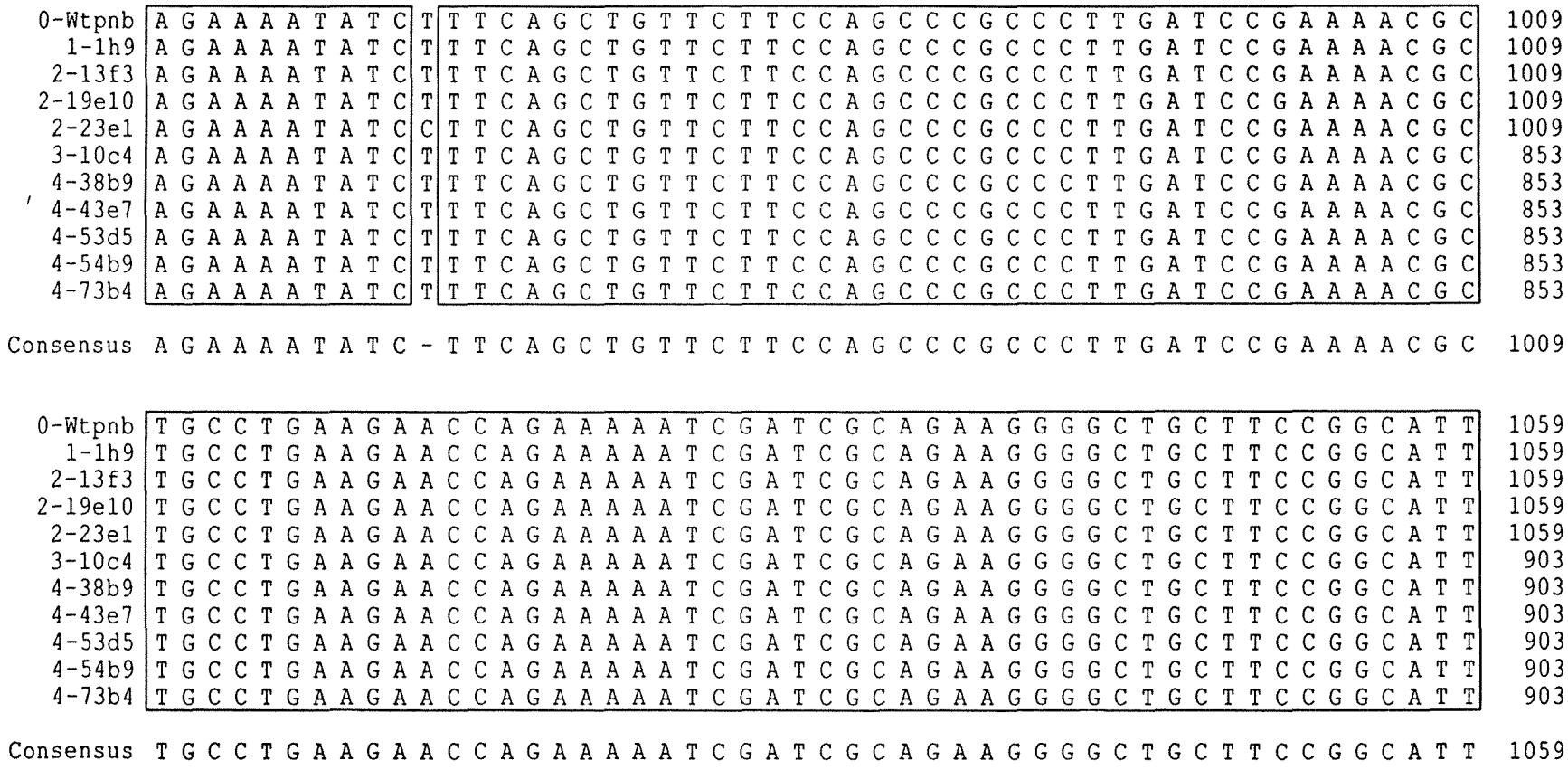

0-Wtpnb G T G T G G A T T C A C G G A G G C G C T T T T T A T C T T G G A G C G G G C A G T G A G C C A T T 509
1-1h9 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T T G G A G C G G G C A G T G A G C C A T T 509
2-13f3 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T T G G A G C G G G C A G T G A G C C A T T 509
2-19e10 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T A G G A G C G G G C A G T G A G C C A T T 509
2-23e1 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T T G G A G C G G G C A G T G A G C C A T T 509
3-10c4 G T G T G G A T T C A C G G A G G C G C T T T T T . A T C T A G G A G C G G G C A G T G A G C C A T T 353
4-38b9 G T G T G G A T T C A C G G A G G C G C T T T T T A . T C T A G G A G C G G G C A G T G A G C C A T T 353
4-43e7 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T A G G A G C G G G C A G T G A G C C A T T 353
4-53d5 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T A G G A G C G G G C A G T G A G C C A T T 353
4-54b9 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T A G G A G C G G G C A G T G A G C C A T T 353
4-73b4 G T G T G G A T T C A C G G A G G C G C T T T T T A T C T A G G A G C G G G C A G T G A G C C A T T 353
Consensus G T G T G G A T T C A C G G A G G C G C T T T T T A T C T - G G A G C G G G C A G T G A G C C A T T 509
0-Wtpnb G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 559
1-1h9 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 559
2-13f3 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 559
2-19e10 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 559
2-23e1 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T C A C A T 559
3-10c4 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 403
4-38b9 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 403
4-43e7 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 403
4-53d5 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 403
4-54b9 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 403
4-73b4 G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T T A C A T 403
Consensus G T A T G A C G G A T C A A A A C T T G C G G C A C A G G G A G A A G T C A T T G T C G T - A C A T 559

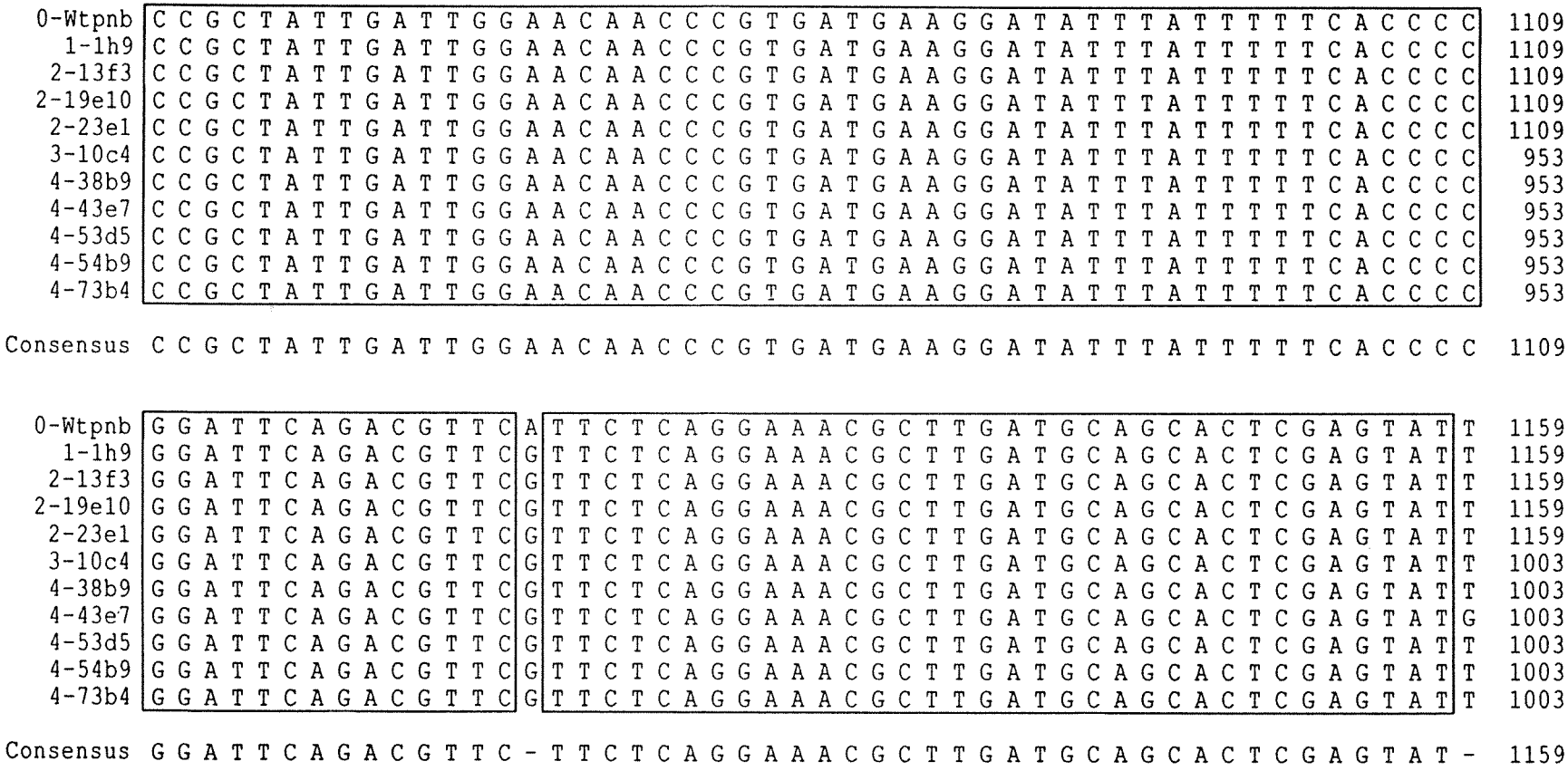
```
0-Wtpnb   TGAACTATCGGCTGGGGCCGTTTGGCTTTTGCACTTGTCTTCGTTTGAT  609
  1-1h9   TGAACTATCGGCTGGGGCCGTTTGGCTTTTGCACTTGTCTTCGTTTGAT  609
 2-13f3   TGAACTATCGGCTGGGGCCGTTTGGCTTTTGCACTTGTCTTCGTTTGAT  609
2-19e10   TGAACTATCGGCTGGGGCCGTTTGGCTTTTGCACTTGTCTTCGTTTGAT  609
 2-23e1   TGAACTATCGGCTGGGGCCGTTTGGCTTTTGCACTTGTCTTCGTTTGAT  609
 3-10c4   TGAACTATCGGCTGGGGCCGTTTGGCTTTAGCACTTGTCTTCGTTTGAT  453
 4-38b9   TGAACTATCGGCTGGGGCCGTTTGGCTTTAGCACTTGTCTTCGTTTGAT  453
 4-43e7   TGAACTATCGGCTGGGGCCGTTTGGCTTTAGCACTTGTCTTCGTTTGAT  453
 4-53d5   TGAACTATCGGCTGGGGCCGTTTGGCTTTAGCACTTGTCTTCGTTTGAT  453
 4-54b9   TGAACTATCGGCTGGGGCCGTTTGGCTTTAGCACTTGTCTTCGTTTGAT  453
 4-73b4   TGAACTATCGGCTGGGGCCGTTTGGCTTTAGCACTTGTCTTCGTTTGAT  453

Consensus TGAACTATCGGCTGGGGCCGTTTGGCTTT-TGCACTTGTCTTCGTTTGAT  609


0-Wtpnb   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  659
  1-1h9   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  659
 2-13f3   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  659
2-19e10   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  659
 2-23e1   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  659
 3-10c4   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  503
 4-38b9   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  503
 4-43e7   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  503
 4-53d5   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  503
 4-54b9   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  503
 4-73b4   GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  503

Consensus GAGGCGTATTCCGATAACCTTGGGGCTTTTAGACCAAGCCGCCGCGCTGAA  659
```

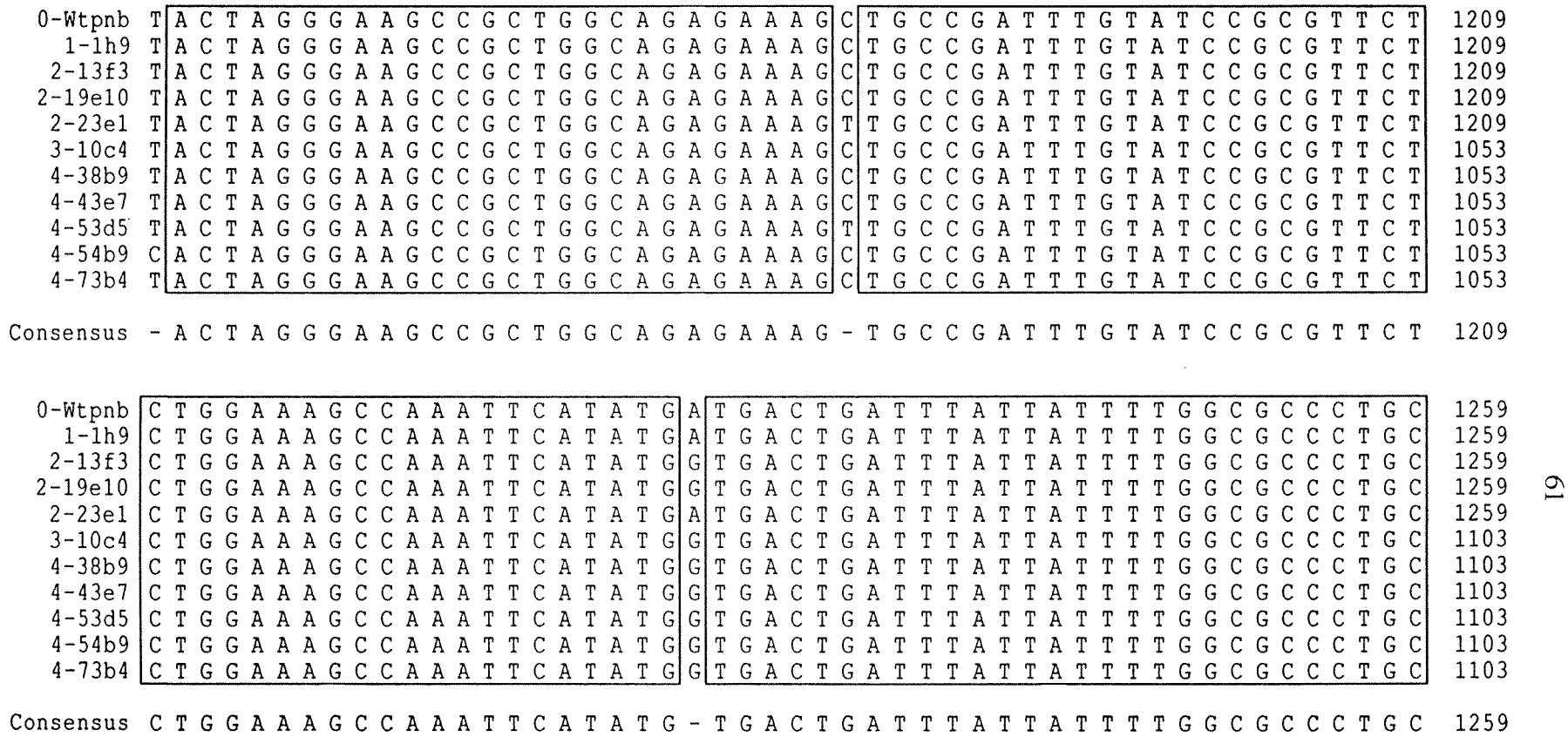
```
0-Wtpnb   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   709
  1-1h9   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   709
 2-13f3   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   709
2-19e10   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCG-ATAACGTAA   709
 2-23e1   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   709
 3-10c4   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   553
 4-38b9   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   553
 4-43e7   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   553
 4-53d5   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   553
 4-54b9   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   553
 4-73b4   ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   553

Consensus ATGGGTGCGGGAGAATATCTCAGCGTTTGGCGGTGATCCCGATAACGTAA   709


0-Wtpnb   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   759
  1-1h9   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   759
 2-13f3   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   759
2-19e10   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   759
 2-23e1   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   759
 3-10c4   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   603
 4-38b9   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   603
 4-43e7   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   603
 4-53d5   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   603
 4-54b9   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   603
 4-73b4   CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   603

Consensus CAGTATTTGGAGAATCCGCCGGCGGCATGAGCATTGCCGCGCTGCTCGCT   759
```


0-Wtpnb A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 809
1-1h9 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 809
2-13f3 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 809
2-19e10 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 809
2-23e1 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 809
3-10c4 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 653
4-38b9 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 653
4-43e7 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 653
4-53d5 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 653
4-54b9 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 653
4-73b4 A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 653
Consensus A T G C C T G C G G C A A A A G G C C T G T T C C A G A A A G C G A T C A T G G A A A G C G G C G C 809
0-Wtpnb T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 859
1-1h9 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 859
2-13f3 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 859
2-19e10 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 859
2-23e1 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 859
3-10c4 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 703
4-38b9 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 703
4-43e7 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 703
4-53d5 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 703
4-54b9 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 703
4-73b4 T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 703
Consensus T T C C C G A A C A A T G A C A A A A G A A C A A G C G G C A A G C A C T G C G G C T G C C T T T T 859

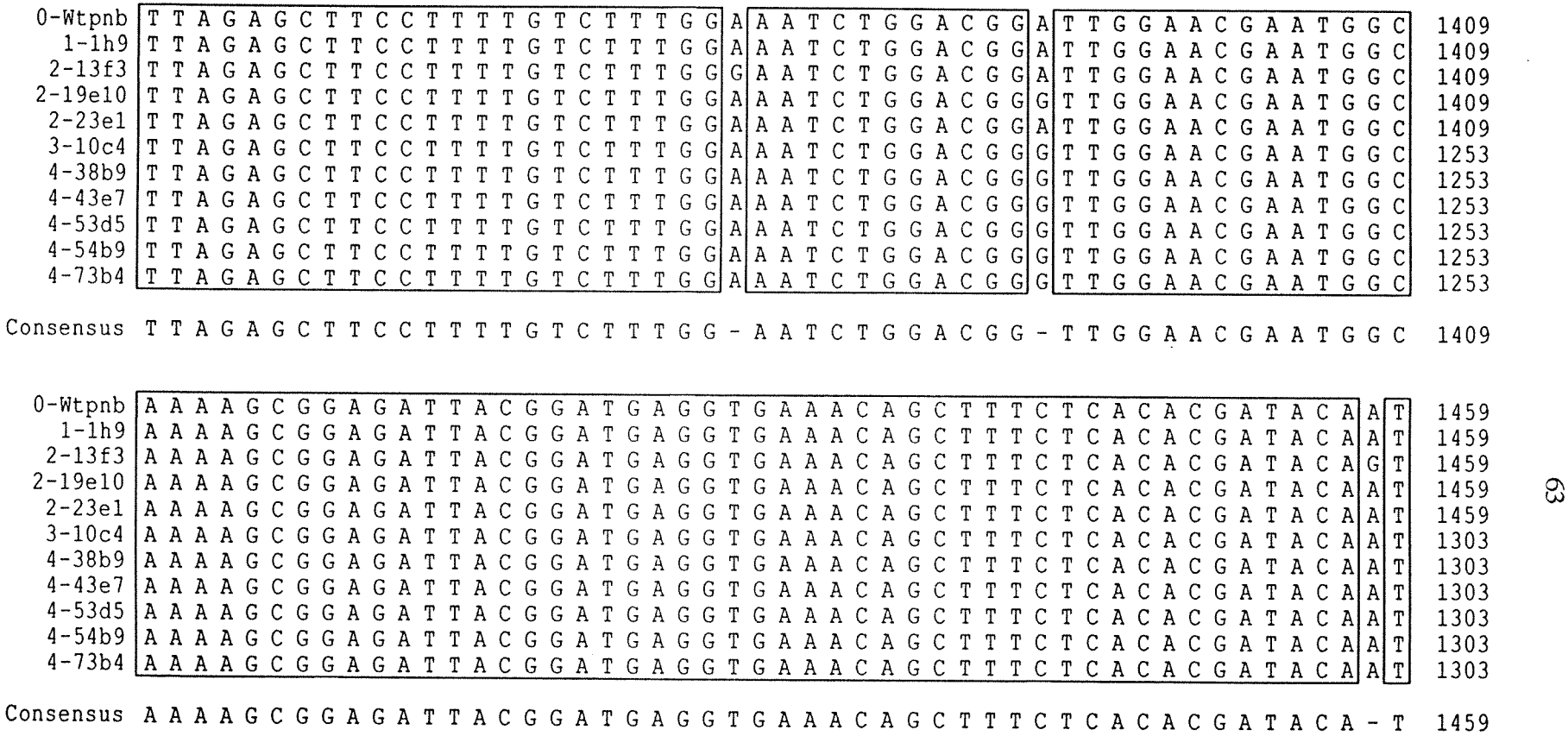
```
0-Wtpnb   T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   909
  1-1h9   T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   909
  2-13f3  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   909
 2-19e10  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   909
  2-23e1  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   909
  3-10c4  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   753
  4-38b9  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   753
  4-43e7  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   753
  4-53d5  T A C A G G T C C T T G G G A T C A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   753
  4-54b9  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   753
  4-73b4  T A C A G G T C C T T G G G A T T A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   753

Consensus T A C A G G T C C T T G G G A T - A A T G A G A G C C A G C T G G A C A G A T T G C A T A C T G T A   909


0-Wtpnb   G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   959
  1-1h9   G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   959
  2-13f3  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   959
 2-19e10  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   959
  2-23e1  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   959
  3-10c4  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   803
  4-38b9  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   803
  4-43e7  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   803
  4-53d5  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   803
  4-54b9  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A G   803
  4-73b4  G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A A   803

Consensus G C A G C G G A A G A T T T G C T T A A A G C G G C C G A T C A G C T T C G G A T T G C A G A A A -   959
```


0-Wtpnb A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 1009
1-1h9 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 1009
2-13f3 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 1009
2-19e10 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 1009
2-23e1 A G A A A A T A T C C T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 1009
3-10c4 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 853
4-38b9 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 853
4-43e7 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 853
4-53d5 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 853
4-54b9 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 853
4-73b4 A G A A A A T A T C T T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 853
Consensus A G A A A A T A T C - T T C A G C T G T T C T T C C A G C C C G C C C T T G A T C C G A A A A C G C 1009
0-Wtpnb T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 1059
1-1h9 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 1059
2-13f3 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 1059
2-19e10 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 1059
2-23e1 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 1059
3-10c4 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 903
4-38b9 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 903
4-43e7 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 903
4-53d5 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 903
4-54b9 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 903
4-73b4 T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 903
Consensus T G C C T G A A G A A C C A G A A A A A T C G A T C G C A G A A G G G G C T G C T T C C G G C A T T 1059

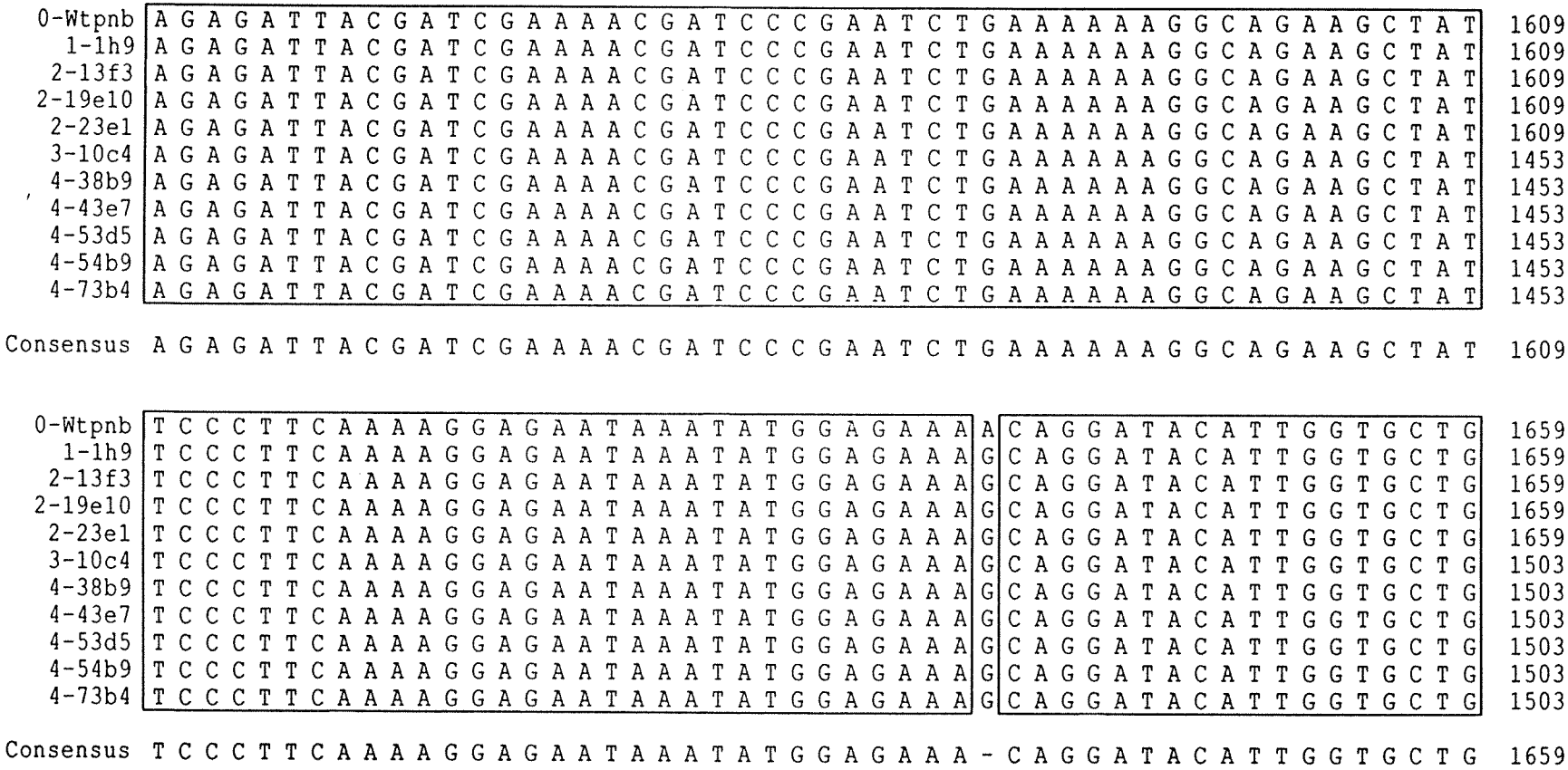

0-Wtpnb C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 1109
1-1h9 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 1109
2-13f3 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 1109
2-19e10 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 1109
2-23e1 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 1109
3-10c4 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 953
4-38b9 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 953
4-43e7 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 953
4-53d5 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 953
4-54b9 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 953
4-73b4 C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 953
Consensus C C G C T A T T G A T T G G A A C A A C C C G T G A T G A A G G A T A T T T A T T T T T C A C C C C 1109
0-Wtpnb G G A T T C A G A C G T T C A T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1159
1-1h9 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1159
2-13f3 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1159
2-19e10 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1159
2-23e1 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1159
3-10c4 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1003
4-38b9 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1003
4-43e7 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T G 1003
4-53d5 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1003
4-54b9 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1003
4-73b4 G G A T T C A G A C G T T C G T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T T 1003
Consensus G G A T T C A G A C G T T C - T T C T C A G G A A A C G C T T G A T G C A G C A C T C G A G T A T - 1159

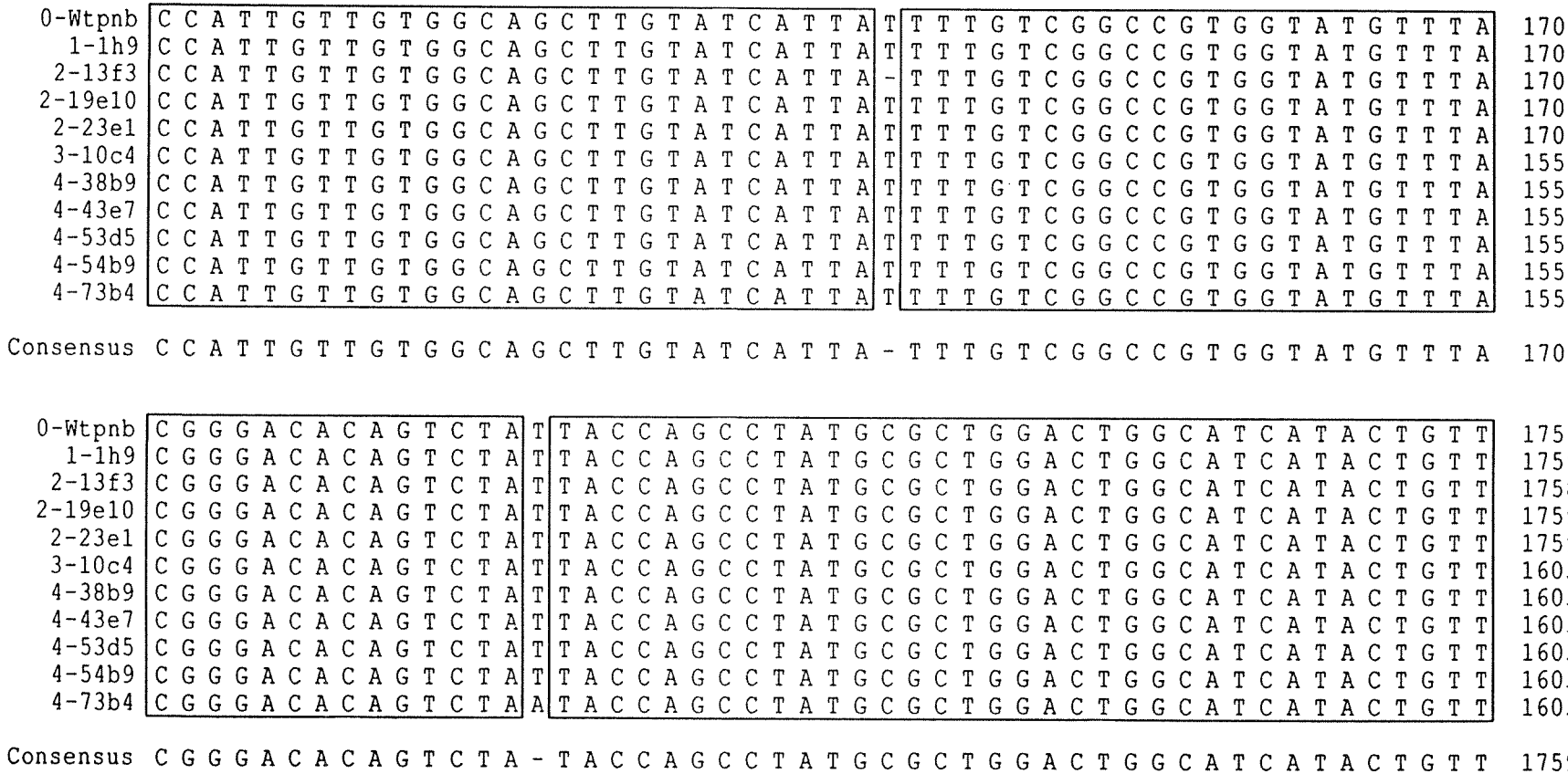

0-Wtpnb T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1209
1-1h9 T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1209
2-13f3 T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1209
2-19e10 T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1209
2-23e1 T A C T A G G G A A G C C G C T G G C A G A G A A A G T T G C C G A T T T G T A T C C G C G T T C T 1209
3-10c4 T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1053
4-38b9 T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1053
4-43e7 T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1053
4-53d5 T A C T A G G G A A G C C G C T G G C A G A G A A A G T T G C C G A T T T G T A T C C G C G T T C T 1053
4-54b9 C A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1053
4-73b4 T A C T A G G G A A G C C G C T G G C A G A G A A A G C T G C C G A T T T G T A T C C G C G T T C T 1053
Consensus - A C T A G G G A A G C C G C T G G C A G A G A A A G - T G C C G A T T T G T A T C C G C G T T C T 1209
0-Wtpnb C T G G A A A G C C A A A T T C A T A T G A T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1259
1-1h9 C T G G A A A G C C A A A T T C A T A T G A T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1259
2-13f3 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1259
2-19e10 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1259
2-23e1 C T G G A A A G C C A A A T T C A T A T G A T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1259
3-10c4 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1103
4-38b9 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1103
4-43e7 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1103
4-53d5 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1103
4-54b9 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1103
4-73b4 C T G G A A A G C C A A A T T C A T A T G G T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1103
Consensus C T G G A A A G C C A A A T T C A T A T G - T G A C T G A T T T A T T A T T T T G G C G C C C T G C 1259

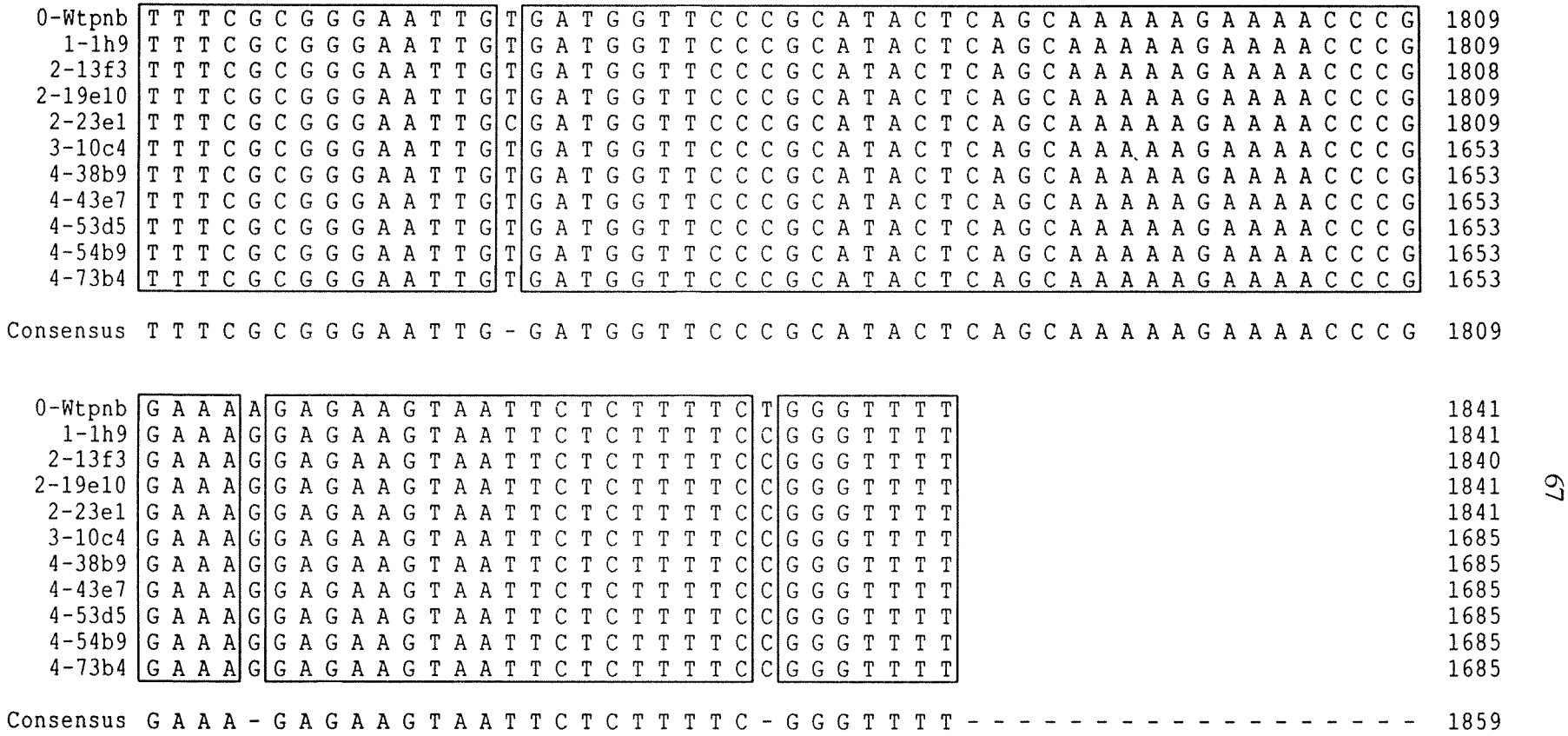
```
0-Wtpnb  C G T C G C C T A T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1309
  1-1h9  C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1309
  2-13f3 C G T C G C C T T T G C A T C C G C G C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1309
2-19e10  C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1309
  2-23e1 C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1309
  3-10c4 C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1153
  4-38b9 C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1153
  4-43e7 C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1153
  4-53d5 C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1153
  4-54b9 C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1153
  4-73b4 C G T C G C C T T T G C A T C C G C A C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1153

Consensus C G T C G C C T - T G C A T C C G C - C A G T C T C A T T A C G C C C C T G T C T G G A T G T A C C  1309

0-Wtpnb  G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1359
  1-1h9  G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1359
  2-13f3 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1359
2-19e10  G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1359
  2-23e1 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1359
  3-10c4 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1203
  4-38b9 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1203
  4-43e7 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1203
  4-53d5 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1203
  4-54b9 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1203
  4-73b4 G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1203

Consensus G G T T C G A T T G G C A C C C G G A G A A G C C G C C G T A C A A T A A A G C G T T T C A C G C A  1359
```


0-Wtpnb T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G A T T G G A A C G A A T G G C 1409
1-1h9 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G A T T G G A A C G A A T G G C 1409
2-13f3 T T A G A G C T T C C T T T T G T C T T T G G G A A T C T G G A C G G A T T G G A A C G A A T G G C 1409
2-19e10 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G G T T G G A A C G A A T G G C 1409
2-23e1 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G A T T G G A A C G A A T G G C 1409
3-10c4 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G G T T G G A A C G A A T G G C 1253
4-38b9 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G G T T G G A A C G A A T G G C 1253
4-43e7 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G G T T G G A A C G A A T G G C 1253
4-53d5 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G G T T G G A A C G A A T G G C 1253
4-54b9 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G G T T G G A A C G A A T G G C 1253
4-73b4 T T A G A G C T T C C T T T T G T C T T T G G A A A T C T G G A C G G G T T G G A A C G A A T G G C 1253
Consensus T T A G A G C T T C C T T T T G T C T T T G G - A A T C T G G A C G G - T T G G A A C G A A T G G C 1409
0-Wtpnb A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1459
1-1h9 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1459
2-13f3 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A G T 1459
2-19e10 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1459
2-23e1 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1459
3-10c4 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1303
4-38b9 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1303
4-43e7 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1303
4-53d5 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1303
4-54b9 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1303
4-73b4 A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A A T 1303
Consensus A A A A G C G G A G A T T A C G G A T G A G G T G A A A C A G C T T T C T C A C A C G A T A C A - T 1459

```
0-Wtpnb  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1509
  1-1h9  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1509
 2-13f3  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1509
2-19e10  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1509
 2-23e1  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1509
 3-10c4  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1353
 4-38b9  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1353
 4-43e7  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1353
 4-53d5  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1353
 4-54b9  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1353
 4-73b4  C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1353

Consensus C C G C G T G G A T C A C G T T C G C T A A A A C A G G A A A C C C A A G C A C C G A A G C T G T G  1509


0-Wtpnb  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1559
  1-1h9  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1559
 2-13f3  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1559
2-19e10  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1559
 2-23e1  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1559
 3-10c4  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1403
 4-38b9  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1403
 4-43e7  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1403
 4-53d5  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1403
 4-54b9  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1403
 4-73b4  A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1403

Consensus A A T T G G C C G G C G T A T C A T G A A G A A A C G A G A G A G A C G G T G A T T T T A G A C T C  1559
```


0-Wtpnb A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1609
1-1h9 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1609
2-13f3 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1609
2-19e10 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1609
2-23e1 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1609
3-10c4 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1453
4-38b9 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1453
4-43e7 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1453
4-53d5 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1453
4-54b9 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1453
4-73b4 A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1453
Consensus A G A G A T T A C G A T C G A A A A C G A T C C C G A A T C T G A A A A A A G G C A G A A G C T A T 1609
0-Wtpnb T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A C A G G A T A C A T T G G T G C T G 1659
1-1h9 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1659
2-13f3 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1659
2-19e10 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1659
2-23e1 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1659
3-10c4 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1503
4-38b9 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1503
4-43e7 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1503
4-53d5 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1503
4-54b9 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1503
4-73b4 T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A A G C A G G A T A C A T T G G T G C T G 1503
Consensus T C C C T T C A A A A G G A G A A T A A A T A T G G A G A A A - C A G G A T A C A T T G G T G C T G 1659


0-Wtpnb C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1709
1-1h9 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1709
2-13f3 C C A T T G T T G T G G C A G C T T G T A T C A T T A - T T T G T C G G C C G T G G T A T G T T T A 1708
2-19e10 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1709
2-23e1 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1709
3-10c4 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1553
4-38b9 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1553
4-43e7 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1553
4-53d5 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1553
4-54b9 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1553
4-73b4 C C A T T G T T G T G G C A G C T T G T A T C A T T A T T T G T C G G C C G T G G T A T G T T T A 1553
Consensus C C A T T G T T G T G G C A G C T T G T A T C A T T A - T T T G T C G G C C G T G G T A T G T T T A 1709
0-Wtpnb C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1759
1-1h9 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1759
2-13f3 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1758
2-19e10 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1759
2-23e1 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1759
3-10c4 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1603
4-38b9 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1603
4-43e7 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1603
4-53d5 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1603
4-54b9 C G G G A C A C A G T C T A T T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1603
4-73b4 C G G G A C A C A G T C T A A T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1603
Consensus C G G G A C A C A G T C T A - T A C C A G C C T A T G C G C T G G A C T G G C A T C A T A C T G T T 1759


0-Wtpnb TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1809
1-1h9 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1809
2-13f3 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1808
2-19e10 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1809
2-23e1 TTTCGCGGGAATTGCGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1809
3-10c4 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1653
4-38b9 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1653
4-43e7 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1653
4-53d5 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1653
4-54b9 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1653
4-73b4 TTTCGCGGGAATTGTGATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1653
Consensus TTTCGCGGGAATTG-GATGGTTCCCGCATACTCAGCAAAAAGAAAACCCG 1809
0-Wtpnb GAAAAGAGAAGTAATTCTCTTTTCTGGGTTTT 1841
1-1h9 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1841
2-13f3 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1840
2-19e10 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1841
2-23e1 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1841
3-10c4 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1685
4-38b9 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1685
4-43e7 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1685
4-53d5 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1685
4-54b9 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1685
4-73b4 GAAAGGAGAAGTAATTCTCTTTTCCGGGTTTT 1685
Consensus GAAA-GAGAAGTAATTCTCTTTTC-GGGTTTT---------- 1859

column of DNA bases is not boxed, and a dash is indicated in the consensus sequence at the bottom of each set of rows. Table 3 summarizes the positions of the DNA base changes in the enzyme variants with respect to the wild type pNB esterase gene sequence for all the variants sequenced during this study. Reverse type indicates the substitutions not present in the previous generation parent enzyme. Horizontal lines indicate the begin and end of the open reading frame, which starts at base position 1 and ends at position 1470. All together the sequences contain 31 substitutions, of which 29 are unambiguously unique. The two substitutions which are unlikely to be unique are those where identical substitutions were found in two different variants of the same generation: the A to G substitutions at position 1075 in 2-13F3 and 2-19E10, and the A to G substitution at position 181 in 4-38B9 and 4-73B4. Figure 18 plots the frequency of substitutions as a function of substitution within the gene sequence. This plot demonstrates that the locations of substitutions leading to enhanced enzyme activity are well distributed throughout the target sequence.

The types of substitutions generated, however, are not well distributed. Of the 29 unique substitutions, 25 were substitutions changing an A or T, 4 were substitutions from C, and none were substitutions of G. These bases were changed almost half of the time to G, with A to G changes making up the majority (9/10) of these substitutions. This non-randomness is also shown in the number of transition (purine to purine changes - A to G or C to T) to transversion (purine to pyrimidine changes - A to T/G or C to A/G or T to A/G) substitutions, where the transitions outnumber the transversions 24 to 5.

Also included in the DNA sequence information is the amino acid substitution information. Figure 19 presents the amino acid sequence alignment of the pNB esterase family as translated from the DNA sequence alignment in Figure 17. As before, the variants are listed in order by generation; the sequences start with amino acid one; and the DNA bases conserved between all members of this pNB esterase family are boxed. Where a mutation has occurred, the column of amino acid residues is not boxed, and a dash is

| DNA Pos. | Gen. 1 | Generation 2 | | | Gen. 3 | Generation 4 | | | | | Gen. 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1H9 | 13F3 | 19E10 | 23E1 | 10C4 | 38B9 | 43E7 | 53D5 | 54B9 | 73B4 | 1A12 |
| -12 | | | **A ⇒ T** | | A → T | A → T | A → T | A → T | A → T | A → T | A → T |
| 87 | **C ⇒ T** | C → T | C → T | C → T | C → T | C → T | C → T | C → T | C → T | C → T | C → T |
| 102 | | | | | | | | | **T ⇒ C** | | |
| 181 | | | | | | **A ⇒ G** | | | | **A ⇒ G** | A → G |
| 255 | | | | | | | | **T ⇒ C** | | | |
| 283 | | | | | | | **A ⇒ G** | | | | |
| 290 | | **A ⇒ G** | | | | | | | | | |
| 291 | | | | | **T ⇒ C** | T → C | T → C | T → C | T → C | T → C | T → C |
| 333 | | | **T ⇒ A** | | T → A | T → A | T → A | T → A | T → A | T → A | T → A |
| 399 | | | | **T ⇒ C** | | | | | | | |
| 433 | | | | | **T ⇒ A** | T → A | T → A | T → A | T → A | T → A | T → A |
| 720 | | | | | | | | **T ⇒ C** | | | |
| 803 | | | | | | | | | **A ⇒ G** | | |
| 814 | | | | **T ⇒ C** | | | | | | | |
| 968 | **A ⇒ G** | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G |
| 1003 | | | | | | | **T ⇒ G** | | | | |
| 1004 | | | | | | | | | **T ⇒ C** | | T → C |
| 1031 | | | | **C ⇒ T** | | | | **C ⇒ T** | | | |
| 1075 | | **A ⇒ G** | **A ⇒ G** | | A → G | A → G | A → G | A → G | A → G | A → G | A → G |
| 1112 | **A ⇒ T** | A → T | A → T | A → T | A → T | A → T | A → T | A → T | A → T | A → T | A → T |
| 1122 | | **A ⇒ G** | | | | | | | | | |
| 1239 | | | **A ⇒ G** | | A → G | A → G | A → G | A → G | A → G | A → G | A → G |
| 1302 | | **A ⇒ G** | | | | | | | | | |
| 1485 | **A ⇒ G** | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G |
| 1568 | | | | | | | | | | **T ⇒ A** | |
| 1618 | | | | **T ⇒ C** | | | | | | | |
| 1658 | **A ⇒ G** | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G | A → G |
| 1678 | **T ⇒ C** | T → C | T → C | T → C | T → C | T → C | T → C | T → C | T → C | T → C | T → C |
| 1745 | | | **C ⇒ T** | | C → T | C → T | C → T | C → T | C → T | C → T | C → T |

Table 2.3: DNA mutations found in evolved pNB esterases. Bold-faced horizontal lines mark the beginning and end of the open reading frame (bp 1 and 1470). Mutations indicated in reverse type are those new to that generation. Normal type indicates mutations accumulated in previous generations.

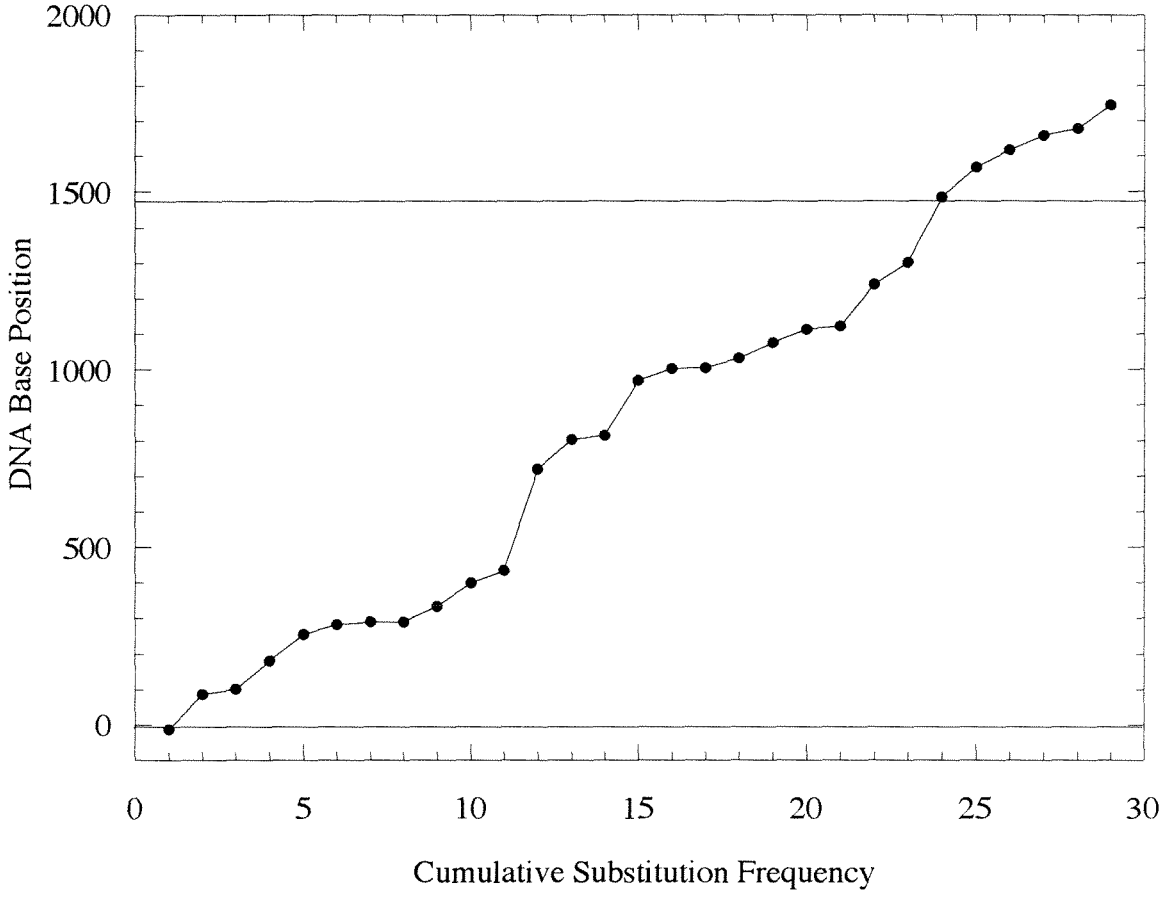



Figure 2.18: The location of DNA substitutions reported in table 3 are shown plotted against their cumulative frequency. Horizontal lines mark the open reading frame.

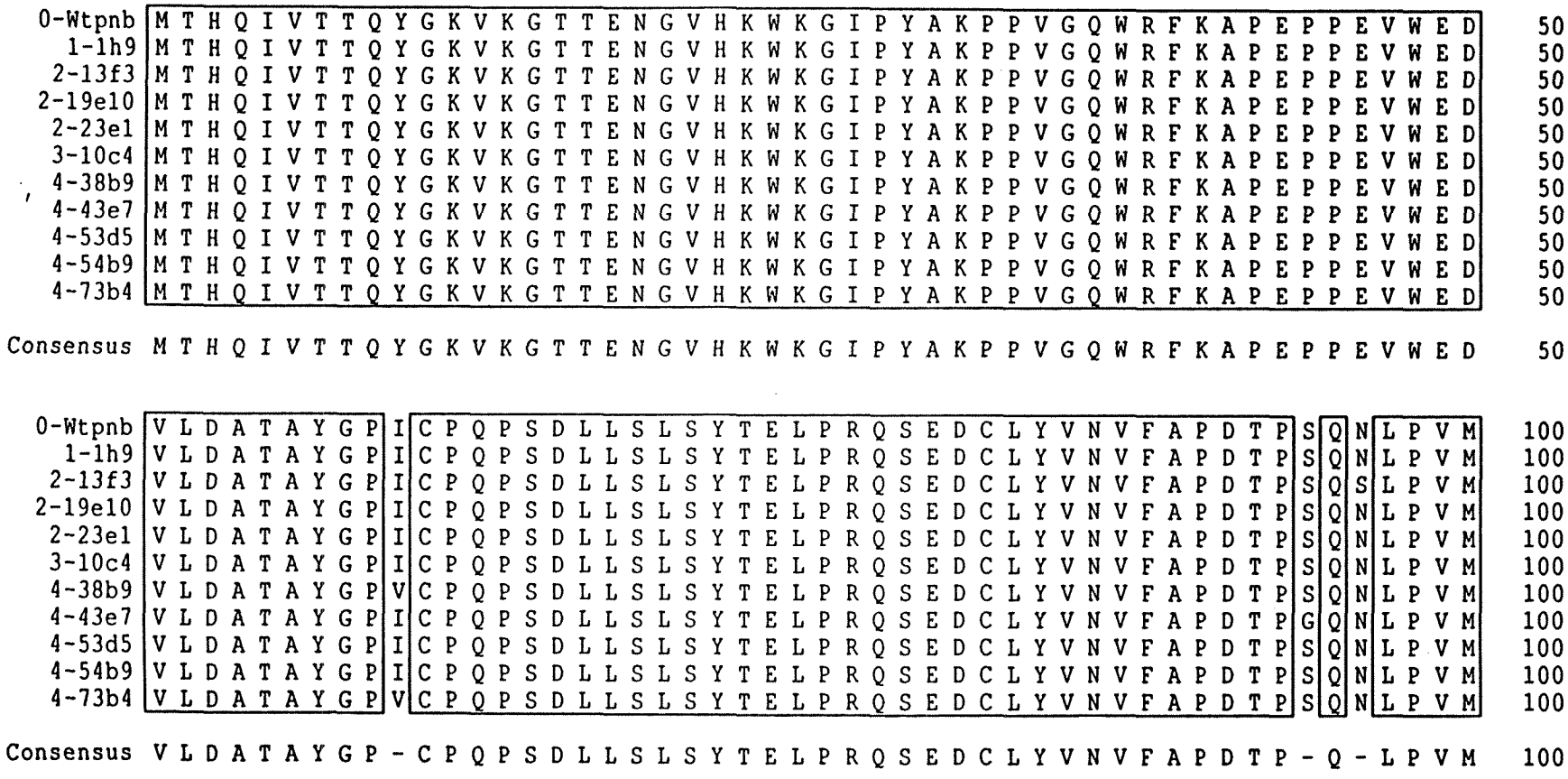
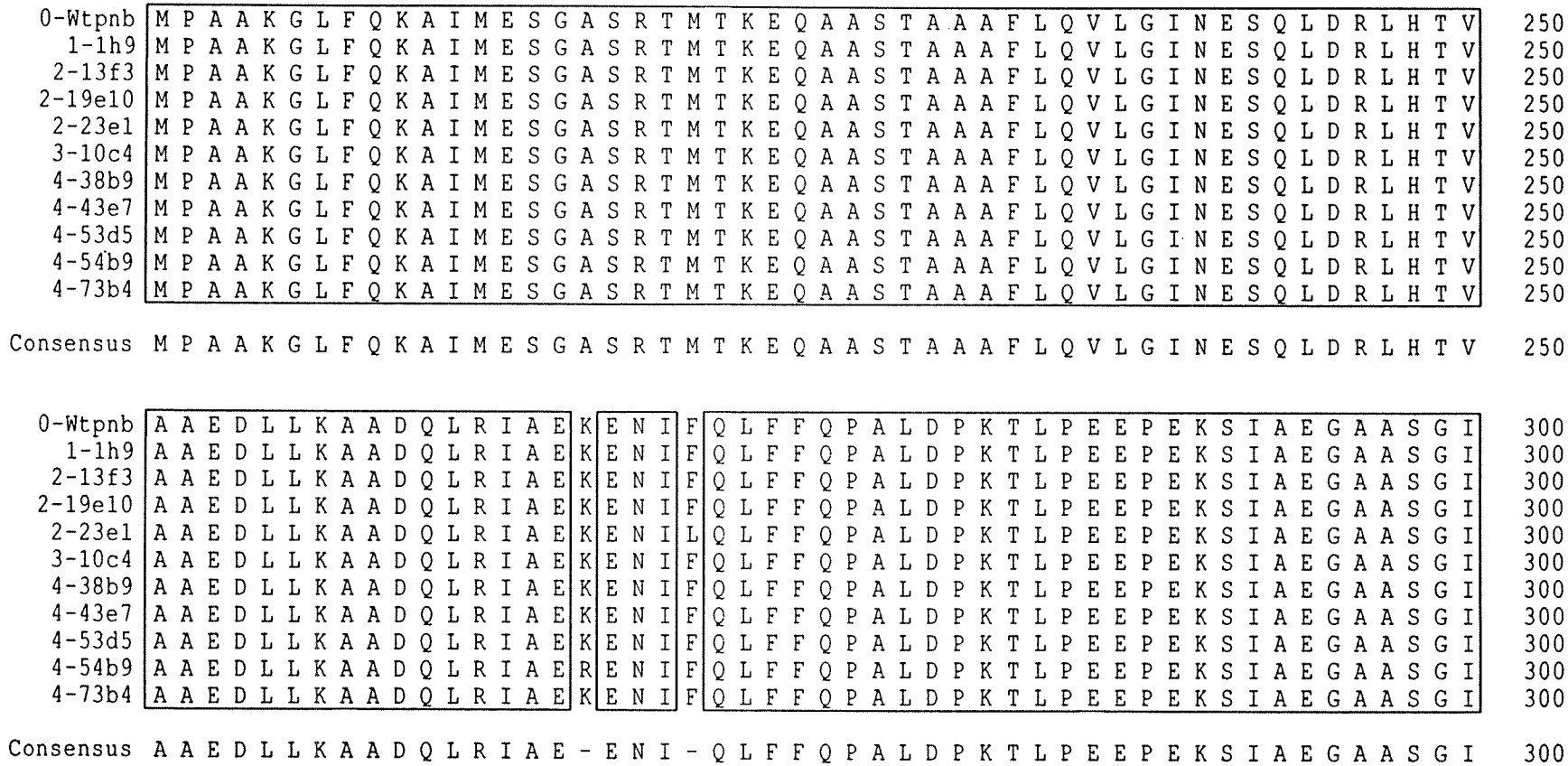
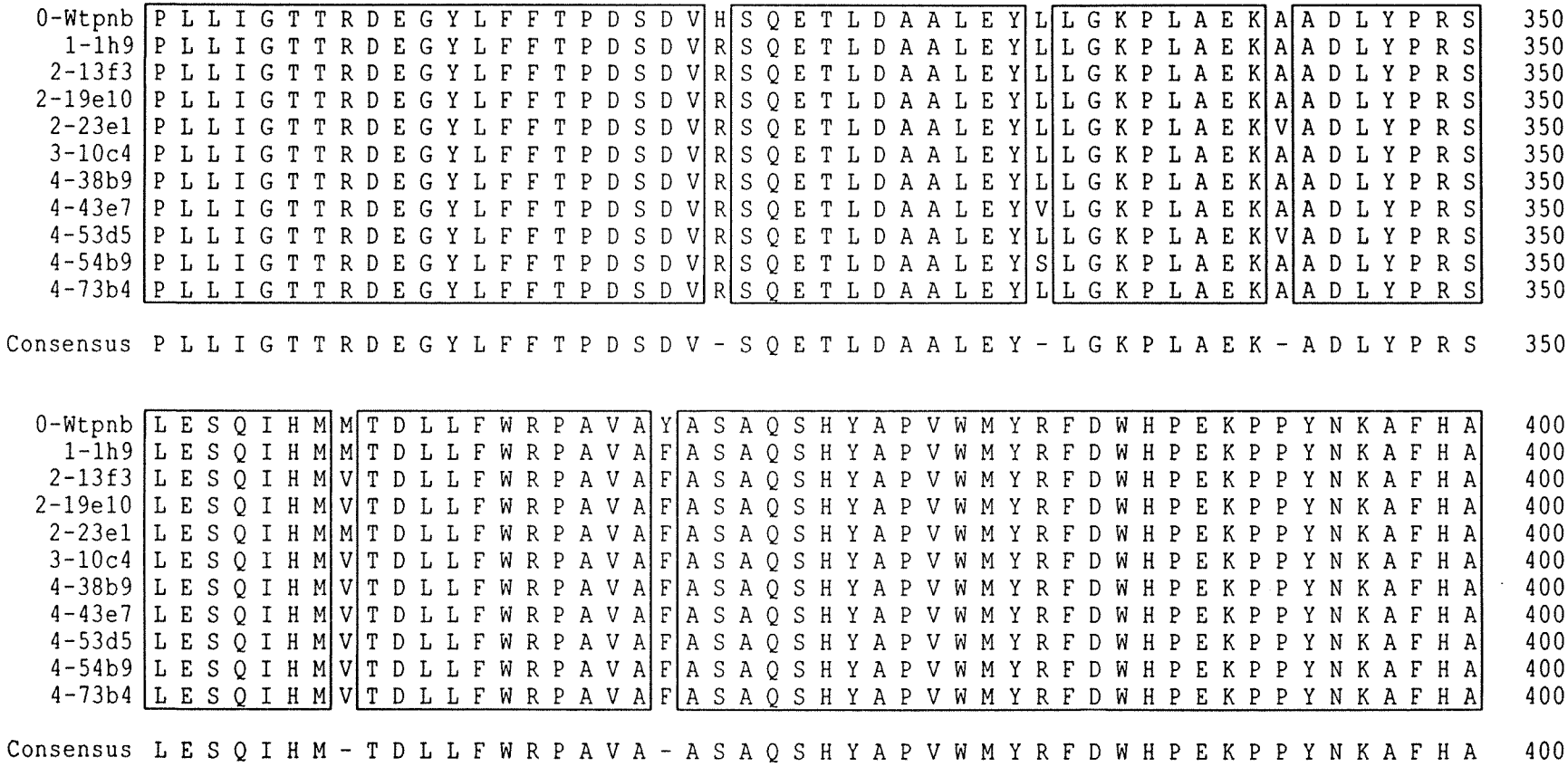



Figure 2.19: Amino acid sequence alignment of all enzyme variants sequenced in this study. The variants are listed from top to bottom by generation. Boxed regions indicate amino acid sequence regions where all variants are identical. The columns of amino acids not boxed are those where at least one mutation has occurred. These locations are further indicated by dashes in the consensus sequence. This figure was generated using PILEUP and PRETTYPLOT from the GCG software package.

```
0-Wtpnb  V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F L H L S S F D  150
1-1h9    V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F L H L S S F D  150
2-13f3   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F L H L S S F D  150
2-19e10  V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F L H L S S F D  150
2-23e1   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F L H L S S F D  150
3-10c4   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F M H L S S F D  150
4-38b9   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F M H L S S F D  150
4-43e7   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F M H L S S F D  150
4-53d5   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F M H L S S F D  150
4-54b9   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F M H L S S F D  150
4-73b4   V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F M H L S S F D  150

Consensus V W I H G G A F Y L G A G S E P L Y D G S K L A A Q G E V I V V T L N Y R L G P F G F - H L S S F D  150


0-Wtpnb  E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
1-1h9    E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
2-13f3   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
2-19e10  E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
2-23e1   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
3-10c4   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
4-38b9   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
4-43e7   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
4-53d5   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
4-54b9   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
4-73b4   E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200

Consensus E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V T V F G E S A G G M S I A A L L A  200
```


0-Wtpnb M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
1-1h9 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
2-13f3 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
2-19e10 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
2-23e1 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
3-10c4 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
4-38b9 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
4-43e7 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
4-53d5 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
4-54b9 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
4-73b4 M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
Consensus M P A A K G L F Q K A I M E S G A S R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V 250
0-Wtpnb A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
1-1h9 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
2-13f3 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
2-19e10 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
2-23e1 A A E D L L K A A D Q L R I A E K E N I L Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
3-10c4 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
4-38b9 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
4-43e7 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
4-53d5 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
4-54b9 A A E D L L K A A D Q L R I A E R E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
4-73b4 A A E D L L K A A D Q L R I A E K E N I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300
Consensus A A E D L L K A A D Q L R I A E - E N I - Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I 300


0-Wtpnb P L L I G T T R D E G Y L F F T P D S D V H S Q E T L D A A L E Y L L G K P L A E K A A D L Y P R S 350
1-1h9 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K A A D L Y P R S 350
2-13f3 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K A A D L Y P R S 350
2-19e10 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K A A D L Y P R S 350
2-23e1 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K V A D L Y P R S 350
3-10c4 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K A A D L Y P R S 350
4-38b9 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K A A D L Y P R S 350
4-43e7 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y V L G K P L A E K A A D L Y P R S 350
4-53d5 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K V A D L Y P R S 350
4-54b9 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y S L G K P L A E K A A D L Y P R S 350
4-73b4 P L L I G T T R D E G Y L F F T P D S D V R S Q E T L D A A L E Y L L G K P L A E K A A D L Y P R S 350
Consensus P L L I G T T R D E G Y L F F T P D S D V - S Q E T L D A A L E Y - L G K P L A E K - A D L Y P R S 350
0-Wtpnb L E S Q I H M M T D L L F W R P A V A Y A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
1-1h9 L E S Q I H M M T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
2-13f3 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
2-19e10 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
2-23e1 L E S Q I H M M T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
3-10c4 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
4-38b9 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
4-43e7 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
4-53d5 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
4-54b9 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
4-73b4 L E S Q I H M V T D L L F W R P A V A F A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400
Consensus L E S Q I H M - T D L L F W R P A V A - A S A Q S H Y A P V W M Y R F D W H P E K P P Y N K A F H A 400

```
0-Wtpnb  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
  1-1h9  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 2-13f3  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
2-19e10  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 2-23e1  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 3-10c4  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 4-38b9  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 4-43e7  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 4-53d5  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 4-54b9  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450
 4-73b4  L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450

Consensus L E L P F V F G N L D G L E R M A K A E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V  450


0-Wtpnb  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
  1-1h9  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 2-13f3  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
2-19e10  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 2-23e1  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 3-10c4  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 4-38b9  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 4-43e7  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 4-53d5  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 4-54b9  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489
 4-73b4  N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  489

Consensus N W P A Y H E E T R E T V I L D S E I T I E N D P E S E K R Q K L F P S K G E -  490
```

indicated in the consensus sequence at the bottom of each set of rows. Figure 20 is an "evolution tree" summarizing the amino acid substitutions and positions resulting from the amino acid sequence information in Figure 19. Only the changes listed in this tree have the ability to alter the properties of pNB esterase. Because several of the DNA substitutions lie outside the open reading frame, and even more DNA substitutions do not result in an amino acid change, the number of amino acid substitutions is far smaller than the number of DNA substitutions. For example, the variant isolated from the first round of mutagenesis and screening, 1-1H9, contains six DNA substitutions, of which only three lie within the pNB esterase open reading frame. Of the three remaining substitutions, two lead to amino acid changes and one does not (a silent mutation). This mutated gene served as the parent sequence for the second generation. All three clones sequenced from the second generation contain the substitutions observed in the 1-1H9 sequence, as well as a few new additional substitutions. This can be seen at the DNA level in Table 3 and at the protein level in Figure 20. Two of the second generation clones contain the same DNA substitution, an A to G substitution at position 1075, which gives rise to the substitution of methionine at position 358 by valine. This residue must be responsible for the increased specific activity of these two clones, as it is the only non-silent mutation in clone 2-19E10, and its catalytic activity is slightly better than 2-13F3. 2-19E10 was chosen to parent the third round of mutagenesis. One silent and one translated substitution are added to the evolutionary sequence in 3-10C4; a T to A substitution at position 433 gives rise to leucine 144 substituted by methionine. The remaining five sequences are progeny of 3-10C4 from the fourth generation. All have the eight DNA base substitutions of 3-10C4 in common as the cumulative result of the three previous rounds of evolution. 4-38B9 and 4-73B4 each contain an identical open reading frame substitution, an A to G change at position 181, leading to isoleucine at position 60 changed to valine. As a result, these two variants are listed together in Figure 20. 4-43E7 and 4-54B9 each contain two substitutions; one

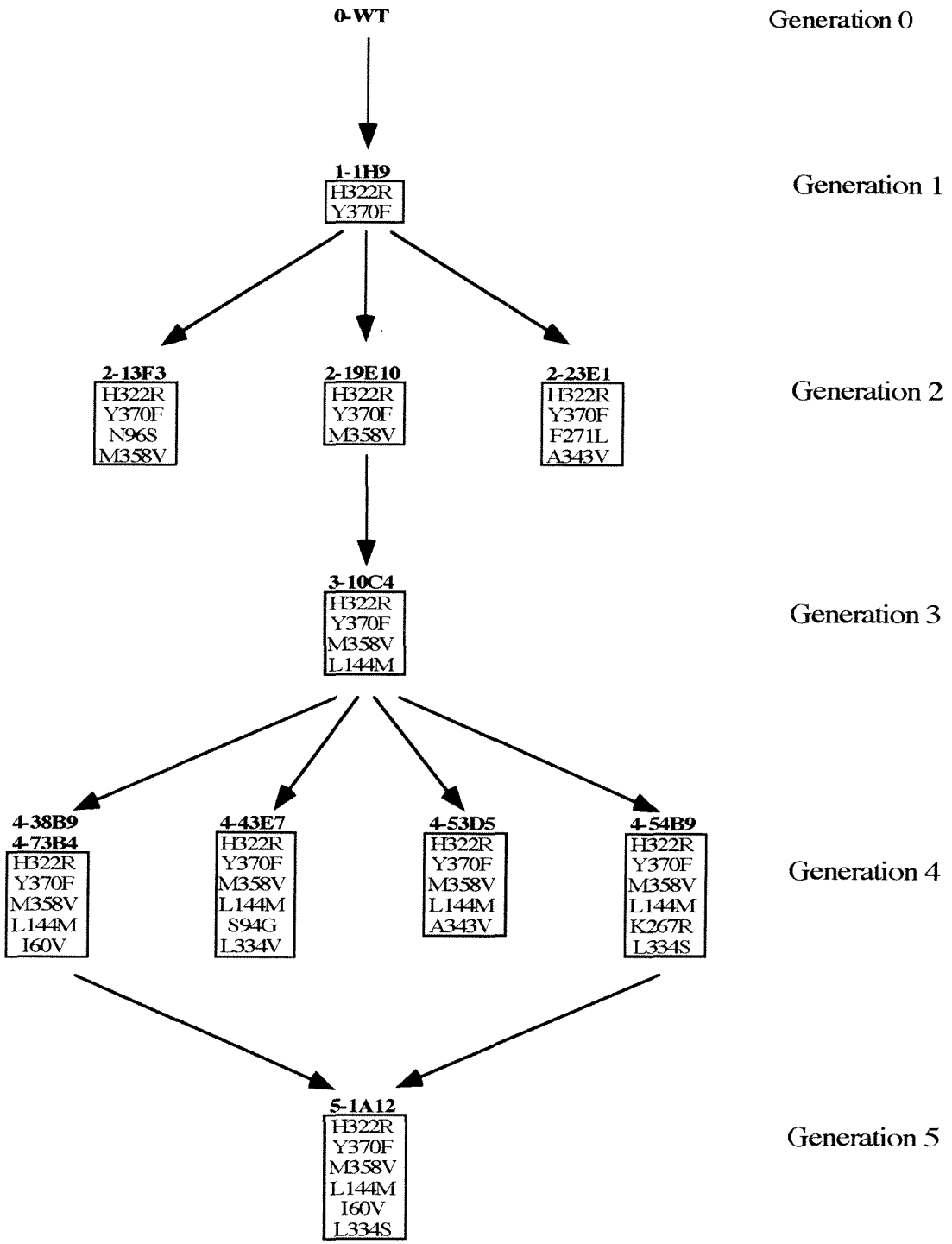



Figure 2.20: Tree showing amino acid substitution progression in pNB esterase variants. The esterase variants are labeled in bold-type and boxed beneath each label are the substitutions present in the variant. The amino acid substitutions were determined by translation of DNA sequencing results.

occurs within identical codons in the DNA sequence, giving rise to changes in leucine 334 to valine in 4-43E7 and serine in 4-54B9.

Re-examining the location versus frequency of substitution plot in Figure 18 with the translated substitutions highlighted, the locations of substitutions resulting in amino acid changes appears distinctly non-random (Figure 21). Fully half of the translated substitutions DNA mutations lie within a 144 base pair stretch of DNA (less than 10% of the open reading frame). Presumably this is due to the fact that a non-random selection of enzymes were chosen for sequencing (only those exhibiting improved activity). This suggests that this region of the amino acid sequence is important for substrate recognition and enzyme activity.

*Structural Modeling*

All of the enzymes sharing significant homology with pNB esterase belong to the $\alpha/\beta$ hydrolase enzyme family and serve to place pNB esterase also in this family. This family contains proteins organized around a large $\alpha$-helix/$\beta$-sheet structure with much of the variability in structure between family members located in the loops that connect the $\alpha$-helices and $\beta$-strands (Ollis, 1992). A sequence alignment of all the enzymes listed in Table 1 is presented in Figure 22, where the identical and similar residues are boxed and indicated in the consensus sequence. Aligning the sequence of pNB esterase with these members of the $\alpha/\beta$ hydrolase enzyme family shows several features (Cygler, 1993). First, most members of this family are serine esterases with a catalytic triad similar to serine proteases. The catalytic triad consists of serine, histidine and glutamate in these serine esterases, while the glutamate is replaced by aspartate in the serine proteases. According to the sequence alignment (Figure 22), pNB esterase serine 189 aligns with active site serines, glutamate 300 with active site glutamates and histidine 399 with active site histidines. This suggests that the mechanism for pNB ester hydrolysis proceeds through an acyl enzyme intermediate common to many of the serine protease and esterase enzymes (Taylor and

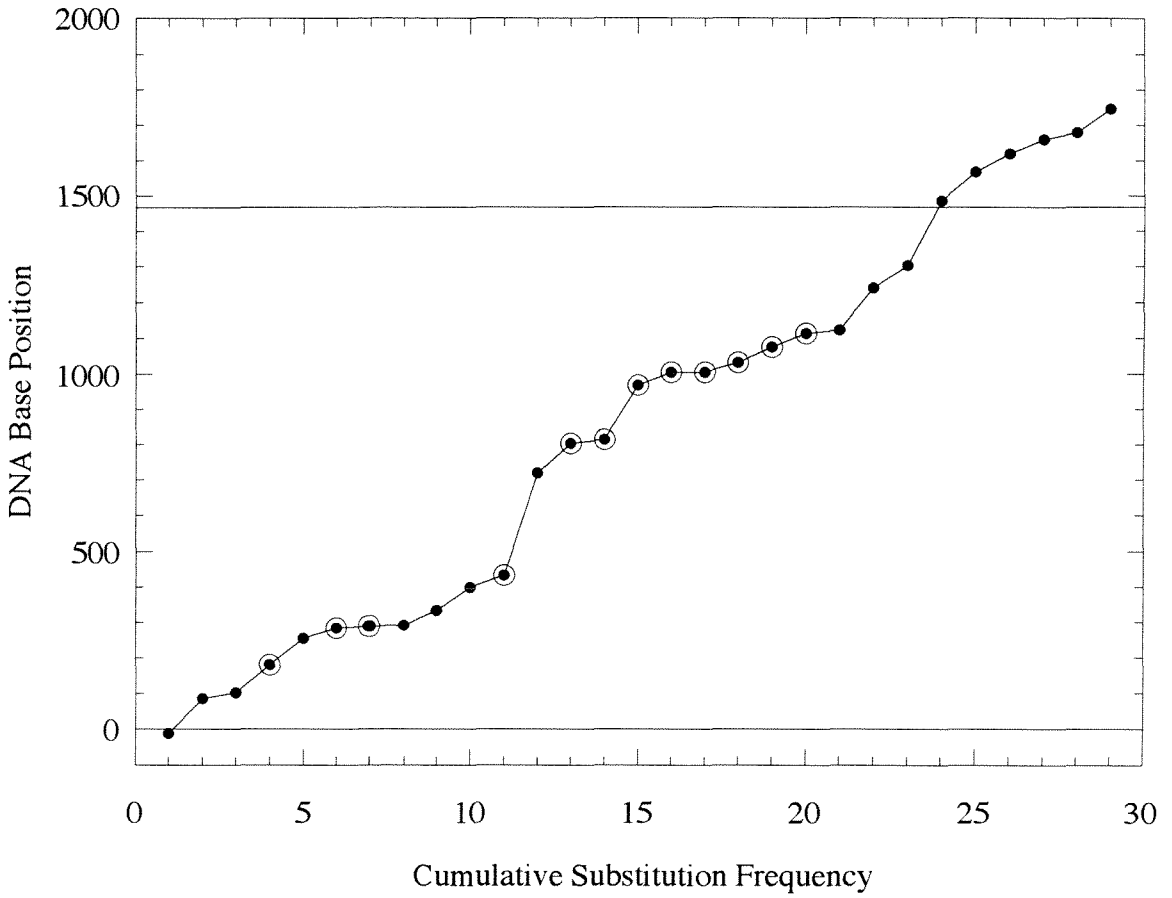
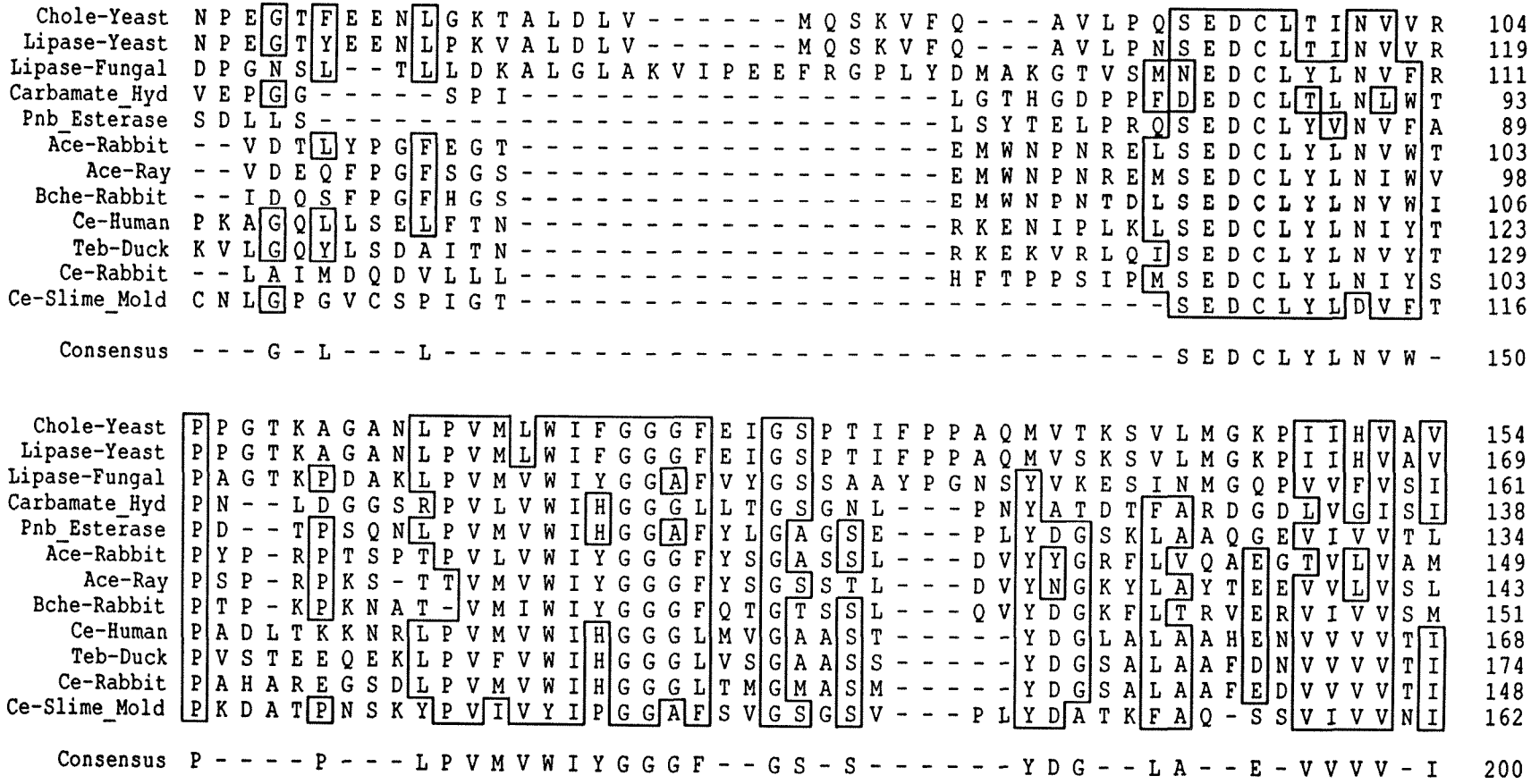
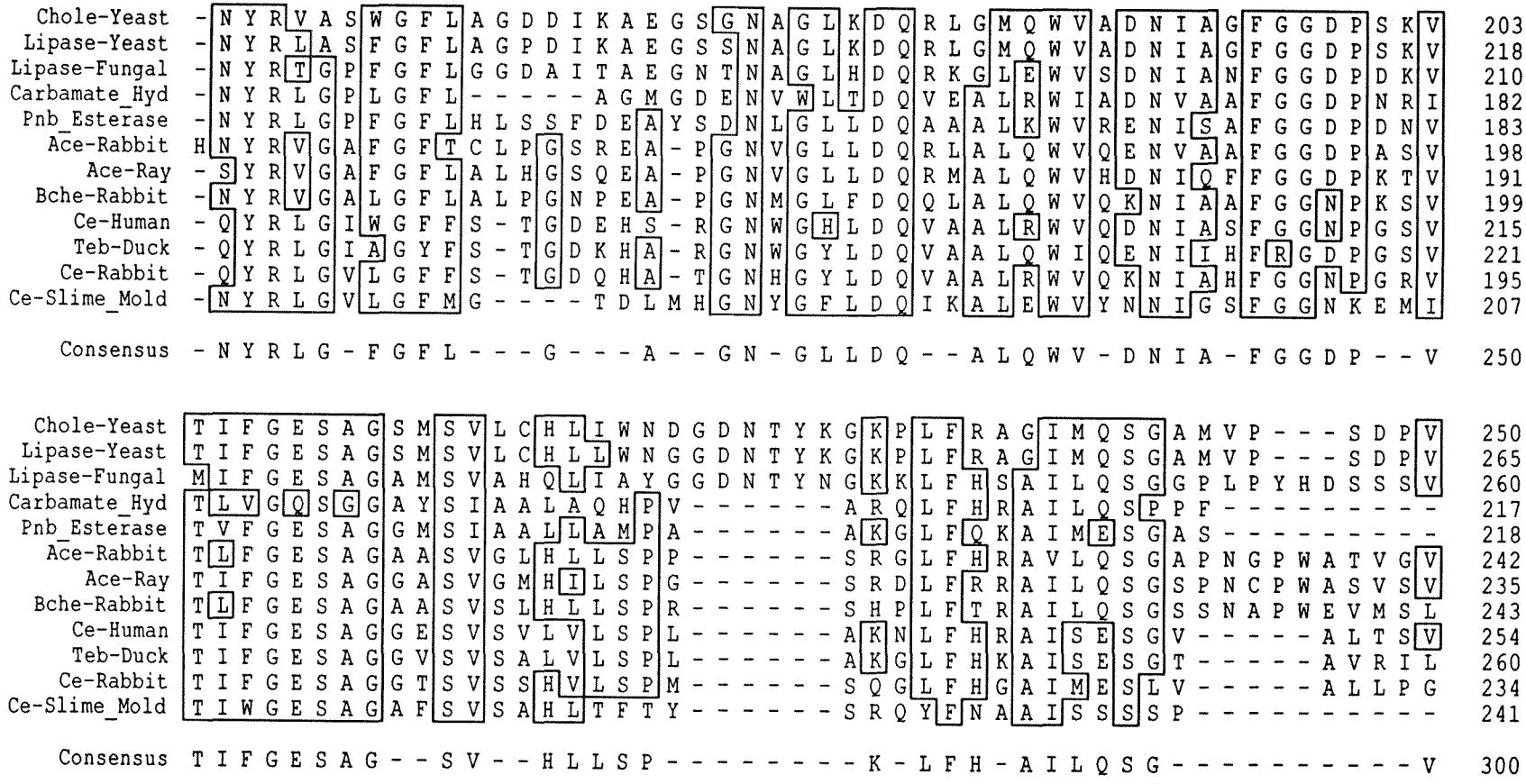



Figure 2.21: The location of DNA substitutions reported in Table 3 are shown plotted against their cumulative frequency. The data circled are the DNA substitutions which lead to amino acid changes. Horizontal lines demark the boundaries of the open reading frame.

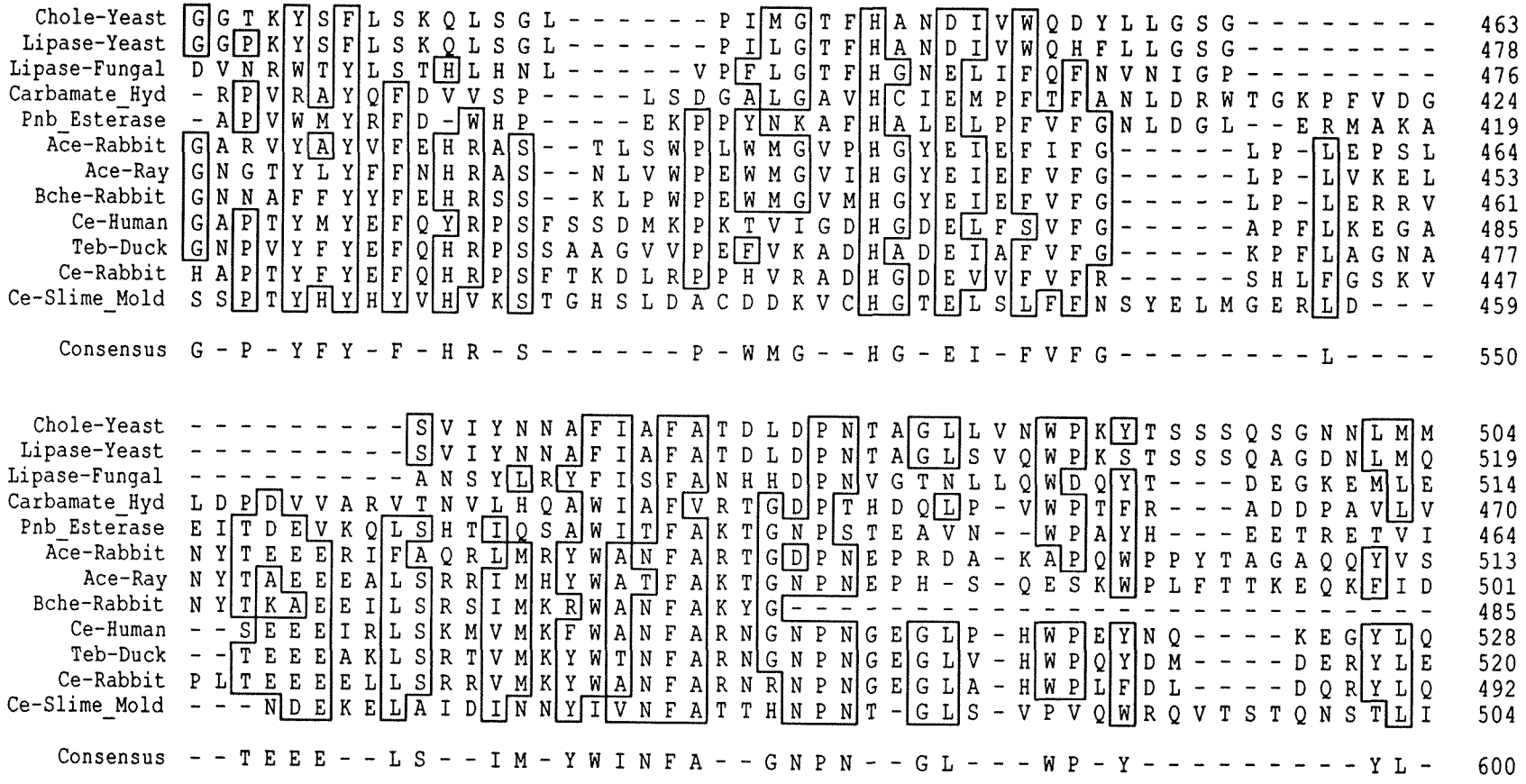

Figure 2.22: Sequence alignment of pNB esterase and all enzymes listed in Table 3. Boxed regions highlight identical and highly conserved residues between the proteins. Dashes indicate gaps in the sequence alignment where residues in one sequence have presumably been removed in the second sequence. The numbers on the right indicate the amino acid position of the last amino acid in the row, and the consensus sequence is the residue represented in more than half of the sequences, if any.

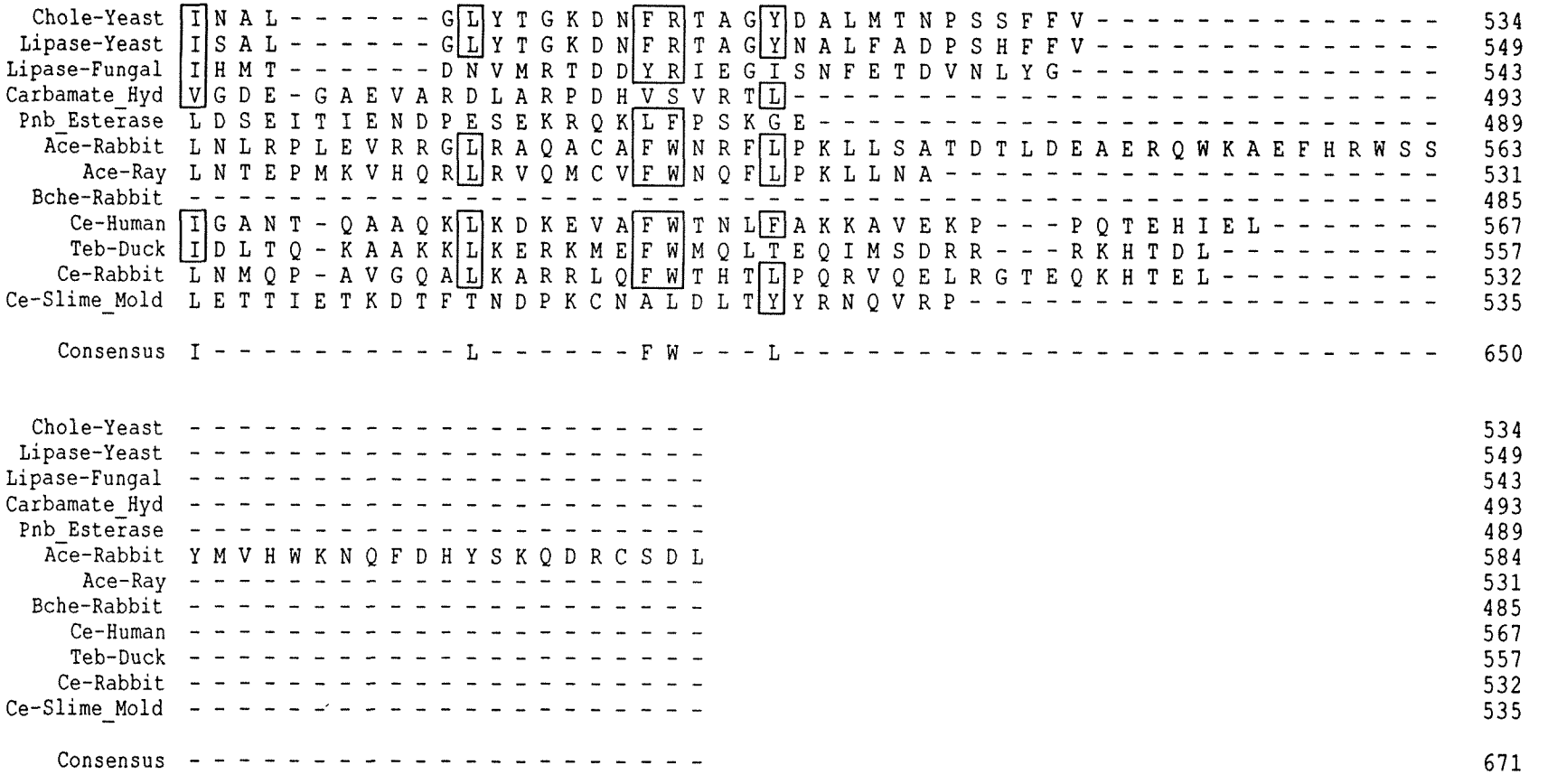

Chole-Yeast N P E G T F E E N L G K T A L D L V - - - - - - M Q S K V F Q - - - A V L P Q S E D C L T I N V V R 104
Lipase-Yeast N P E G T Y E E N L P K V A L D L V - - - - - - M Q S K V F Q - - - A V L P N S E D C L T I N V V R 119
Lipase-Fungal D P G N S L - - T L L D K A L G L A K V I P E E F R G P L Y D M A K G T V S M N E D C L Y L N V F R 111
Carbamate_Hyd V E P G G - - - - - S P I - - - - - - - - - - - - - - - - - L G T H G D P P F D E D C L T L N L W T 93
Pnb_Esterase S D L L S - - - - - - - - - - - - - - - - - - - - - - - - - L S Y T E L P R Q S E D C L Y V N V F A 89
Ace-Rabbit - - V D T L Y P G F E G T - - - - - - - - - - - - - - - - - - E M W N P N R E L S E D C L Y L N V W T 103
Ace-Ray - - V D E Q F P G F S G S - - - - - - - - - - - - - - - - - - E M W N P N R E M S E D C L Y L N I W V 98
Bche-Rabbit - - I D Q S F P G F H G S - - - - - - - - - - - - - - - - - E M W N P N T D L S E D C L Y L N V W I 106
Ce-Human P K A G Q L L S E L F T N - - - - - - - - - - - - - - - - - - R K E N I P L K L S E D C L Y L N I Y T 123
Teb-Duck K V L G Q Y L S D A I T N - - - - - - - - - - - - - - - - - - R K E K V R L Q I S E D C L Y L N V Y T 129
Ce-Rabbit - - L A I M D Q D V L L L - - - - - - - - - - - - - - - - - - H F T P P S I P M S E D C L Y L N I Y S 103
Ce-Slime_Mold C N L G P G V C S P I G T - - - - - - - - - - - - - - - - - - - - - - - - - - - - S E D C L Y L D V F T 116
Consensus - - - G - L - - - L - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - S E D C L Y L N V W - 150
Chole-Yeast P P G T K A G A N L P V M L W I F G G G F E I G S P T I F P P A Q M V T K S V L M G K P I I H V A V 154
Lipase-Yeast P P G T K A G A N L P V M L W I F G G G F E I G S P T I F P P A Q M V S K S V L M G K P I I H V A V 169
Lipase-Fungal P A G T K P D A K L P V M V W I Y G G A F V Y G S S A A Y P G N S Y V K E S I N M G Q P V V F V S I 161
Carbamate_Hyd P N - - L D G G S R P V L V W I H G G G L L T G S G N L - - - P N Y A T D T F A R D G D L V G I S I 138
Pnb_Esterase P D - - T P S Q N L P V M V W I H G G A F Y L G A G S E - - - P L Y D G S K L A A Q G E V I V V T L 134
Ace-Rabbit P Y P - R P T S P T P V L V W I Y G G G F Y S G A S S L - - - D V Y Y G R F L V Q A E G T V L V A M 149
Ace-Ray P S P - R P K S - T T V M V W I Y G G G F Y S G S S T L - - - D V Y N G K Y L A Y T E E V V L V S L 143
Bche-Rabbit P T P - K P K N A T - V M I W I Y G G G F Q T G T S S L - - - Q V Y D G K F L T R V E R V I V V S M 151
Ce-Human P A D L T K K N R L P V M V W I H G G G L M V G A A S T - - - - - Y D G L A L A A H E N V V V V T I 168
Teb-Duck P V S T E E Q E K L P V F V W I H G G G L V S G A A S S - - - - - Y D G S A L A A F D N V V V V T I 174
Ce-Rabbit P A H A R E G S D L P V M V W I H G G G L T M G M A S M - - - - - Y D G S A L A A F E D V V V V T I 148
Ce-Slime_Mold P K D A T P N S K Y P V I V Y I P G G A F S V G S G S V - - - P L Y D A T K F A Q - S S V I V V N I 162
Consensus P - - - - P - - - L P V M V W I Y G G G F - - G S - S - - - - - - Y D G - - L A - - E - V V V V - I 200


Chole-Yeast - N Y R V A S W G F L A G D D I K A E G S G N A G L K D Q R L G M Q W V A D N I A G F G G D P S K V 203
Lipase-Yeast - N Y R L A S F G F L A G P D I K A E G S S N A G L K D Q R L G M Q W V A D N I A G F G G D P S K V 218
Lipase-Fungal - N Y R T G P F G F L G G D A I T A E G N T N A G L H D Q R K G L E W V S D N I A N F G G D P D K V 210
Carbamate_Hyd - N Y R L G P L G F L - - - - - A G M G D E N V W L T D Q V E A L R W I A D N V A A F G G D P N R I 182
Pnb_Esterase - N Y R L G P F G F L H L S S F D E A Y S D N L G L L D Q A A A L K W V R E N I S A F G G D P D N V 183
Ace-Rabbit H N Y R V G A F G F T C L P G S R E A - P G N V G L L D Q R L A L Q W V Q E N V A A F G G D P A S V 198
Ace-Ray - S Y R V G A F G F L A L H G S Q E A - P G N V G L L D Q R M A L Q W V H D N I Q F F G G D P K T V 191
Bche-Rabbit - N Y R V G A L G F L A L P G N P E A - P G N M G L F D Q Q L A L Q W V Q K N I A A F G G N P K S V 199
Ce-Human - Q Y R L G I W G F F S - T G D E H S - R G N W G H L D Q V A A L R W V Q D N I A S F G G N P G S V 215
Teb-Duck - Q Y R L G I A G Y F S - T G D K H A - R G N W G Y L D Q V A A L Q W I Q E N I I H F R G D P G S V 221
Ce-Rabbit - Q Y R L G V L G F F S - T G D Q H A - T G N H G Y L D Q V A A L R W V Q K N I A H F G G N P G R V 195
Ce-Slime_Mold - N Y R L G V L G F M G - - - - T D L M H G N Y G F L D Q I K A L E W V Y N N I G S F G G N K E M I 207
Consensus - N Y R L G - F G F L - - - G - - - A - - G N - G L L D Q - - A L Q W V - D N I A - F G G D P - - V 250
Chole-Yeast T I F G E S A G S M S V L C H L I W N D G D N T Y K G K P L F R A G I M Q S G A M V P - - - S D P V 250
Lipase-Yeast T I F G E S A G S M S V L C H L L W N G G D N T Y K G K P L F R A G I M Q S G A M V P - - - S D P V 265
Lipase-Fungal M I F G E S A G A M S V A H Q L I A Y G G D N T Y N G K K L F H S A I L Q S G G P L P Y H D S S S V 260
Carbamate_Hyd T L V G Q S G G A Y S I A A L A Q H P V - - - - - - - A R Q L F H R A I L Q S P P F - - - - - - - - - - 217
Pnb_Esterase T V F G E S A G G M S I A A L L A M P A - - - - - - - A K G L F Q K A I M E S G A S - - - - - - - - - - 218
Ace-Rabbit T L F G E S A G A A S V G L H L L S P P - - - - - - - S R G L F H R A V L Q S G A P N G P W A T V G V 242
Ace-Ray T I F G E S A G G A S V G M H I L S P G - - - - - - - S R D L F R R A I L Q S G S P N C P W A S V S V 235
Bche-Rabbit T L F G E S A G A A S V S L H L L S P R - - - - - - - S H P L F T R A I L Q S G S S N A P W E V M S L 243
Ce-Human T I F G E S A G G E S V S V L V L S P L - - - - - - - A K N L F H R A I S E S G V - - - - - - A L T S V 254
Teb-Duck T I F G E S A G G V S V S A L V L S P L - - - - - - - A K G L F H K A I S E S G T - - - - - - A V R I L 260
Ce-Rabbit T I F G E S A G G T S V S S H V L S P M - - - - - - - S Q G L F H G A I M E S L V - - - - - - A L L P G 234
Ce-Slime_Mold T I W G E S A G A F S V S A H L T F T Y - - - - - - - S R Q Y F N A A I S S S S P - - - - - - - - - - 241
Consensus T I F G E S A G - - S V - - H L L S P - - - - - - - - - K - L F H - A I L Q S G - - - - - - - - - - V 300


Chole-Yeast - - - - D G T Y G N E I Y D L F V S S A G C G S A S D K L A C L R S A S S D T L L D A T N - - N T P 294
Lipase-Yeast - - - - D G T Y G T Q I Y D T L V A S T G C S S A S N K L A C L R G L S T Q A L L D A T N - - D T P 309
Lipase-Fungal - - - - G P D I S Y N R F A Q Y A G C D T S A S A N D T L E C L R S K S S S V L H D A Q N S Y D L K 306
Carbamate_Hyd - - - G M Q P H T V E E S T A R T K A L A R H L G H D D I E A L R H E P W E R L I Q G T I G V L M E 264
Pnb_Esterase - - - R T M T K E Q A A S T A A A F L Q V L G I N E S Q L D R L H T V A A E D L L K A A - D Q L R I 264
Ace-Rabbit G E A R R R A T L L A R L V V C P P G G A G G N D T E L V A C L R T R P A Q D L V D H E W R V L P Q 292
Ace-Ray A E G R R R A V E L G R N L N C N L N S - - - - D E E L I H C L R E K K P Q E L I D V E W N V L P F 281
Bche-Rabbit H E A R N R T L T L A K F V G C S T E - - - - N E T E I I K C L R N K D A Q E I L L N E V F V V P F 289
Ce-Human L V K K G D V K P L A E Q I A I T A G C K T T T S A V M V H C L R Q K T E E E L L E T T L K M K F L 304
Teb-Duck F T E Q P E E Q - - A Q R I A A A A G C E K S S S A A L V E C L R E K T E A E M E Q I T L K M P P M 308
Ce-Rabbit L I T S S S E - V V S T V V A N L S R C G Q V D S E T L V R C L R A K S E E E M L A I T Q V F M - - 281
Ce-Slime_Mold L T V G L K D K T T A R G N A N R F A T N V G C N I E D L T C L R G K S M D E I L D A Q E K V - - - 288
Consensus - - - - - - - - - - - A - - - A - - - - - - - - - - - - E - V - C L R - K - - E E L L D - T - - - - - - 350
Chole-Yeast G F - - - L A Y S S L R L S Y L P R P D G K N I T D D M Y K L V R D G K Y A S V P V I I G D Q N D E 341
Lipase-Yeast G F - - - L S Y T S L R L S Y L P R P D G A N I T D D M Y K L V R D G K Y A S V P V I I G D Q N D E 356
Lipase-Fungal D L - - - F G L L P Q F L G F G P R P D G N I I P D A A Y E L F R S G R Y A K V P Y I S G N Q E D E 353
Carbamate_Hyd H T K F G - - - - E W P L A F Y P V F D E A T I P R H P I E S I I D - - - S D I E I I I G W T R D E 307
Pnb_Esterase A E K E N - - - - I F Q L F F Q P A L D P K T L P E E P E K S I A E G A A S G I P L L I G T T R D E 310
Ace-Rabbit E S I F R F S - - - - - - - - F V P V V D G D F L S D T P E A L I N A G D F Q G L Q V L V G V V K D E 335
Ace-Ray D S I F R F S - - - - - - - - F V P V I D G E F F P T S L E S M L N S G N F K K T Q I L L G V N K D E 324
Bche-Rabbit D S L L S V N - - - - - - - - F G P T V D G D F L T D M P D T L L Q L G Q L K K T Q I L V G V N K D E 332
Ce-Human S L D L Q G D P R E S Q P L L G T V I D G M L L L K T P E E L Q A E R N F H T V P Y M V G I N K Q E 354
Teb-Duck - - - - - - - - - - - - - - - - F I S A S L D G V F F P K S P R Q L L S E K V I N A V P Y I I G V N N C E 345
Ce-Rabbit - - - - - - - - - - - - - - - - L I P G V V D G V F L P R H P E E L L A L A D F Q P V P S I I G I N N D E 318
Ce-Slime_Mold G L T F G D K I L D A F T I W S P V I D G D I I P M Q P L T A V K E G K T Y D V P T I I G N V K H E 338
Consensus - - - F - - - - - - - - - - - F - P V - D G - F L P D - P E - L - - E G - F - - V P - I I G V - - D E 400


Chole-Yeast GTLFGLSSL----------NVTTNAQARAYFKQSFIHASDAEIDTLMAA 380
Lipase-Yeast GFLFGLSSL----------NTTTEADAEAYLRKSFIHATDADITALKAA 395
Lipase-Fungal GTAFAPVAL----------NATTTPHVKKWLQYIFYDASEASIDRVLSL 392
Carbamate_Hyd GTFPF---------AFDPQV-SQADRD-QVESW------LQKRFGDHAASA 341
Pnb_Esterase GYLFF---------TPDSDVHSQETLDAALEYL------LGKPLAEKAADL 346
Ace-Rabbit GTYFL---VYGAPGFSKDNESFISRAQFLAGVRVGVPQASDLAAEAVVLH 382
Ace-Ray GSFFL---LYGAPGFSKDSESKISREDFMSGVKLSVPHANDLGLDAVTLQ 371
Bche-Rabbit GTAFL---VYGAPGFSKDNTSIITRKEFQEGLKIFFPGVSEFGKESILFH 379
Ce-Human FGWLIPMQLMSYPLSEGQLDQKTAMSLLWKSYPLV--CIAKELIPEATEK 402
Teb-Duck FGWILP-RMMKFPEFTEGLEKDVARQVLQSTLALSFKGAPSDIVDLVYNE 394
Ce-Rabbit YGWIIPKLLLAI-DPQEERDRQAMREIMHQATKQLM--LPPALGDLLMDE 365
Ce-Slime_Mold AIPFIYSFFQDSVGID----------YYRVLVAIVFPLNAMKILPL 374
Consensus G--FL---L----------R--F---L-------A-D---D-V--- 450
Chole-Yeast YPQDITQGSPFDTGIFNAITPQFKRISAVLGDLAFIHARRYFLNHFQ--- 427
Lipase-Yeast YPSDVTQGSPFDTGILNALTPQLKRINAVLGDLTFFTLSRRYFLNHYT--- 442
Lipase-Fungal YPQTLSVGSPFRTGILNALTPQFKRVAAILSDMLFQSPRRVMLSATK--- 439
Carbamate_Hyd YEAHAGDGTS----------PWTVIANV--VGDELFHSAG-YRVADERATR 379
Pnb_Esterase Y----------PRSLESQIHMMTDLLFWRPA-VAYASAQSHY 377
Ace-Rabbit YTDWLHPEDP----------ARLRDALSDVVGDHNVVCPV-AQLAGRLAAQ 422
Ace-Ray YTDWMDDNNG----------IKNRDGLDDIVGDHNVICPL-MHFVNKYTKF 411
Bche-Rabbit YTDWVDEQRP----------ENYREALDDVVGDYNFICPA-LEFTKKFSEW 419
Ce-Human YLG--GTDDT----------VKKKDLFLDLIADVMFGVPS-VIVARNHRDA 440
Teb-Duck YIG--VAENR----------AQVRDGLLDSIADPLFVFS A-VEVARHHRDA 432
Ce-Rabbit YMG--SNEDP----------KHLMAQFQEMMADAMFVMPA-LRVAHLQR-S 402
Ce-Slime_Mold YPAAPRGQDS----------RPILSELITDYLFRCPDRYHTVTNAKKL 412
Consensus Y--------P----------D-L-DVVGD-LF--P----LA------ 500


Chole-Yeast G G T K Y S F L S K Q L S G L - - - - - - P I M G T F H A N D I V W Q D Y L L G S G - - - - - - - - - 463
Lipase-Yeast G G P K Y S F L S K Q L S G L - - - - - - P I L G T F H A N D I V W Q H F L L G S G - - - - - - - - - 478
Lipase-Fungal D V N R W T Y L S T H L H N L - - - - - V P F L G T F H G N E L I F Q F N V N I G P - - - - - - - - - 476
Carbamate_Hyd - R P V R A Y Q F D V V S P - - - - L S D G A L G A V H C I E M P F T F A N L D R W T G K P F V D G 424
Pnb_Esterase - A P V W M Y R F D - W H P - - - - E K P P Y N K A F H A L E L P F V F G N L D G L - - E R M A K A 419
Ace-Rabbit G A R V Y A Y V F E H R A S - - T L S W P L W M G V P H G Y E I E F I F G - - - - - L P - L E P S L 464
Ace-Ray G N G T Y L Y F F N H R A S - - N L V W P E W M G V I H G Y E I E F V F G - - - - - L P - L V K E L 453
Bche-Rabbit G N N A F F Y Y F E H R S S - - K L P W P E W M G V M H G Y E I E F V F G - - - - - L P - L E R R V 461
Ce-Human G A P T Y M Y E F Q Y R P S F S S D M K P K T V I G D H G D E L F S V F G - - - - - A P F L K E G A 485
Teb-Duck G N P V Y F Y E F Q H R P S S A A G V V P E F V K A D H A D E I A F V F G - - - - - K P F L A G N A 477
Ce-Rabbit H A P T Y F Y E F Q H R P S F T K D L R P P H V R A D H G D E V V F V F R - - - - - S H L F G S K V 447
Ce-Slime_Mold S S P T Y H Y H Y V H V K S T G H S L D A C D D K V C H G T E L S L F F N S Y E L M G E R L D - - - 459
Consensus G - P - Y F Y - F - H R - S - - - - - - P - W M G - - H G - E I - F V F G - - - - - - - - - L - - - - 550
Chole-Yeast - - - - - - - - - S V I Y N N A F I A F A T D L D P N T A G L L V N W P K Y T S S S Q S G N N L M M 504
Lipase-Yeast - - - - - - - - - S V I Y N N A F I A F A T D L D P N T A G L S V Q W P K S T S S S Q A G D N L M Q 519
Lipase-Fungal - - - - - - - - - A N S Y L R Y F I S F A N H H D P N V G T N L L Q W D Q Y T - - - D E G K E M L E 514
Carbamate_Hyd L D P D V V A R V T N V L H Q A W I A F V R T G D P T H D Q L P - V W P T F R - - - A D D P A V L V 470
Pnb_Esterase E I T D E V K Q L S H T I Q S A W I T F A K T G N P S T E A V N - - W P A Y H - - - E E T R E T V I 464
Ace-Rabbit N Y T E E E R I F A Q R L M R Y W A N F A R T G D P N E P R D A - K A P Q W P P Y T A G A Q Q Y V S 513
Ace-Ray N Y T A E E E A L S R R I M H Y W A T F A K T G N P N E P H - S - Q E S K W P L F T T K E Q K F I D 501
Bche-Rabbit N Y T K A E E I L S R S I M K R W A N F A K Y G - - - - - - - - - - - - - - - - - - - - - - - - - 485
Ce-Human - - S E E E I R L S K M V M K F W A N F A R N G N P N G E G L P - H W P E Y N Q - - - - K E G Y L Q 528
Teb-Duck - - T E E E A K L S R T V M K Y W T N F A R N G N P N G E G L V - H W P Q Y D M - - - - D E R Y L E 520
Ce-Rabbit P L T E E E E L L S R R V M K Y W A N F A R N R N P N G E G L A - H W P L F D L - - - - D Q R Y L Q 492
Ce-Slime_Mold - - - N D E K E L A I D I N N Y I V N F A T T H N P N T - G L S - V P V Q W R Q V T S T Q N S T L I 504
Consensus - - T E E E - - L S - - I M - Y W I N F A - - G N P N - - G L - - - W P - Y - - - - - - - - - Y L - 600


Chole-Yeast I N A L - - - - - - - G L Y T G K D N F R T A G Y D A L M T N P S S F F V - - - - - - - - - - - - - - - - 534
Lipase-Yeast I S A L - - - - - - - G L Y T G K D N F R T A G Y N A L F A D P S H F F V - - - - - - - - - - - - - - - - 549
Lipase-Fungal I H M T - - - - - - - D N V M R T D D Y R I E G I S N F E T D V N L Y G - - - - - - - - - - - - - - - - 543
Carbamate_Hyd V G D E - G A E V A R D L A R P D H V S V R T L - - - - - - - - - - - - - - - - - - - - - - - - - - - - 493
Pnb_Esterase L D S E I T I E N D P E S E K R Q K L F P S K G E - - - - - - - - - - - - - - - - - - - - - - - - - - - - 489
Ace-Rabbit L N L R P L E V R R G L R A Q A C A F W N R F L P K L L S A T D T L D E A E R Q W K A E F H R W S S 563
Ace-Ray L N T E P M K V H Q R L R V Q M C V F W N Q F L P K L L N A - - - - - - - - - - - - - - - - - - - - - - - - 531
Bche-Rabbit - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 485
Ce-Human I G A N T - Q A A Q K L K D K E V A F W T N L F A K K A V E K P - - - P Q T E H I E L - - - - - - - 567
Teb-Duck I D L T Q - K A A K K L K E R K M E F W M Q L T E Q I M S D R R - - - R K H T D L - - - - - - - - - 557
Ce-Rabbit L N M Q P - A V G Q A L K A R R L Q F W T H T L P Q R V Q E L R G T E Q K H T E L - - - - - - - - - 532
Ce-Slime_Mold L E T T I E T K D T F T N D P K C N A L D L T Y Y R N Q V R P - - - - - - - - - - - - - - - - - - - 535
Consensus I - - - - - - - - - - - L - - - - - - - F W - - - L - - - - - - - - - - - - - - - - - - - - - - - - - - - - 650
Chole-Yeast - - - - - - - - - - - - - - - - - - - - - - 534
Lipase-Yeast - - - - - - - - - - - - - - - - - - - - - - 549
Lipase-Fungal - - - - - - - - - - - - - - - - - - - - - - 543
Carbamate_Hyd - - - - - - - - - - - - - - - - - - - - - - 493
Pnb_Esterase - - - - - - - - - - - - - - - - - - - - - - 489
Ace-Rabbit Y M V H W K N Q F D H Y S K Q D R C S D L 584
Ace-Ray - - - - - - - - - - - - - - - - - - - - - - 531
Bche-Rabbit - - - - - - - - - - - - - - - - - - - - - - 485
Ce-Human - - - - - - - - - - - - - - - - - - - - - - 567
Teb-Duck - - - - - - - - - - - - - - - - - - - - - - 557
Ce-Rabbit - - - - - - - - - - - - - - - - - - - - - - 532
Ce-Slime_Mold - - - - - - - - - - - - - - - - - - - - - - 535
Consensus - - - - - - - - - - - - - - - - - - - - - - 671

Radic, 1994). Second, the alignment allows for the identification of which enzymes are most closely related. Using PHYLIP, a program designed to help identify evolutionary information based on sequence data (Felsenstein, 1989), a tree consisting of members of the α/β hydrolase family was constructed, as shown in Figure 23. The branching and the length of the lines connecting the members indicate the degree to which the members are related. Crystal structures have been solved for two members of this tree, *Torpedo californica* acetylcholinesterase (Ace - Ray) (Sussman *et al.*, 1991) and *Geotrichum candidum* triacylglycerol lipase (lipase - fungus) (Schrag and Cygler, 1993). When the two crystal structures are overlaid, the structural features of α-helices and β-sheets overlap well (RMS deviation of 399 corresponding $C_\alpha$ atoms is 1.90 Å), with regions of non-overlap arising almost exclusively in the loops connecting these large elements of secondary structure (Figure 24) (Ollis, 1992). These loops primarily serve to define the substrate binding site, with the small acetylcholine binding pocket being closed on both ends, and the lipid binding pocket being much more open. Because of the high degree of overlap between the two crystal structures, it is likely that every point on the line connecting these two structures in Figure 23 represents a structure almost identical to these two. Because the pNB esterase "branch" is located between the two crystal structures, the evolutionary distance between pNB esterase and a "known" structure is not as great as the distance to either the acetylcholinesterase or the lipase structure. Furthermore, the overlap of the two crystal structures represents a good approximation of this "known" structure. Third, the sequence alignment shows three large sequence regions of eight or nine amino acid residues in acetylcholinesterase for which no sequence exists in pNB esterase. Since the homology indicates pNB esterase is similar to acetylcholinesterase, the sequence alignment is locating the probable sites for differences in the protein structures. This suggests that only three regions contain large differences, and these differences are deletions from the existing structure. The structure similarity between the two crystal

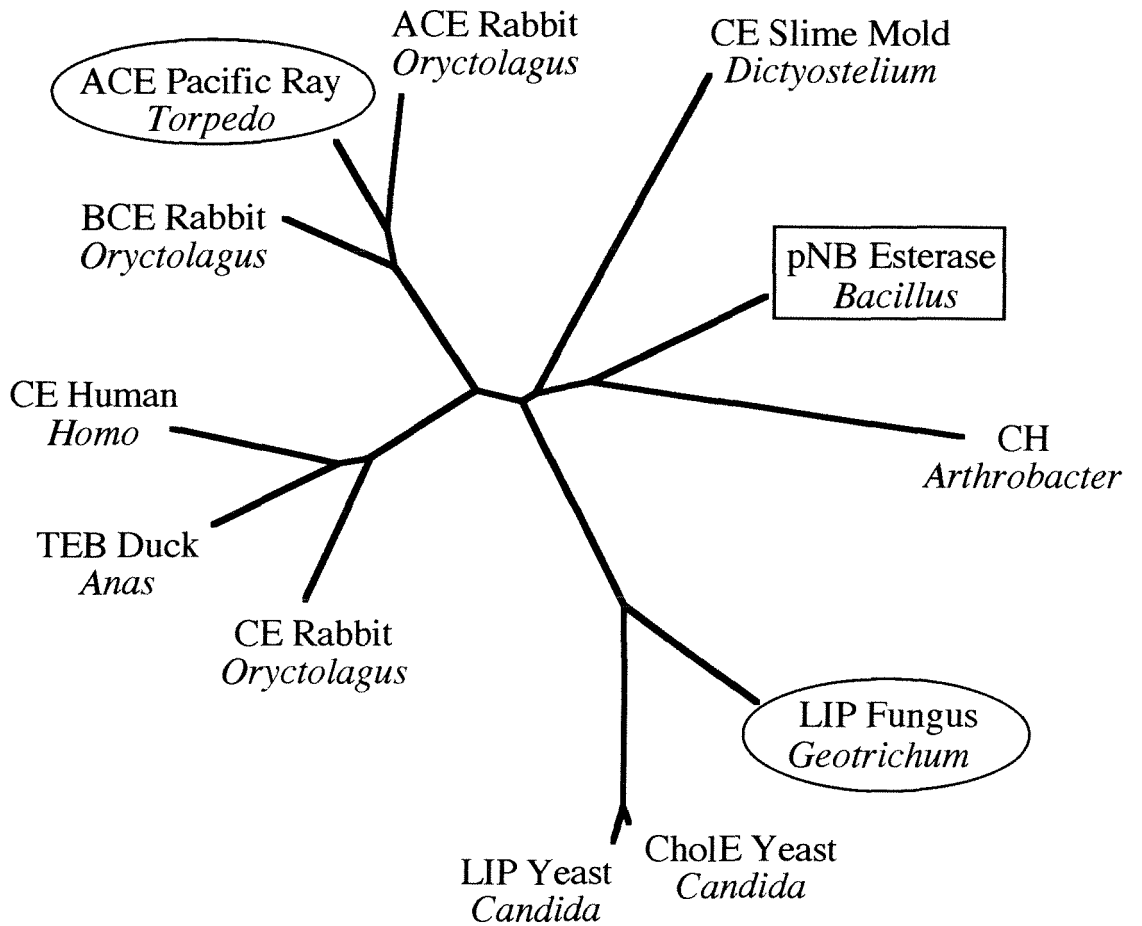



Figure 2.23: Evolutionary tree showing the phylogenetic relationship between the esterases listed in Table1 and pNB esterase (boxed). The enzymes are listed by enzyme name and source. The enzyme abbreviations are: ACE - acetylcholinesterase, BCE - butyrylcholinesterase, LIP - Lipase, CE - carboxylesterase, TEB - thioesterase b, and CH - carbamate hydrolase. The genus names for the common organism names are shown in italics. Crystal structures have been solved for the enzymes labelled "ACE Pacific Ray" and "LIP Fungus" (circled).

structures and the sequence similarity between the crystal structures and pNB esterase suggest that the two known structures can be used to construct a model of pNB esterase.

A computer program MODELLER (Sali and Blundell, 1993, and Sali and Overington, 1994) was used to construct a pNB esterase structural model from the acetylcholinesterase and triacylglycerol hydrolase structures. The two known crystal structures were aligned in three dimensions by Modeller using the crystal structure coordinates. Approximately 80% of the residues in these two crystal structures overlap within 2.0 angstroms (as shown in Figure 24), generating a structure-based sequence alignment. Many of the non-overlapped residues are in a region of the protein where a hinge-type motion has altered the location of residues but not the secondary structural features, and these were manually aligned within the structure-based sequence alignment. The remaining non-overlapped residues are located in loops connecting these conserved elements and usually are caused by insertion of residues. For instance, a loop in acetylcholinesterase might contain four residues while the same loop in triacylglycerol hydrolase might contain ten residues. Initially these loops were left as non-identical portions of the structure-based sequence alignment. Using the original sequence alignment in Figure 22, pNB esterase was aligned with the structure-based alignment. The ACE - Ray and LIP - fungus sequences in the original sequence alignment match almost exactly the true structure-based alignment, and where this is true, the original sequence alignment clearly affixes the pNB esterase sequence to the structure-based sequence. In some loop regions, however, the sequence alignment does not match the structure-based alignment, and when different, pNB esterase was aligned with the structure-based alignment manually. Two main criteria were used. First, if the number of residues in a loop connecting conserved elements in pNB esterase is in between the number of loop residues in each of the two crystal structures, the two structures were overlapped in the structure-based alignment, and the program used both loops to calculate an "average" pNB esterase loop (Srinivasan and Blundell, 1993). If the number of residues were identical to either

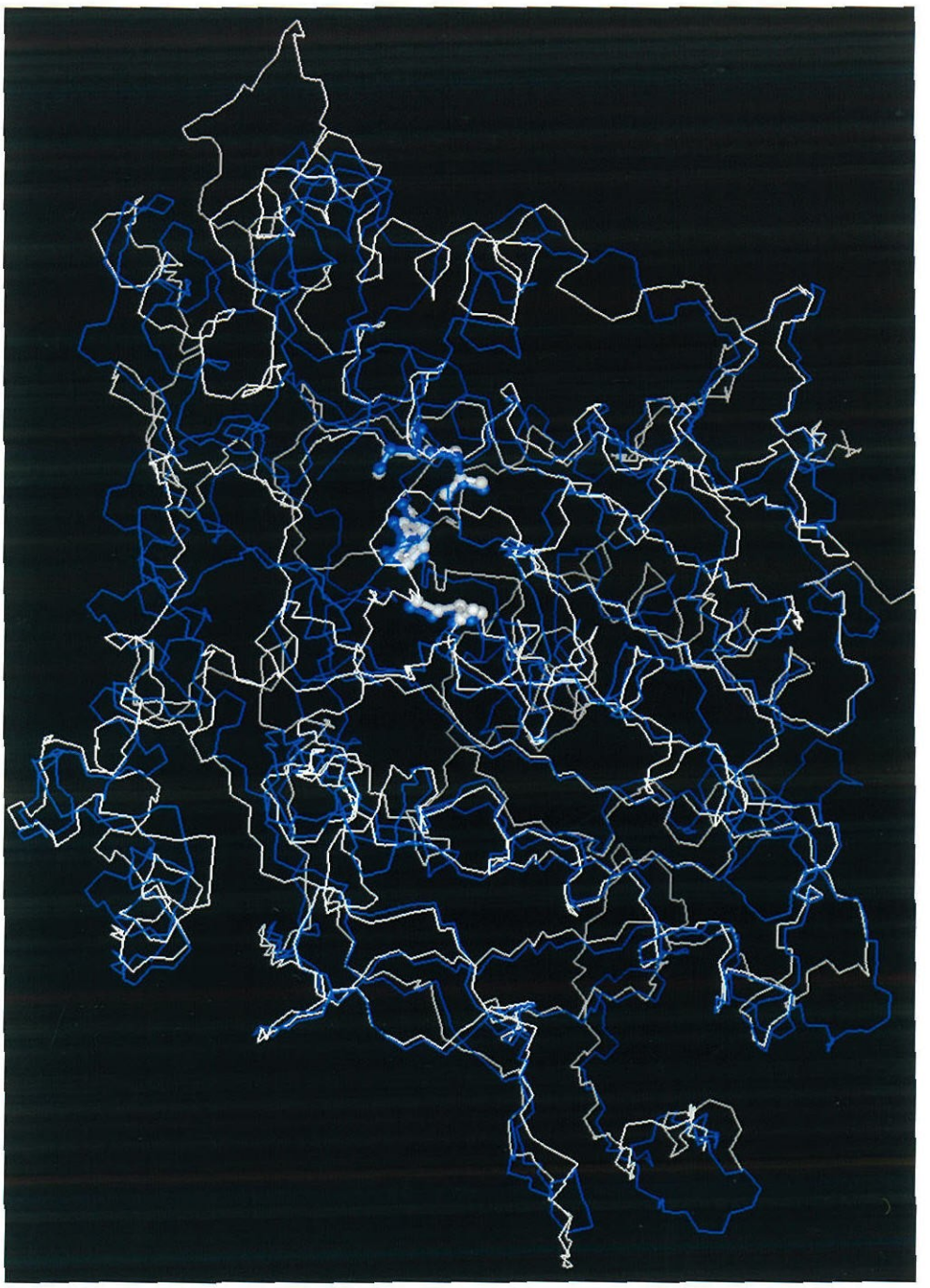


Figure 2.24: Superposition of the crystal structures for acetylcholinesterase (in grey) and triacylglycerol lipase (in blue) as determined by Modeller. The active site His, Ser, and Glu are represented by the ball-and-stick structures. The superposition was performed using Insight from Biosym.

structure, the loop in pNB esterase was modeled according to that loop only. Second, the alignments were adjusted to allow for the flexibility to maintain the location and orientation of secondary structural elements. For instance, in locations where not enough residues were present to complete a loop without carbon - carbon bond stretching or moving of large portions of secondary structure, residues that in the sequence alignment were aligned with the end of a helix were shifted into loop residues. This shortened the helix by half of a turn and made a more plausible loop connection. Often prolines and glycines were used as indicators in such instances by trying to position these residues as the last residue in the secondary structure element or the first residue in the loop. Modeller then took the new sequence alignment and the two crystal structures and calculated a structure for pNB esterase.

The acetylcholinesterase crystal structure includes the acetylcholine substrate (Sussman, *et al.*, 1991). Figure 25 shows the resulting calculated pNB esterase structure superimposed with the acetylcholinesterase structure and with the acetylcholine substrate displayed. The acetylcholine substrate was converted to LCN-pNB using Insight v. 2.3 (Biosym) and added to the pNB esterase structure. The orientation and position of the ester bond in the acetylcholine substrate was maintained in the LCN-pNB structure, so that the alcohol-containing choline group was replaced with the alcohol-containing pNB group, and the carboxyl-containing acetyl group was replaced with carboxyl-containing LCN. The resulting substrate was minimized using 100 iterations of the conjugate gradients method in the Discover module of Insight. The minimization was done only on the substrate; the protein was not allowed to move. The minimization repositioned both the LCN and pNB groups to locations where this new substrate did not overlap side chains in the enzyme structure. Figure 26 shows the LCN-pNB substrate in white and the conserved active site residues in red. This figure is oriented so that the pNB leaving group is pointing away and the loracarbef nucleus portion is oriented toward the front.

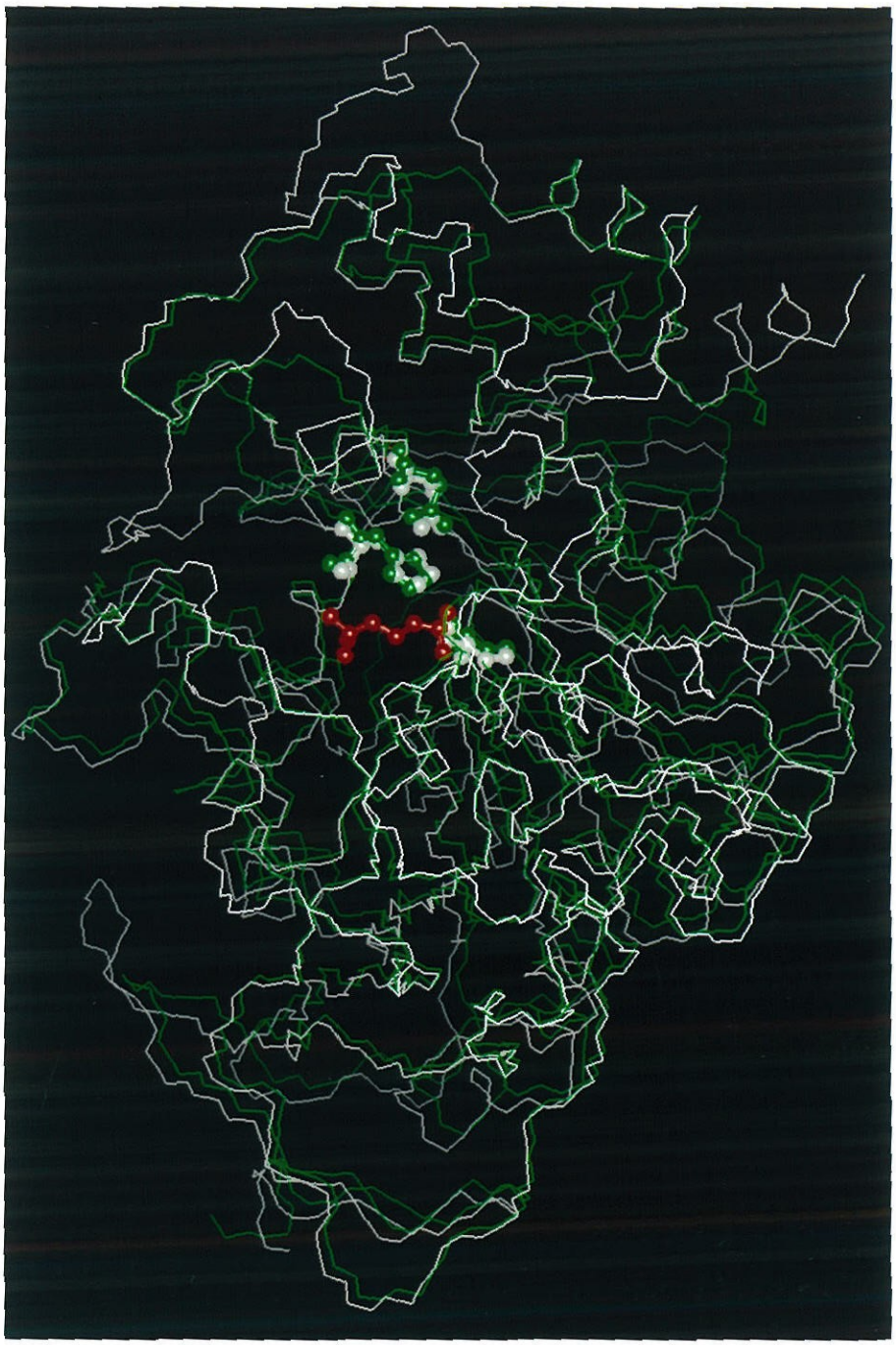


Figure 2.25: The pNB esterase model (in green) as calculated by Modeller superimposed on the acetylcholinesterase crystal structure (in grey). The structure in red is the acetylcholine substrate bound to acetylcholinesterase in the crystal structure. The ball-and-stick residues are the active site residues of both enzymes. The superposition was performed using Insight from Biosym.

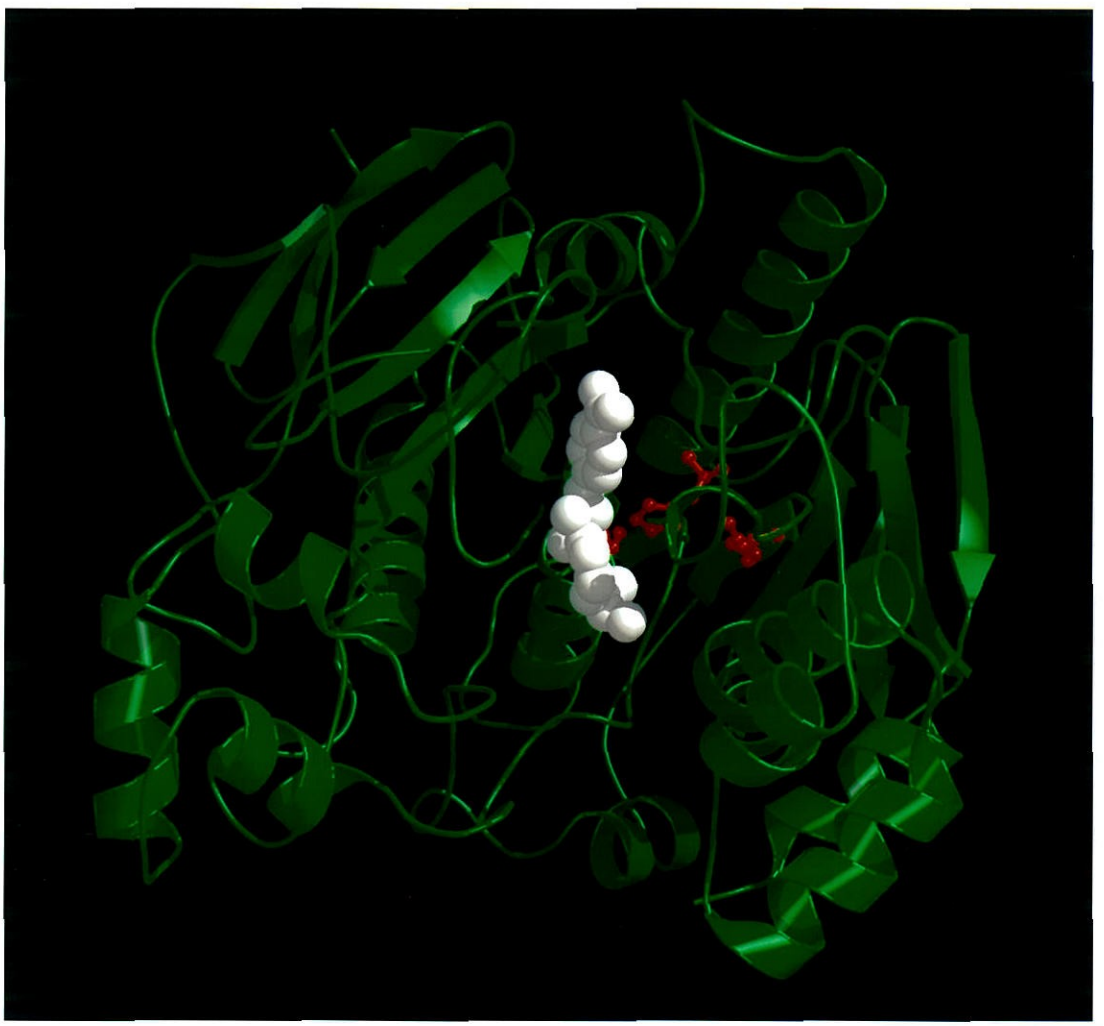


Figure 2.26: The pNB esterase enzyme with the LCN-pNB substrate. The acetylcholine substrate from Figure 23 was incorporated into the pNB esterase model to determine the location and orientation of substrate binding. The acetylcholine was then altered to the LCN-pNB substrate by replacing the acetyl portion of acetylcholine with LCN and the choline portion with pNB. The LCN-pNB substrate was then energy minimized using Insight's Discover module.

Combining this model with the sequence data allows for the positioning of the amino acid substitutions resulting in enhanced activity. With the substitutions from wild type shown in orange, Figure 27 depicts a model of 2-19E10. According to the model structure, valine 358 is located on a helix within the protein core on the edge of the binding pocket. The additional substitutions seen in this structure derive from the 1-1H9 parent. The methionine to valine substitution at position 358 replaces a methionine seen in only one of the homologous enzymes (carboxylesterase from slime mold) to the valine seen in four of the homologous enzymes, as shown in Figure 22. The two substitutions from 1-1H9 are histidine 322 to arginine, a residue not seen in any of the homologous sequences, and tyrosine 370 to phenylalanine, a residue seen in acetylcholinesterase of the Pacific ray and rabbit butyrylcholinesterase.

Figure 28 shows 3-10C4 and the location of the leucine to methionine substitution at position 144 responsible for effecting the enhancement in activity. The additional substitutions (again in orange) reflect changes from wild type arising from 2-19E10, the parent of 3-10C4. This substitution occurs on the same face as previous substitutions but is located somewhat distant from the loracarbef substrate. The original leucine residue is actually the consensus residue among the homologous enzymes, although methionine is seen in carboxylesterase from slime mold, as shown in Figure 22.

Figure 29 shows the best variant produced in this directed evolution experiment (4-54B9) and highlights the two new amino acid substitutions, a lysine to arginine at residue 267 and a leucine to serine at position 334, displayed in the background of the parent 3-10C4. Both substitutions are located in the portion of the binding pocket expected to contain the loracarbef moiety. At least one of these residues is responsible for the large increase in activity associated with this variant. A second fourth generation variant, 4-43E7, has a different substitution also located at residue 334 (leucine to valine), and this may indicate that the amino acid residue at this position is responsible for the activity enhancement. The original leucine at position 334 is the most common residue at this

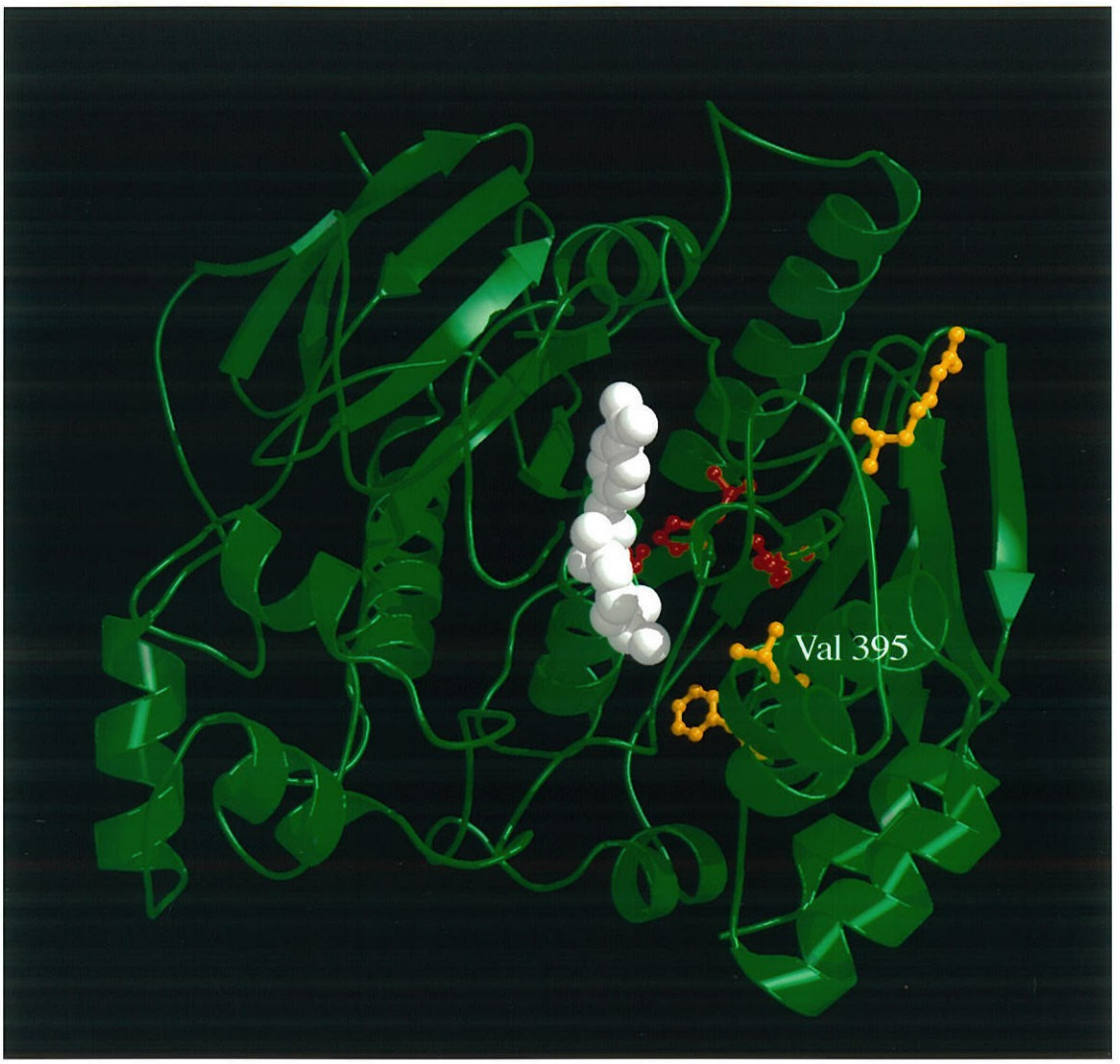



Figure 2.27: Variant 2-19E10 showing the location of the Val 395 substitution (indicated in orange) with respect to the substrate (in grey) and the active site (in red). Also shown are substitutions Arg 322 and Phe 370 (in orange) derived from 1-1H9.

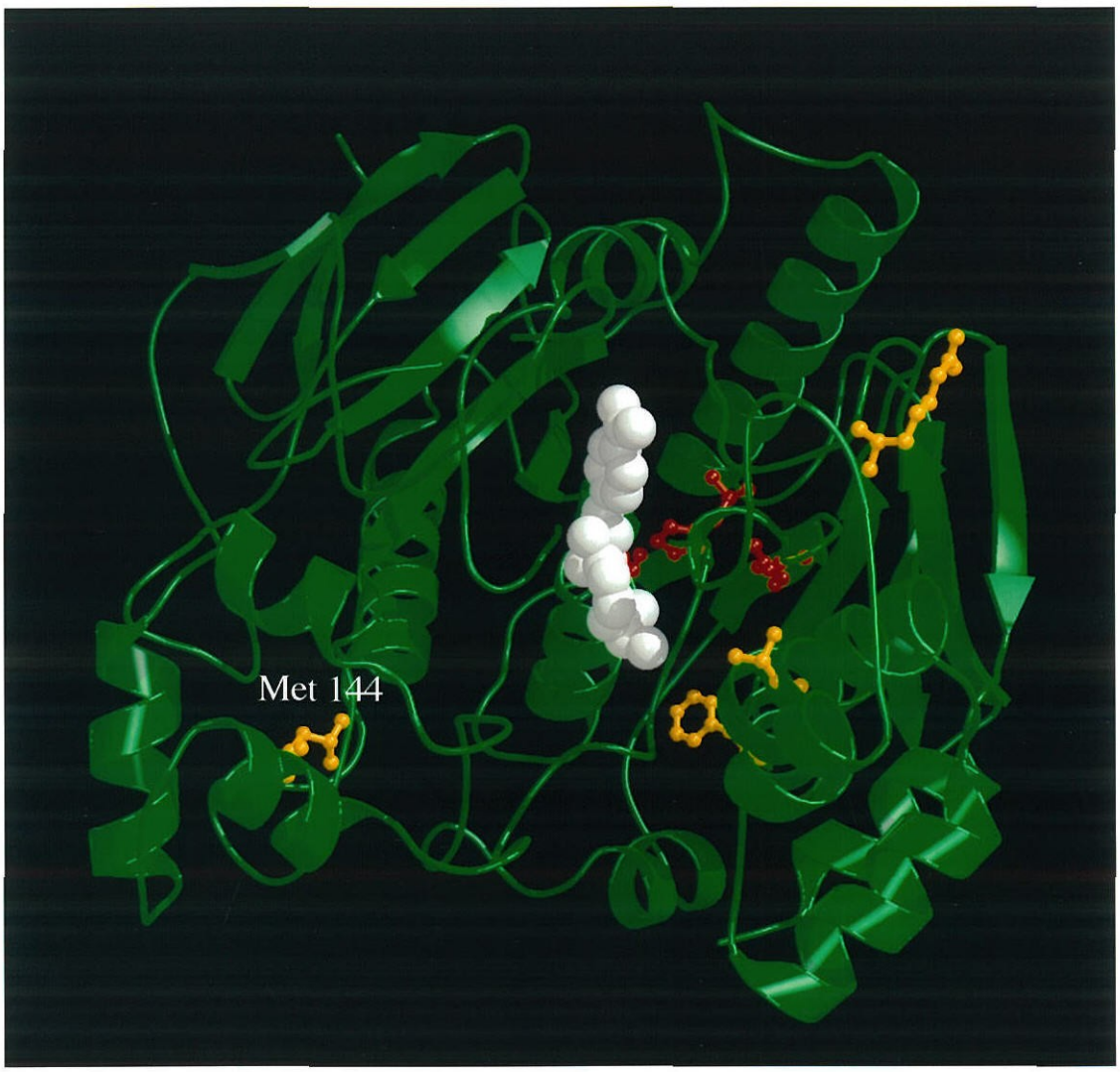



Figure 2.28: Variant 3-10C4 showing the location of the Met 144 substitution (indicated in orange) with respect to the substrate (in grey) and the active site (in red). The remaining substitutions (in orange) are those from this variant's parent, 2-19E10.

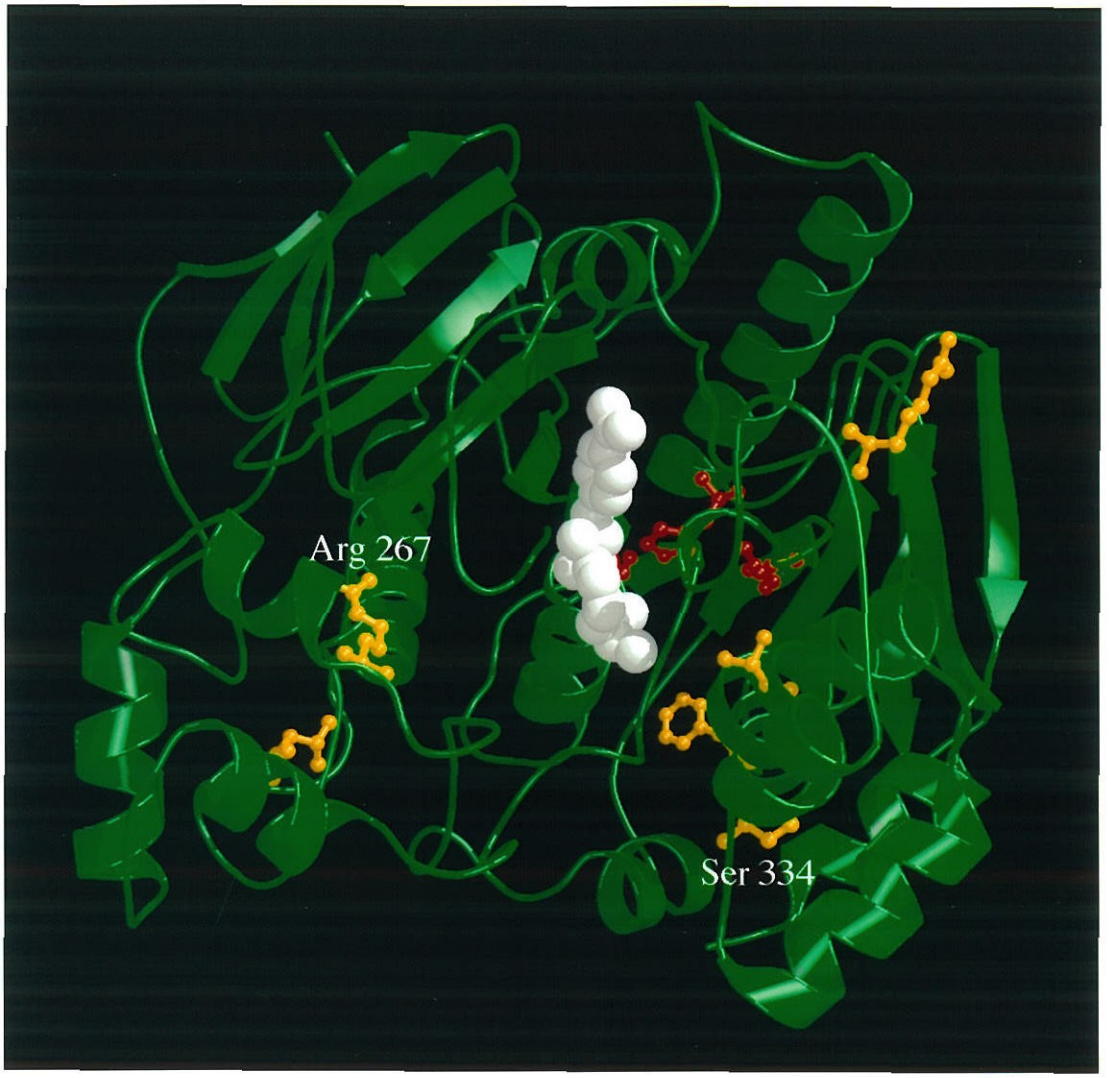



Figure 2.29: Variant 4-54B9 showing the location of the Arg 267 and Ser 334 substitutions (labeled, in orange) with respect to the substrate (grey) and the active site (in red). The remaining substitutions (also in orange) are those from this variant's parent, 3-10C4.

location in the homologous enzyme sequence alignment, with valine appearing in the two acetylcholinesterases and serine not seen in any homologous enzyme. Position 267 is located in a gap region for five of the eleven homologous sequences, and while the original lysine was shared by carbamate hydrolase, arginine is not seen in any of the homologous sequences.

Figure 30 shows another fourth generation variant, 4-38B9 and its substitution at residue 60 from isoleucine to valine, again in the background of 3-10C4. The substitution is on the surface of the enzyme where it should interact easily with solvent, but far removed from the substrate binding pocket where the majority of previous substitutions were found. The original isoleucine residue at position 60 had no corresponding similar residue among the homologous enzyme sequences examined; the change to valine is seen in rabbit acetylcholinesterase at the corresponding position, as seen in Figure 22. Table 4 summarizes all of the information presented on all the substitutions.

## Discussion

*Random Mutagenesis*

The method of random mutagenesis chosen for this study is error-prone PCR (Leung *et al.*, 1989 and Cadwell and Joyce, 1992). This method has been shown both in this work in Figure 18 and in previous work (Cadwell and Joyce, 1992 and Chen and Arnold, 1991) to perform mutagenesis located randomly within the pNB esterase gene and other genes of interest. In addition, conditions can be such that no strong bias for any one type of mutation exists (Leung *et al.*, 1989 and Cadwell and Joyce, 1992); this is exactly opposite to the strong bias for A to G transitions seen in this study in Table 3. The reported conditions regulating the lack of bias in the literature are strongly dependent on $Mn^{2+}$ and unbalanced dNTP concentrations in the reaction mixture. The conditions used in this work are reported in the Materials and Methods and lack $Mn^{2+}$ and unbalanced dNTP concentrations. $Mn^{2+}$ and unbalanced dNTP concentrations were not used in this directed

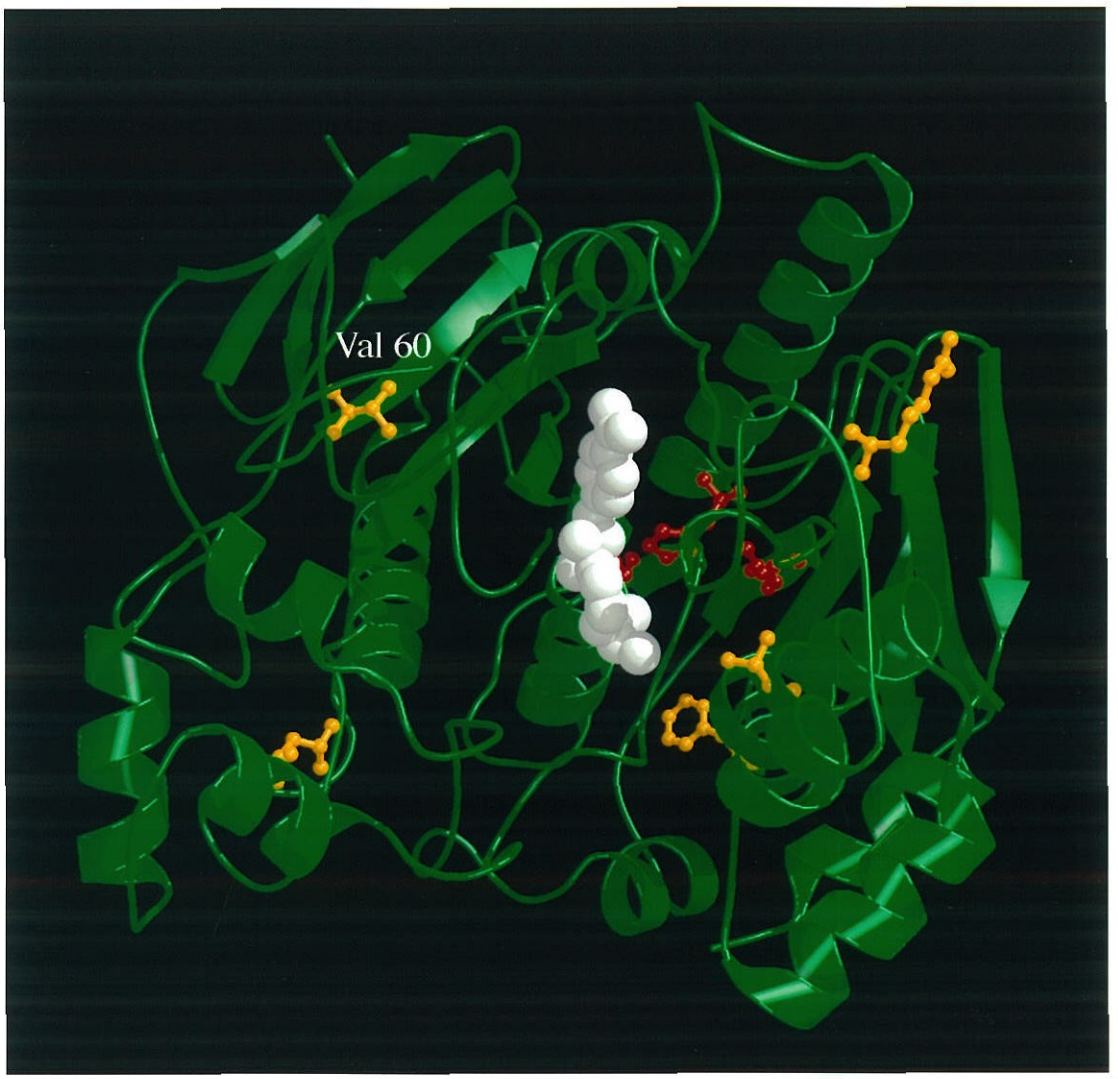



Figure 2.30: Variant 4-38B9 showing the location of the Val 60 substitution (labeled, in orange) with respect to the substrate (in grey) and the active site (in red). The remaining substitutions (also in orange) are those from this variant's parent, 3-10C4.

| amino acid substitution | pNB esterase variant | location | distance from substrate [a] (Å) | element of secondary structure | original residue conserved[b] | replacement residue found [c] | most common residues at this position |
|---|---|---|---|---|---|---|---|
| Ile 60 Val | 4-38B9, 4-73B4 | near surface | 13 | none | No | 2 | M/S |
| Ser 94 Gly | 4-43E7 | surface | 30 | surface loop | No | 4, 8, 9, 10 | G/K |
| Asn 96 Ser | 2-13F3 | surface | 30 | surface loop | No | 10 | N/K |
| Leu 144 Met | 3-10C4 | buried | 20 | surface loop | Yes | 5 | L |
| Lys 267 Arg | 4-54B9 | near surface | 12 | none | No | none | I (6/11)[d] |
| Phe 271 Leu | 2-23E1 | surface | 13 | surface loop | No | 8, 9 | L (6/11)[d] |
| Leu 334 Val | 4-43E7 | surface | 16 | helix | Yes | 1, 2 | L |
| Leu 334 Ser | 4-54B9 | surface | 16 | helix | Yes | none | L |
| Ala 343 Val | 2-23E1, 4-53D5 | near surface | 20 | helix | No | 1, 2, 6, 7 | V/L |
| Met 358 Val | 2-13F3, 2-19E10 | surface | 7 | helix | No | 1, 2, 3, 10 | V/L |
| (His 322 Arg) | 1-1H9 | surface | 24 | surface loop | No | none | E/D (6/11)[d] |
| (Tyr 370 Phe) | 1-1H9 | buried | 15 | helix | No | 1, 3 | L/V |

[a]Shortest distance from this residue to the substrate. [b]Yes = at least half of the homologous proteins contain the residue substituted at the corresponding location. [c]Indicates the homologous proteins from following list which contain the substituted amino acid residue. 1 - Acetylcholinesterase, *Torpedo californica* (Sussman, 1991), 2 - Acetylcholinesterase, *Oryctolagus cuniculus* (Jbilo, 1994), 3 - Butyrylcholinesterase, *Oryctolagus cuniculus* (Jbilo, 1990), 4 - Carboxylesterase, *Oryctolagus cuniculus* (Ozol, 1989), 5 - Carboxylesterase, *Dictyostelium discoideum* (Bomblie, 1990), 6 - Thioesterase, *Anas platyrhynchos* (Hwang, 1993), 7 - Triacylglycerol lipase, *Geotrichum candidum* (Schrag, 1993), 8 - Triacylglycerol lipase, *Candida rugosa* (Lotti, 1993), 9 - Cholesterol esterase, *Candida rugosa* (Kaiser, 1994), 10 - Carbamate hydrolase, *Arthrobacter oxidans* (Pohlenz, 1992). [d]Number of sequences (out of 11) that contain amino acid residues at this location.

Table 2.4: Amino acid substitutions in evolved pNB esterases.

evolution experiment because the addition of $Mn^{2+}$ and unbalanced dNTP concentrations increase the substitution frequency to a minimum of approximately 7 base substitutions per 1000 DNA bases (Cadwell and Joyce, 1992). This should give rise to on average 10 to 11 per gene in the case of pNB esterase and is well outside the desired level. To maintain a lower substitution rate, conditions which gave rise to non-random DNA substitutions were required. The conditions used in this study (differing from normal PCR conditions by increases in all four dNTP concentrations) give a reported error-rate of between 1.5 and 4 base substitutions per 1000 bases. Under these conditions, the substitutions should be predominantly transitions (e.g., A to G), no transversions (e.g., A to T), and a small fraction of insertion or deletions (Leung *et al.*, 1989 and Cadwell and Joyce, 1992). The sequence data presented here demonstrate a significant bias towards A to G substitutions, especially in the first two generations. We found that the substitution frequency was towards the higher limit of the expected range, with the 1-1H9 variant demonstrating a substitution rate of 3.5 substitutions per 1000 bases, and highly populated with A to G changes. The third and fourth PCR reactions yielded lower mutation frequencies, with only 1 to 2.5 substitutions per 1000 bases. Transition substitutions still outnumbered transversion substitutions by 3 to 1. In no case was a G mutated. Although this mutagenesis produced biased mutations, the conditions used allowed for a decreased mutagenic rate required for a sequence of this length and does not prevent the discovery of many positive variants.

Substitution bias in the resulting protein sequences also arises from the fact that the DNA sequence is translated to the amino acid sequence through the triplet-coding ribosomes. Twenty amino acids are encoded by 61 triplet DNA codons; the distribution of these codons is far from even. For example, tryptophan is coded for by only one of the 64 triplet codons (TGG), while leucine is encoded by 6. Single base changes within a codon are the only type of base changes we can expect to see, as the probability of making two random substitutions within one codon is vanishingly small. On average, only five to six

new amino acids are available to replace each amino acid in the original sequence by single base substitution. Thus, the translation process introduces non-randomness in the amino acid sequence. Additionally, the sets of codons for each amino acid are designed conservatively, so that single base changes within a triplet codon will often generate substitutions at the protein level which are similar in properties to the original amino acid residues. This may be useful for directed evolution, when one does not want to radically alter an enzyme that has some catalytic activity, but rather wants to incorporate small changes which enhance the enzyme's ability to perform. This bias towards conservative amino acid substitutions was observed in the current study: most changes (although not all) were conservative ones, such as tyrosine 370 to phenylalanine in 1-1H9, phenylalanine 271 to leucine in 2-23E1, and isoleucine 60 to valine in 4-38B9.

*Screening Strategy*

The screening of the fourth generation clones serves to validate the screening strategy used in this study. Using 3-10C4 as the parent for the fourth generation, 7400 clones were screened, a much larger number of clones than in any previous generation. Clones which appeared to outperform 3-10C4 either in total activity or in the ratio of activities in high vs. low DMF concentrations were assayed on both LCN-pNP and the target substrate LCN-pNB. From these 64 clones, a measure of how well the activity of these enzymes on the screening substrate relates to activity on the substrate of interest was established (Figure 7). The overall trend demonstrates a good correlation between activities on the screening and actual substrates. The points would all lie on the 45° line if increases in activity on one substrate correlated exactly with increases in activity on the other. The distribution is skewed slightly toward the screening substrate, as demonstrated by the trend of data points around the 45° line. The slope of the best-fit line through the data is 0.7, indicating a slightly less than one-to-one correspondence between activity on the desired substrate to the activity on the screening substrate. The strength of this correlation

(correlation constant, R = 0.8) is an important test of the screening strategy designed for this work. This graph clearly validates the screening strategy premise that a structurally similar LCN-pNP can successfully replace LCN-pNB, the hydrolysis of which is difficult to measure. Contrast this with the original pNPA substrate, whose structure does not mimic the actual substrate well and whose activities do not seem to correlate well with activity on the hybrid or actual substrates.

*Kinetic Parameters*

The kinetic parameters shown in Table 3 and the catalytic efficiencies plotted in Figure 13 demonstrate not only the progress of directed evolution, but also the effects of the strategy chosen to direct the evolution. For instance, aside from the first generation variants, the substrate p-nitrophenyl acetate was not used for screening. In the absence of DMF, none of the enzyme variants show an increase in catalytic efficiency on this substrate (Figure 13a). However, presumably because the screening was performed in DMF, the enzyme variants' efficiencies on the pNPA substrate become better than the wild type pNB esterase in high DMF concentrations.

Unlike the pNPA substrate, p-nitrophenyl loracarbef nucleus was used for screening generations two through four. In these generations a steady increase in the catalytic efficiencies is observed (Figure 13b). The $K_M$ values for the enzyme variants on the LCN-pNP substrate also reflects the evolution strategy. In 1% DMF wild-type pNB esterase's $K_M$ is 0.066 (Table 2), and most of the variants do not improve this value. This is because there was little or no screening pressure to do so. The substrate concentrations used during screening at the lower concentrations of DMF were always 0.3 mM or approximately five times $K_M$. As a result, improvements in activity are largely in $k_{cat}$. However, when the DMF concentration increases to levels which force $K_M$ above 0.8 mM, the variants now all demonstrate $K_M$ values better than wild-type.

Figure 7 has shown that activity on LCN-pNP correlates well with activity on LCN-pNB, and the kinetic parameters reinforce this observation. The trends observed in the kinetic parameters on LCN-pNP are mirrored almost without exception in the enzyme variants' behavior on LCN-pNB (Figure 13c). The $K_M$ values for LCN-pNB are similar to those for LCN-pNP and follow the $K_M$ trends for LCN-pNP. The turnover numbers for LCN-pNB are considerably lower than those for LCN-pNP, presumably reflecting the more difficult catalysis that the pNB protecting group generates. The values exhibit the same trend as on LCN-pNP, however, increasing from low to high as the generation number increases.

*Homology Studies and Structural Modeling*

The homology search revealed similarities between pNB esterase and other esterases in the $\alpha/\beta$ hydrolase family. When enzymes share 30% sequence identity, the fraction of residues correctly aligned in sequence alignments is approximately 80% (Johnson and Overington, 1993). This 80% can be improved upon by aligning multiple sequences with similar sequence identity (Johnson and Overington, 1993), and the accuracy of the alignment is important to the calculation of a model structure. Table 1 shows that the various native esterase enzymes share a great deal of sequence similarity, and from the conservation of function and similarity between these enzymes, it is logical to conclude that the three-dimensional fold is highly conserved. These sequence alignments and the two crystal structures from *Torpedo californica* acetylcholinesterase and *Geotrichum candidum* triacylglycerol lipase were used to approximate a three-dimensional structure for pNB esterase. This was accomplished utilizing the program MODELLER, which used the alignments, the crystal structures, and a large array of probabilistic distance, dihedral angle, and spatial restraints derived from a database of three-dimensional structures, to calculate the pNB esterase structure (Sali, 1995). Examining the major differences between the acetylcholinesterase and pNB esterase structures compared in

Figure 24, the most noticeable and most telling feature is that the substrate binding pockets are built differently, especially in terms of size. The differences observed between the overlaid acetylcholinesterase and pNB esterase structures serve mainly to enlarge the binding pocket in pNB esterase to accommodate the larger substrate shown in Figure 26. Enzymes showing homology were tested for their ability to hydrolyze the LCN-pNB substrate to examine the possibility that other naturally available enzymes could perform the desired reaction. From the structure presented in Figure 25, that pNB esterase is the only enzyme capable of performing this reaction can be rationalized. The crystal structure shows that a small number of amino acid differences between acetylcholinesterase and pNB esterase significantly alter the binding pocket. The sequence alignments of the remaining homologous enzymes show similar variability in the sizes and residue compositions of loops responsible for defining the substrate binding pocket. The wide array of substrates recognized by the different esterases reflects sequence and structural variations in this region. It also means that significant changes in substrate selectivity can be expected as a result of accumulated alterations in the amino acids surrounding the binding pocket.

*Analysis of Variants*

Figure 31 shows top and side views of wild-type pNB esterase with the LCN-pNB substrate (grey). Residues in the catalytic triad are indicated in red. A total of 12 amino acid substitutions at 11 different sites were found in the variants with improved activity towards various substrates and in the presence of DMF. These substitutions are highlighted in Figure 31 and summarized in Table 4, along with an indication of whether the amino acid is buried or surface-accessible, whether the amino acid is found in an element of secondary structure, and whether or not the particular amino acid is conserved among the homologous enzymes. Similar to what was observed during the directed evolution of subtilisin for activity in dimethylformamide (Chen and Arnold, 1993, You and

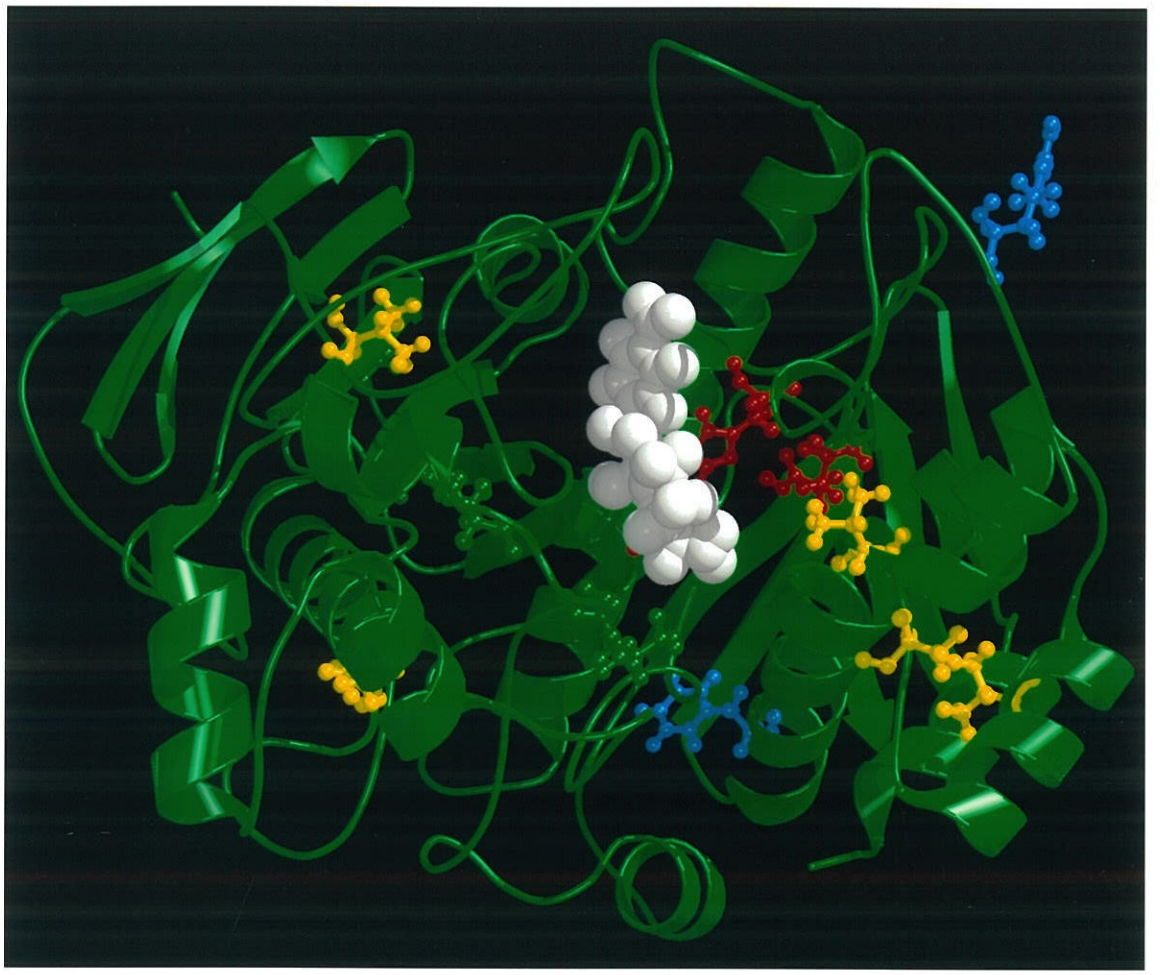


Figure 2.31a: Top view of pNB esterase structure with the LCN-pNB substrate (in grey). Amino acid substitutions found in evolved pNB esterase variants are highlighted: substitutions found in 1-1H9 are shown in cyan, substitutions believed to be neutral are indicated in green, while substitutions that enhance specific esterase activity are shown in yellow. This view highlights the radial distribution of mutations around the active site.

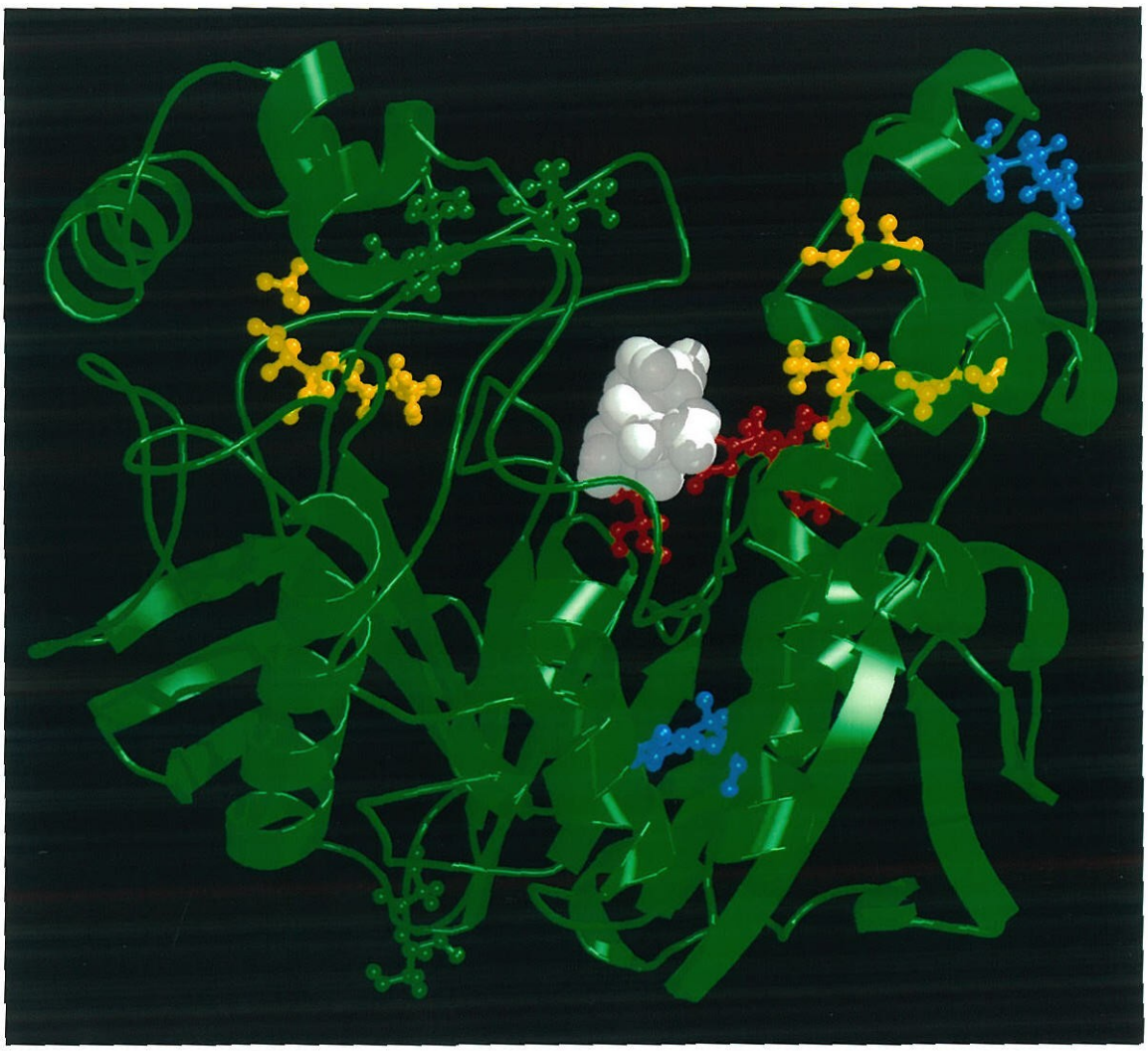


Figure 2.31b: Side view of pNB esterase structure with the LCN-pNB substrate (grey). Active site residues are shown in red. Amino acid substitutions found in evolved pNB esterase variants are highlighted: substitutions found in 1-1H9 are shown in cyan, substitutions believed to be neutral are indicated in green, while substitutions that enhance specific esterase activity are shown in yellow. This view shows that the effective mutations are clustered in the top half of the enzyme.

Arnold, 1995) the majority of the amino acid substitutions lie on or near the surface of the enzyme and/or are clustered around the substrate binding pocket.

Of the 12 amino acid substitutions identified, 2 were found in 1-1H9, an enzyme variant expressed at higher levels but without increased specific activity (His 322 Arg and Tyr 370 Phe, listed in parentheses in Table 4 and indicated in blue in Figure 31). Expression level is not necessarily a property of the amino acid sequence; alterations in the DNA sequence may play an important role in transcriptional regulation of expression. The two amino acid changes or any of the six DNA mutations could be responsible for increasing the total activity of this clone.

Of the remaining ten amino acid substitutions, the four at positions 94, 96, 267, and 271 are believed to be neutral or deleterious (shown in green in Figure 31). The Asn 96 Ser substitution occurs in variant 2-13F3. 2-13F3 and 2-19E10 share Met 358 Val, which is the only mutation found in the more active 2-19E10 variant. It can therefore be argued that position 358 is responsible for the improved activity, while the substitution at position 96 is neutral, or even possibly deleterious. Similarly, since variants 2-23E1 and 4-53D5 share Ala 343 Val, and this is the only substitution found in going from 3-10C4 to 4-53D5, the additional Phe 271 Leu substitution in variant 2-23E1 is probably incidental. The substitutions at residues 94 and 267 were found accompanied by one additional amino acid substitution each in 4-43E7 and 4-54B9, respectively. These two variants share substitutions at position 334, and this residue is likely to be responsible for the enhanced activity.

Most mutations are neutral or deleterious. The size of the library containing all double amino acid substitutions is larger than the sequence length squared. Screening on the order of a thousand colonies covers only a tiny fraction of this mutant space, and is, therefore, very unlikely to uncover a sequence containing two mutations, both of which positively contribute to activity enhancement. Thus it is argued that only one of the two mutations in 4-43E7 and 4-54B9 is responsible for these variants increase in activity, and

that one is the one they have in common, at position 334. This is further supported by the recombination experiment (*vide infra*) which combined Val 60 from 4-38B9 and Ser 334 from 4-54B9 while omitting Arg 267, to create 5-1A12, a variant with twice the apparent activity of 4-54B9.

The neutral substitutions at residues 94, 96, 267, and 271 are all located at or near the surface of the enzyme in flexible surface loops and outside discernible elements of secondary structure. In addition, these positions are not conserved among the homologous enzymes, suggesting that the pNB esterase can tolerate variability at these locations.

The six remaining substitutions are those believed to be responsible for enhancements in specific esterase activity and are shown in yellow. Four of these (at residues 334, 343, and 358) are positioned in α-helices, located on the LCN end of the elongated substrate binding pocket and on one side of the entrance to the binding pocket, as seen in Figure 31. Val 358 is the closest to the substrate, at 7 Å from the nearest point of the substrate, while the others are up to 20 Å away (Table 3). Thus it appears that none of these four mutations can directly interact with bound substrate. The amino acids at these locations in the wild type and homologous enzymes are typically hydrophobic; the large, hydrophobic amino acid Leu predominates at these positions in the homologous enzymes. The evolved pNB esterase variants contain smaller valine and serine. Ala 343 Val in variant 2-23E1 is responsible for an increase in activity on the screening substrate LCN-pNP but does not improve activity toward LCN-pNB. Why this mutation should affect activity toward one substrate and not the other is unclear.

The remaining amino acid substitutions which enhance esterase activity are at positions 60 and 144. Ile 60 Val (in 4-38B9) is responsible for a significant improvement in catalytic efficiency, as indicated in Table 2 and Figure 13. This residue is located near the surface of the enzyme, 13 Å from the closest point of the substrate. This is the only activity-enhancing mutation that is closer to the leaving group (pNP or pNB) than to the LCN portion of the substrate. Leu 144 Met is the only one of the six that is buried deep

beneath the solvent-accessible enzyme surface. It is also located relatively far from the substrate, at 20 Å from the LCN. This substitution's contribution to the activity of 3-10C4 is greatly enhanced in the presence of DMF (Table 2, Figure 13). The only other variant showing similar behavior is 1-1H9 toward pNPA. 1-1H9 is also the only other variant to have a completely buried substitution (Tyr 370 Phe).

The arrangement of mutations is similar to that seen with previous experiments with subtilisin (Chen and Arnold, 1993, You and Arnold, 1996). The mutations as seen from a top view in Figure 31a are radially distributed about the substrate just as found in the previous subtilisin experiments. In Figure 31b, the non-random distribution of mutations enhancing activity is clearly seen. The pNB esterase is different from the previous subtilisin example in that the pNB esterase catalysis of the substrate occurs deep within the enzyme structure and not near the enzyme's surface. Correspondingly, the mutations which enhance activity are located not clustered on one surface face of the enzyme as in the subtilisin example, but rather, in a region more closely associated with the depth of the substrate within the enzyme. In general, it is difficult to explain the mechanism(s) by which these amino acid substitutions enhance the catalytic activity of the evolved pNB esterases toward any of these substrates. None of the effective amino acid substitutions lie in segments of the enzyme predicted to interact directly with the bound substrate. To predict in advance that these substitutions enhance pNB esterase activity would be practically impossible.

**Conclusions**

The pNB esterase enzyme is useful both in the deprotection of antibiotic pNB esters as well as potentially for the deprotection of more general pNB-protected substrates. By providing a biodegradable alternative to the current environmentally toxic, non-biodegradable catalyst responsible for the removal of the pNB-protecting group, the enzyme also has the potential to reduce the amount of environmentally toxic material

disposed of as a result of pNB deprotection reactions. By utilizing the histidine to metal interaction common to IMAC chromatography, the purification of this enzyme and its variants has been simplified. The structure of pNB esterase has been approximated using homologous sequences and a model of its three-dimensional structure established along with the probable location of binding for the substrate of this reaction. The mechanism of hydrolysis has been inferred through homology as proceeding through an acyl enzyme intermediate common to the serine hydrolase family, and moreover that this enzyme belongs to a special sub-class of this family called the $\alpha/\beta$ hydrolase family. Many members of this family utilize glutamate instead of aspartate as one of the members of the catalytic triad, and pNB esterase is believed to be in this category.

Directed evolution has allowed the improvement of this enzyme for performing the specific task of deprotecting the antibiotic loracarbef-pNB ester, and six variants were identified as possessing this activity enhancement. One variant in particular is 15 times more active than the original, 24 times more efficient, and performs as well in 30% DMF as the wild type does in aqueous solution. The substitutions which confer this increase in activity were located in the model structure and can be used to locate regions for additional mutagenesis, should further improvements be required. The family of enzyme variants created have been shown to hydrolyze pNB-protected esters of non-antibiotic structures, and one or more of the intermediates between the wild type enzyme and 4-54B9 may be a good starting candidate for evolution of activities toward other substrates of choice and for which existing available enzymes might not be suitable. Simple recombination between this variant and one other variant further increased the activity. The mutations responsible for the enhanced activity are located radially about the substrate binding pocket and at a depth consistent with the substrate.

More generally, this work demonstrates the feasibility of conducting molecular evolution experiments in situations where the substrate of interest is not a reasonable choice for large scale screening and investigates the tradeoffs in screening with alternative

substrates. Additionally, the ability of directed evolution experiments to alter the pH profile of enzyme variants was demonstrated, and these features further broaden the scope of the power of molecular evolution to provide catalysts meeting specific requirements.

## Materials and Methods

*pNB Esterase Expression System*

Plasmid pNB106R was provided by Steve Queener at Eli Lilly & Co. This plasmid contains the pNB esterase gene under the control of an altered λ promoter, *pL*106 (U.S. Patent Application No. 07,739,280) (Zock *et al.*, 1994). The plasmid also contains a temperature-sensitive λ CI repressor which inactivates the *pL*106 promoter below 35 °C. Further, the plasmid contains an *E. coli* origin of replication, a plasmid copy control gene, and a tetracycline resistance gene.

*Structural Modeling*

Homology searches and sequence alignments were performed at the California Institute of Technology's Sequence Analysis Facility using the GCG Sequence Analysis Software Package version 8.0 from the Genetics Computer Group (University Research Park, Madison, Wisconsin). BLAST searches of the Brookhaven Protein Data Bank (PDB), the SWISS-PROT database, the Protein Information Resource database, and the translated GenBank database were performed at the National Center for Biotechnology Information using a BLAST network service (Altschul *et al.*, 1990). PHYLIP was used at the same facility to construct evolutionary trees. The computer program MODELLER was used to align crystal structures and to calculate pNB esterase structures based on the crystal structure and sequence alignments (Sali and Blundell, 1993 and Sali and Overington, 1994). Protein modeling was performed using the Insight software package from Biosym Technologies for Silicon Graphics Inc. Indigo workstations.

*Activity of Commercial Enzymes towards LCN-pNB*

Acetylcholinesterase (*Electrophorus electricus*), butyrylcholinesterase (horse serum), cholesterol esterase (porcine pancreas), and elastase (EC 3.4.21.36, porcine pancreas) were purchased chromatographically pure from Worthington Biochemical

Corporation. Lipase (*Chromobacterium viscosum*) and subtilisin A (EC 3.4.21.62, *Bacillus licheniformis*) were purchased chromatographically pure from Calbiochem (Torrance, CA). These enzymes and pNB esterase were adjusted to 0.2 μmol/mL in 0.1 M PIPES pH 7.0. 10 μL of enzyme solution were added to 990 μL of a 0.125 mM LCN-pNB, 0.1 M PIPES pH 7.0, 20 mM $MgSO_4$ 20 mM $CaCl_2$ and 1% DMF solution in a quartz cuvette. The reaction was monitored at 289 nm for 2 min., and the change in absorbance vs. time data were used to determine reaction rates.

*Restriction of DNA*

Plasmid and fragment DNA when required were cut with Bam HI, Xba I and where indicated Xho I (Boerhinger Mannheim, Indianapolis, IN) in restriction buffer B at 37 °C for one hour. The resulting linear DNA was then run on a 1% agarose gel and separated into bands according to size. The appropriately sized band was excised from the gel and extracted using either the GeneClean (Bio101, La Jolla, CA) or Qiagen (Qiagen, Chatsworth, CA) method. In both cases purified DNA was eluted in Tris-EDTA buffer.

*Random Mutagenesis*

Random mutagenesis was performed in a 2000 base pair region surrounding the entire pNB esterase open reading frame (Leung *et al.*, 1989). Random mutagenesis primers 3'-GAGCACATCAGATCTATTAAC-5' (forward primer) and 3'-GGAGTGGC-TCACAGTCGGTGG-5' (reverse primer) were designed to complement regions of this plasmid just outside the restriction sites Xba I upstream of the open reading frame start and Bam HI downstream of the open reading frame end. A solution containing 1 mM dNTPs, Tris-HCl, β-mercaptoethanol, forward and reverse primers, plasmid pNBE, and Taq DNA polymerase (Perkin Elmer-Cetus, Emeryville, CA) in a total volume of 100 μL were covered with 2-3 drops of light mineral oil (Sigma, St. Louis, MO). The sample was then placed in a well containing 2-3 drops of mineral oil of a Precision Scientific (Chicago, IL)

thermal cycler for polymerase chain reaction (PCR). The thermal cycler repeats the following steps: 1 minute at 94 °C, 2 minutes at 42 °C, and 1 minute at 72 °C for 25 cycles. These conditions should generate an error frequency of approximately one substitution per 1000 bases, or approximately 1.5 substitutions per gene copy (Leung *et al.*, 1989). The fragment of DNA amplified by this technique was then subjected to a phenol/chloroform extraction and ethanol precipitation. The DNA was then restricted and purified as described above.

*Competent Cell Preparation*

Competent TG1 cells were prepared according to the $CaCl_2$ method (Sambrook). TG1 cells were grown overnight at 37 °C in a 3 mL culture of LB broth. The cells were diluted 1:200 in fresh LB and allowed to grow to an $OD_{600}$ of 0.35 to 0.40. They were placed on ice for 1 hour and spun at maximum speed in a 4 °C Beckman tabletop centrifuge. The cell pellet was resuspended in 0.5 volumes of 0.1 M $CaCl_2$ and allowed to sit on ice for 30 min. to 1 hour and recentrifuged as before. The cell pellets were resuspended in sterile 0.02 volumes of 0.1 M $CaCl_2$, 10% v/v glycerol and frozen at -70 °C until use.

*Ligation and Transformation*

Ligation reactions were performed using T4 DNA ligase (Boerhinger-Mannheim). Vector DNA (the entire pNB106R plasmid excluding the pNB esterase gene between Xba I and Bam HI), insert DNA (the pNB esterase gene between Xba I and Bam HI or where indicated the two fragments Xba I to Xho I and Xho I to Bam HI), 10X ligation buffer, water and enzyme were combined and incubated at 4 °C overnight (12-16 hours). The solution was incubated with previously prepared competent cells on ice for 1 hour. The cells were then heat shocked at 42 °C for 1 minute, supplied with an equal volume of LB

media, and incubated at 30 °C for 45 minutes. This solution was then plated onto LB plates containing tetracycline to 20 μg/mL (LB Tet plates).

*Screening*

TG1 transformants arising from ligations of pNBE vector and randomly mutagenized inserts were allowed to grow for 36 to 48 hours before shifting to 42 °C to induce expression of the pNB esterase gene. After an eight hour induction period, each colony was picked with a sterile toothpick and resuspended in a unique well of a 96 well plate containing 200 μL of 0.1 M Tris-HCl, pH 7.0. The turbidity of each well was measured as the absorbance at 620 nm adjusted by a cell-free reference well by a SInstruments 96 well plate reader. A 20 μL aliquot from each well was pipetted into a second 96 well plate, to which was added 200 μL of a substrate solution containing 0.8 mM para-nitrophenyl acetate (pNPA) and 0.4% (v/v) acetonitrile or 0.3 to 0.8 mM para-nitrophenyl loracarbef nucleus (LCN-pNP), 0.1M Tris-HCl pH 7.0, and between 0 and 30% v/v dimethylformamide (DMF - EM Science Guaranteed Reagent grade). The resulting reaction was monitored using the 96 well plate reader at 405 nm. Reactions were typically monitored for 11 data points varying from 15 seconds between data points for 0% DMF measurements to 180 seconds between data points for 30 % DMF. The slopes of the best-fit lines through the resulting 11 data points for each of the 96 wells were normalized by the corresponding absorbance at 620 nm measured previously. These normalized values were compared, and the wells exhibiting the highest activity to turbidity ratios were plated onto LB Tet plates. Two single colonies from these plates were restreaked on LB Tet plates to provide single colony isolates for further testing. Two single colonies from each of these second plates (four colonies total) were then arrayed onto LB Tet plates using sterile toothpicks. This collection of potential variants was then rescreened using the activity to turbidity ratio assay again. Of those that showed better activity to turbidity than wild type, the best three were chosen for 1 L culture and purification.

Crude screening was performed on the fourth generation variants using the LCN-pNB substrate (the actual substrate) using a similar whole cell assay. 100 µl samples of the resuspended colonies were removed from each well of the 96 well plate and added to a quartz cuvette containing a 1000 µl reaction solution consisting of 2.5% DMF, 0.1 M Tris-HCl pH 7.0, and 0.25 mM LCN-pNB. The absorbance at 291 nm of each sample was measured for 2.5 min. using a UV spectrophotometer. Initial rates were measured for both the LCN-pNP substrate in the 96 well plate assay (above) and the LCN-pNB substrate in quartz cuvettes. Both sets of slopes generated from the initial rate data were normalized to the turbidity measurements at 620 nm.

*Cell Culture*

Single colonies were inoculated into 5 mL LB Tet culture tubes and allowed to grow overnight at 30 °C. The contents of these tubes were then used to inoculate a one-liter culture of LB Tet and allowed to grow to maximum turbidity. These one-liter cultures were decanted into sterile centrifuge bottles and spun at 6000 rpms in a JA-10 rotor for 15 minutes in a Beckman centrifuge. The cell pellets were resuspended in LB Tet pre-warmed to 42 °C. The flasks were placed in a 42 °C incubator and allowed to shake for 8 hours (Zock *et al.*, 1994). The cells were harvested by similar centrifugation and resuspended in a centrifuge tube in 25 mL of Buffer A (Lysis Buffer), consisting of 10 mM potassium phosphate, 1 mM β-mercaptoethanol, and 0.5 mM EDTA, pH 7.0.

*Cell Lysis*

A French Press was used to lyse the harvested cells. The lysis was accomplished by placing the chilled sample into a steel housing, which was compressed to 20,000 atmospheres. A small needle valve was then opened and the cells were released to ambient conditions, causing the cells to rupture. This process was repeated three times to insure

complete lysis. The steel housing was kept chilled prior to use at 4 °C and the samples were stored before and after on ice.

*Purification*

After lysis the cell debris was pelleted by centrifugation at 12,000g in a JA-20 rotor for 15 min. at 4 °C (Chen *et al.*, 1995). The cell lysate supernatant was adjusted to pH 5.0 with HCl, and the newly formed precipitate was removed by centrifugation at 12,000g in a JA-20 rotor for 30 min. at 4 °C. The supernatant volume was measured, and ammonium sulfate was dissolved to 45% saturation at 0 °C. For reference, the ammonium sulfate saturation amount used for calculations was 41.22 g/100 mL solution at 0 °C. The solution was chilled to 0 °C on ice for 5 minutes and centrifuged in a JA-20 rotor at 12,000g for 30 min. at 4 °C. The supernatant was transferred to a new centrifuge tube, where ammonium sulfate was added to bring the final amount to 85% saturation at 0 °C. Centrifugation was performed as before, and the supernatant discarded. The pellet was redissolved in Buffer B (10 mM Tris- HCl, 50 mM NaCl, 1 mM β-mercaptoethanol, and 0.5 mM EDTA, pH 8.5), placed in an Amicon (Beverly, MA) spin filtration unit (Centricon - 10) and buffer exchanged three times with Buffer B to remove the ammonium sulfate. The resulting protein sample was applied to a DEAE-sepharose column (2.5 cm ID x 10 cm high) pre-equilibrated in Buffer B. The column was rinsed with buffer B until the baseline was restored. The column was then rinsed with buffer C (10 mM Tris-HCl, 50 mM NaCl, pH 7.0) until the pH reached 7.0. An NaCl gradient from 50 to 500 mM in buffer C (300 mls total volume) was passed through the column and fractions collected. Those fractions containing activity were pooled and then applied to an immobilized metal affinity chromatography (IMAC) column (2.5 cm ID x 10 cm, Fast-flow Chelating Sepharose, Pharmacia, Uppsala, Sweden) loaded with $Cu^{2+}$ and prepared as per the manufacturer's instructions. The column was first pre-treated by rinsing with three column volumes of (one column volume was approximately 50 ml) 0.5 M NaCl, 50 mM EDTA pH 8.5 to

remove all chelated metal ions, 2 M NaCl to remove any ionically bound material, and 1 M NaOH to remove any denatured protein. Copper as 100 mM copper sulfate in 100 mM sodium acetate pH 4.6 was loaded onto the IMAC column, washed with 20 mM sodium phosphate, 0.5 M NaCl, 50 mM imidazole pH 7.2 until pH 7.2, and finally equilibrated with five column volumes of 20 mM sodium phosphate, 0.5 M NaCl, 1 mM imidazole for sample loading. The sample was applied to the column and the column washed with the 1 mM imidazole solution until the baseline was restored. A linear gradient formed by 100 mls of 1 mM and 10 mM imidazole solutions (200 mls total volume) was applied, and fractions were collected. All tubes demonstrating higher than background activity were pooled, concentrated, and buffer exchanged into 0.1 M Tris-HCl, pH 7.0 in the Amicon Centricon-10 units as before.

*SDS-Page Gels*

SDS-Page gels were used to determine purity of protein solutions. Separating gels were made of 10% acrylamide and allowed to gel under butanol. After gelling the butanol was removed and a 4% acrylamide stacking gel was poured on top of the separating gel. Up to 5 μL of concentrated protein samples were mixed with 20 μL loading buffer (10% glycerol, 1% SDS, etc.) and boiled for 4 minutes. The 25 μL samples were loaded onto the gel and run at 200 V for approximately 30 minutes, at which time the loading buffer dye reached the bottom of the gel. The gel was removed from the apparatus and stained using a Coomasie blue stain solution. After staining a minimum of 45 minutes, the Coomasie blue stain was poured off and destain was added. This was allowed to incubate until the solution approximated the color of the gel, at which time the destain was poured off and new destain was added. The gel was then dried and sealed in plastic for further handling.

*Protein Concentration Assays*

Protein samples were assayed using the Bio-Rad (Hercules, CA) Protein Assay Reagent. The reagent was diluted 1:4 in water and filtered to remove any particulates. 20 μL of protein sample was combined with 980 μL of dilute reagent in a 2 mL spectrophotometer cuvette and allowed to incubate for 10-30 minutes. The sample's absorbance was then measured at 595 nm. A similarly treated series of standards consisting of dilutions of a purified and known concentration of wild type pNB esterase (supplied by Eli Lilly & Co.) was used to identify the protein concentration of the unknown sample.

*Kinetic Assays*

Kinetic assays were performed on three substrates: pNPA, LCN-pNP, and p-nitrobenzyl loracarbef nucleus (LCN-pNB). For pNPA and LCN-pNP substrates, final concentrations varying from 0.0625 mM to 16 mM and 0, 15, and 30% DMF in 0.1 M PIPES (Sigma), pH 7.0 were combined with equal volumes of enzyme samples. These samples were mixed simultaneously in a 96 well plate using an eight channel pipetteman and monitored with the 96 well plate reader. The absorbance values were recorded, initial slopes calculated by linear regression analysis, and used for calculating kinetic parameters. For the LCN-pNB substrate, final concentrations varying from 0.0156 mM to 8.0 mM and 0, 15 and 30% DMF in PIPES, pH 7.0 were combined with enzyme samples in a quartz cuvette and were measured in a spectrophotometer at 289 nm. All assays were measured in triplicate.

Assays on pNB containing substrates L-glutamate-pNB ester and p-nitrobenzyl benzoate were performed by adding a reaction mix containing 1.0 mM substrate in 1 to 20% DMF and 0.1 M phosphate buffer, pH 7.0 to a small volume of enzyme solution, incubating at room temperature for 20 to 60 minutes, and then stopping the reaction with an equal volume of acetonitrile. The samples were then injected into an HPLC containing a

C18 chromatography column (VYDAC). Reaction products were separated using a gradient between 95% 1 mM triethylamine pH 2.5/ 5% methanol and 100% methanol. The resulting peaks were monitored at 270 nm and recorded on an IBM PC data acquisition system. These peaks were then numerically integrated and used for comparison between enzyme samples.

*Sequencing*

Plasmid DNA was prepared using QIAGEN's plasmid prep kit. This DNA was sequenced at Eli Lilly & Co.

**References**

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Arnold, F. H. (1991). Metal-affinity separations - a new dimension in protein processing. Bio/Technology 9: 151-156.

Bomblies, L., Biegelmann, E., Doering, V., Gerisch, G., Krafft-Czepa, H., Noegel, A. A., Schleicher, M., and Humbel, B. M. (1990). Membrane-enclosed crystals in *Dictyostelium discoideum* cells, consisting of developmentally regulated proteins with sequence similarities to known esterases. J. Cell Biol. 110:669-679.

Brannon, D. R., Mabe, J. A., and Fukuda, D. S. (1976). De-esterification of cephalosporin *para*-nitrobenzyl esters by microbial enzymes. J. Antibiotics 29:121-124.

Cadwell, R. C. and Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. PCR Methods and Appl. 2:28-33.

Chen, K. and Arnold, F. (1991). Enzyme engineering for nonaqueous solvents - random mutagenesis to enhance activity of subtilisin-e in polar organic media. Bio/Technology 9:1073-1077.

Chen, K. and Arnold, F. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc. Natl. Acad. Sci. USA 90:5618-5622.

Chen, Y., Usui, S., Queener, S. W., and Yu, C. (1995). Purification and properties of *p*-nitrobenzyl esterase from *Bacillus subtilis* . J. Ind. Micro. 15:10-18.

Cooper, R. D. G. (1992). The carbacephems: a new beta-lactam antibiotic class. Am. J. Med. 92:6A/2S-6A/6S.

Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., and Doctor, B. P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. Protein Science 2:366-382.

Eckert, K. A. and Kunkel, T. A. (1991). DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Applic. 1:17-24.

Felsenstein, J. (1989). PHYLIP -- Phylogeny Inference Package (version 3.2). Cladistics 5:164-166.

Harel, M., Sussman, J. L., Krejci, E., Bon, S., Chanal, P., Massoulie, J., and Silman, I. (1992). Conversion of acetylcholinesterase to butyrylcholinesterase: Modeling and mutagenesis. Proc. Natl. Acad. Sci. USA 89:10827-10831.

Hasinoff, B. B. (1982). Kinetics of acetylcholine binding to electric eel acetylcholinesterase in glycerol/water solvents of increased viscosity. Biochim. Biophys. Acta 87: 955-975.

Hwang, C-S. and Kolattukudy, P. E. (1993). Molecular cloning and sequencing of thioesterase B cDNA and stimulation of expression of the thioesterase B gene associated with hormonal induction of peroxisomal proliferation. J. Biol. Chem. 268:14278-14284.

Jbilo, O. and Chatonnet, A. (1990). Complete sequence of rabbit butyrylcholinesterase. Nucleic Acids Res. 18:3990.

Jbilo, O., L'Hermite, Y., Talesa, V., Toutant, J. P., and Chatonnet, A. (1994). Acetylcholinesterase and butyrylcholinesterase expression in adult-rabbit tissues and during development. Eur. J. Biochem. 225:115-124.

Johnson. M. S. and Overington, J. P. (1993). A structural basis for sequence comparisons: an evaluation of scoring methodologies. J. Mol. Biol. 233:716-738.

Kaiser, R., Erman, M., Duax, W. L., Ghosh, D., and Joernvall, H. (1994). Monomeric and dimeric forms of cholesterol esterase from *Candida cylindracea*. Primary structure, identity in peptide patterns, and additional microheterogeneity. FEBS Lett. 337:123-127.

Leung, D. W., Chen, E., and Goeddel, D. V. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1:11-15.

Lotti, M., Grandori, R., Fusetti, F., Longhi, S., Brocca, S., Tramontano, A., and Alberghina, L. (1993). Cloning and analysis of *Candida cylindracea* lipase sequences. Gene 124:45-55.

Ozols, J. (1989). Isolation, properties, and the complete amino-acid sequence of a 2nd form of 60-kDa glycoprotein esterase - orientation of the 60-kDa proteins in the microsomal membrane. J. Biol. Chem. 264:12533-12545.

Pohlenz, H. D., Boidol, W., Schuttke, I., and Streber, W. R. (1992). Purification and properties of an *Arthrobacter oxydans* p52 carbamate hydrolase specific for the herbicide phenmedipham and nucleotide-sequence of the corresponding gene.
J. Bact. 174:6600-6607.

Radic, Z., Pickering, N. A., Vellom, D. C., Camp, S., and Taylor, P. (1993). Three distinct domains in the cholinesterase molecule confer selectivity for acetyl- and butyryl-cholinesterase inhibitors. Biochemistry 32:12074-12084.

Sali, A. (1995). Modelling mutations and homologous proteins. Curr. Opin. Biotech. 6:437-451.

Sali, A. and Blundell, T. L. (1993). Comparative modelling by satisfaction of spatial restraints. J. Mol. Biol. 234:779-815.

Sali, A. and Overington, J. P. (1994). Derivation of rules for comparative modeling from a database of protein structure alignments. Protein Science 3:1582-1596.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, pp. 1.82-1.84.

Schrag, J. D. and Cygler, M. (1993). 1.8 angstroms refined structure of the lipase from *Geotrichum candidum*. J. Mol. Biol. 230: 575-591.

Shibata, F., Takagi, Y., Kitajima, M., Kuroda, T., and Omura, T. (1993). Molecular cloning and characterization of a human carboxylesterase gene. Genomics 17:7-82.

Srinivasan, N. and Blundell, T. L. (1993). An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. Protein Eng. 6:501-512.

Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., and Siliman, I. (1991). Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. Science 253:872-879.

Sussman, J. L., Harel, M., and Silman, I. (1993). Three-dimensional structure of acetylcholinesterase and of its complexes with anticholinesterase drugs. Chem. Biol. Interactions, 87:187-197.

Todd, R. J., Johnson, R. D., and Arnold, F. H. (1994). Multiple-site binding interactions in metal-affinity chromatography. 1. Equilibrium binding of engineered histidine-containing cytochrome-c. J. Chromat A 662:13-26.

Taylor, P. and Radic, Z. (1994). The cholinesterases: from genes to proteins. Annu. Rev. Pharmacol. Toxicol. 34:281-320.

You, L. and Arnold, F. H. (1996). Directed evolution of *Subtilisin E* in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. Prot. Eng. 9:77-83.

Zock, J., Cantwell, C., Swartling, J., Hodges, R., Pohl, T., Sutton, K., Rosteck Jr., P., McGilvray, D., and Queener, S. (1994) The *Bacillus subtilis* pnbA gene encoding p-nitrobenzyl esterase - cloning, sequence and high-level expression in *Escherichia coli.* Gene 151:37-43.

# Chapter 3

# Sexual Recombination

**Introduction**

Reproduction is essential for the propagation of all living things and occurs either sexually, through the combining of genetic information from two parents to produce new progeny, or asexually, through the formation of progeny directly from a single parent. In sexual reproduction, genetic recombination is the process by which parts or all of DNA molecules from two separate sources are exchanged or brought into a single unit (Brock et al., 1994, p. 239). The purpose of this recombination can serve several beneficial functions: to combine positive traits of two parents into one offspring, to eliminate the passage of a negative trait in one parent to progeny by replacing the negative trait with the normal trait of a second parent; and to maintain a large potential genetic variability in a population which contains only a small amount of variability in each individual, as evolutionary insurance against a changing environment (Crow, 1988). This natural technique reduces the time required for a population to become evolutionarily "fit" to a few generations as opposed to the large number of generations required if such positive trait information were not shared (Maynard Smith, 1988, Muller, 1932).

The process of combining DNA from two separate locations within a single organism has important effects on evolution and arises as a result of recombination processes similar to those in sexual recombination. These effects include the regulation of expression of DNA (by changing the location and orientation of a gene in a chromosome), the generation of new combinations of genes (for instance, keeping many genes in a pathway linked so that they are expressed together) (Stryer, 1988, p. 687), and the generation of new protein functions through gene duplication (Hames and Glover, 1990). Gene duplications are a result of recombination and allow for one of the duplicate genes to undergo mutation without selection (the original gene still functions normally). As the duplicate undergoes genetic drift, it accumulates multiple mutations and can acquire new functions. An example of gene duplication is seen in the hemoglobin gene family, where the standard hemoglobin protein contains two $\alpha$ and two ß subunits. In two separate

instances in human evolution, the ß gene was duplicated, and the genes diverged. What resulted were the γ and ε genes which replace the ß subunit in hemoglobin during embryonic and fetal developments (Rawn, 1989).

Directed evolutionists have used studies in natural evolution processes to guide efforts to tailor enzymatic properties in the laboratory. One of the successful approaches inspired by natural evolution has been random mutagenesis and selection (or screening) for desired functional properties. Enzymes with enhanced properties have been evolved in the laboratory using mutation-induced DNA replication and screening of the resulting protein variants. Enzymes more active on non-natural substrates and/or in organic solvent (Arnold, 1993, Chen and Arnold, 1993, Moore and Arnold, 1996), proteins more stable to denaturation (Berkenpas et al., 1995, Okada et al., 1995, Rellos and Scopes, 1994, Yabuta et al., 1995), and changes in substrate specificity (Gulick and Fahl, 1995, Light and Lerner, 1995) have all benefited from imitating evolution's mutation and selection process. In particular, the enzymes in our lab (Chen and Arnold, 1993, Moore and Arnold, 1996) show activities on non-natural substrates and/or under non-natural conditions two to three orders of magnitude greater than the wild type enzyme's and thus demonstrate the power of imitating this part of the natural evolution process.

As outlined above, recombination is also a powerful force in natural evolution. Recently, a method for conveniently accomplishing sexual recombination *in vitro* has been described under the name 'DNA shuffling' or 'sexual PCR.' This method expands the tools by which directed evolution can be performed (Stemmer, 1994b). The DNA shuffling technique is outlined in Figure 1. Initially, the DNA from a collection of enzyme variants to be shuffled is amplified by PCR. The resulting DNA is randomly digested into a pool of small fragments by DNase I. A PCR-based reassembly step takes advantage of the overlapping regions of sequence identity between adjacent fragments to direct in primer/template fashion the synthesis of complete sequences. This is the step which performs the recombination as fragments with and without mutations serve as templates for

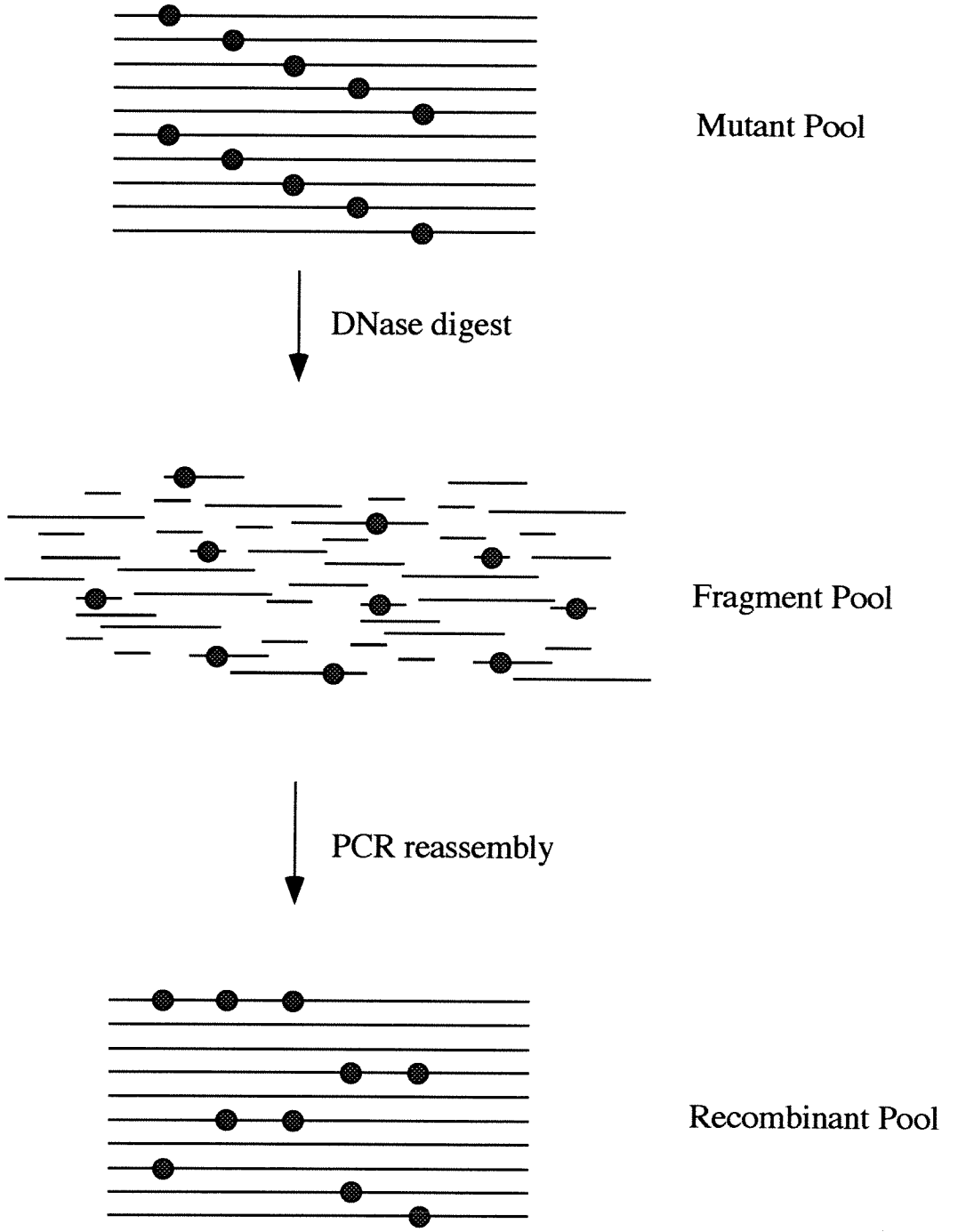



Figure 3.1: The DNA shuffling procedure. DNA containing mutations is digested into small fragments by DNase I. The resulting fragments are reassembled using PCR methods to generate a pool of recombinant clones.

other priming fragments containing or lacking mutations. A final PCR amplification step provides large quantities of the recombinant pool for cloning. This technique is particularly useful because sequence information is not required to generate recombinant clones. Most of the directed evolution efforts performed using DNA shuffling involve the use of natural selection to find positive variants. Where selection could be employed in testing this methodology, such as the laboratory evolution of antibiotic resistance, huge progress has been made in enzymatic activity, as represented by a 16,000 fold increase in the amount of antibiotic required to inhibit bacterial growth (Stemmer, 1994). This is a direct result of the extraordinarily large diversity this process can create and natural selection can search ($10^9$ or more different bacteria per experiment).

This recombination method differs from the natural case of sexual recombination in two distinct features. First, the *in vitro* method allows for sexual recombination between more than two parents. This allows for the discovery of the best combinations of mutations and does not assume that the best combination contains all the mutations in a population. Second, it relies on DNA polymerase to generate a substantial number of full length genes at the reassembly step. An effect of forcing the standard Taq DNA polymerase to synthesize full length genes from the small fragment DNA pool is an additional background mutagenesis rate. This mutagenesis resulted in genetic variability, the variability on which sexual recombination acted in order to enhance antibiotic resistance (Stemmer, 1994b).

Screening of enzymatic properties (as opposed to selection) is a more general approach to directed evolution, as many interesting enzyme properties are not directly selectable. However, screening processes search genetic diversity much more slowly ($10^3$ to $10^4$ different bacteria per experiment), and this creates new problems for the use of recombination. We do not know whether DNA shuffling is an efficient way to recombine positive mutations when screening is required to find positives. We also need to know how the sexual recombination of several parents differs from that of two parents. Are there

approaches to optimize DNA shuffling populations in order to more efficiently create positive variants? Is the mutagenic rate associated with DNA shuffling too large for efficient use when populations are screened?

We have chosen to examine some of these issues using a population of pNB esterases generated previously (Moore and Arnold, 1996). Random mutagenesis and the screening of 7500 clones from the fourth generation of directed evolution provided five positive variants appearing at least 50% more active than the parent clone (labeled 3-10C4) on the man-made substrate p-nitrobenzyl loracarbef nucleus (LCN-pNB). Fifteen variants (including the five above) showed at least 20% improvement in activity. An earlier generation (generation 2) also generated two positive variants. These populations were sexually recombined in order to investigate the issues listed above and to generate further increases in pNB esterase activity.

**Recombination Constraints**

Recombination is a distinctly different process from random mutagenesis, and has its own advantages and disadvantages. Recombination requires a collection of sequences containing mutations in order to operate, and this requirement is a disadvantage when screening is required. To generate a population for recombination requires finding a population of positive clones. Screening requires that each clone be examined one at a time with the consequence that the positive clones are found sequentially. Ideally, all the positive mutations will complement one another, such that the variant containing all of the mutations will be the most useful (cumulative or additive effects). Since the positives are found sequentially, one could easily imagine making the mutations sequentially (i.e., find positive variant A, make mutations to A, find positive AB, make mutations to AB, find ABC, etc.). Finding ABC in this example requires screening the same number of colonies as finding three single mutation sequences A, B, and C. Recombining the A, B, and C population in the effort to generate the ABC variant requires additional screening. As an

example, in the pNB esterase screening, approximately 1 in every 1500 clones was a positive variant (50% or greater improvement). In order to generate a population of three single mutants by screening, the number of clones needed to be examined is 3 times 1500, or 4500. These clones could have been created in sequential fashion (screening 1500 clones to find a single mutant A; mutating this variant and screening another 1500 clones to find a double mutant AB; and finally mutating this double mutant and screening a final 1500 clones) to generate the triple mutant, ABC. This triple mutant would require the same 4500 screened clones that finding the three single mutants require. The three single mutants still need to undergo the recombination process and screening to generate and find the desired triple mutant.

A second major disadvantage to DNA shuffling is the relative frequency of wild type fragment and mutation-containing fragment sequences, which compete with each other for appearance in progeny sequences. The three single mutant population in the previous example implicitly contained three unique variants. Therefore, the mutation in the sequence of variant 1 responsible for the improved property is not found in variants 2 and 3. More importantly, this strongly suggests that variants 2 and 3 contain the wild type sequence at the location of the mutation in variant 1. This argument holds true irregardless of which variant is considered, with the consequence that when these sequences are shuffled, a mutation will be incorporated into a progeny sequence only in one-third of the cases. In the best case, the mutations are distributed far from one another on the gene so that recombination occurs often between any two mutations. Then the appearance of any mutation at any location in progeny sequences will be random. Figure 2 demonstrates this situation and the resulting consequences. The probability of generating a wild type sequence is two in three at each position, corresponding to an overall probability of generating a wild type sequence of $(2/3)^3$ or 30%. The probability of generating a single mutant sequence is one in three (the probability of a mutation at position 1) times two in three (the probability of no mutation at position 2) times two in three (the probability of no

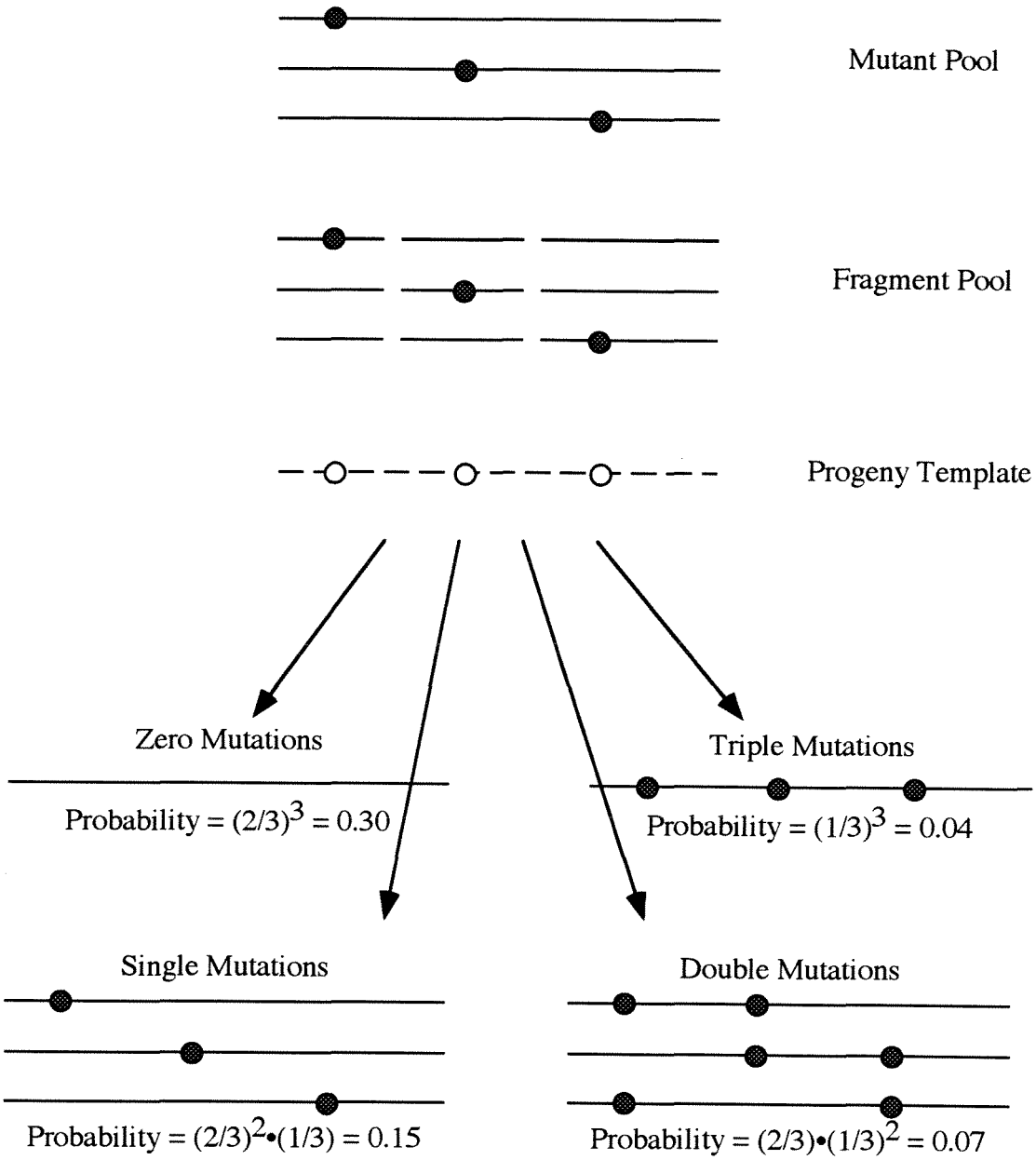



Figure 3.2: The possible combinations of progeny and their associated probabilities when shuffling three single mutation parents. At the location of each mutation in the mutant pool, a number of sequences contain the wild type nucleotide. When reduced to fragments in the fragment pool, these sequences compete for occurrence in the progeny. The possible combinations and their associated probabilities are shown.

mutation at position 3), or 15%. Because there are three such sequences, the probability of generating a single mutant sequence is 45%. The effect on screening the resulting recombinants is that 75% of the clones screened are not novel - they are variants already in the evolutionist's possession. Additionally, the triple mutant occurs only 4% of the time, and this percentage rapidly decreases as the number of mutations and sequences increase. Figure 3 demonstrates a continuation of this analysis for an increasing number of single mutation parent sequences. From this graph, the probability that screening progeny will return either the zero mutation 'grandparent' or single mutation parent sequences remains constant between 73% and 75%, so that in each case only 25% of the clones screened result in new variants. The ability to find a particular progeny sequence decreases rapidly as the number of parents increases, as shown in Figure 4. Together, Figures 3 and 4 may illustrate one reason why nature has gone to great lengths to insure that only two parents are responsible for the genetic composition of progeny. It is also the main reason why selection (rather than screening) of recombinant populations has dominated the use of this technique: *recombination tends to generate large amounts of parent and grandparent sequences.*

This apparent disadvantage of recombination is also an advantage under certain conditions. Often, multiple mutations are discovered in positive variants generated by error-prone PCR. These positive variants are believed to contain one mutation responsible for improved behavior and one or more extraneous mutations which may be neutral or deleterious. Extraneous mutations can have negative effects on properties not explicitly required under the screening or selection conditions. For instance, pNB esterase was screened for increased activity on p-nitrophenyl loracarbef nucleus (LCN-pNP) in dimethylformamide (DMF). The screening was performed at 30 °C. The original wild type enzyme denatures at temperatures above 55 °C, and we can imagine that variants with enhanced activity containing multiple extraneous mutations could very easily be less thermostable. DNA shuffling with an excess of wild type DNA, a technique known as

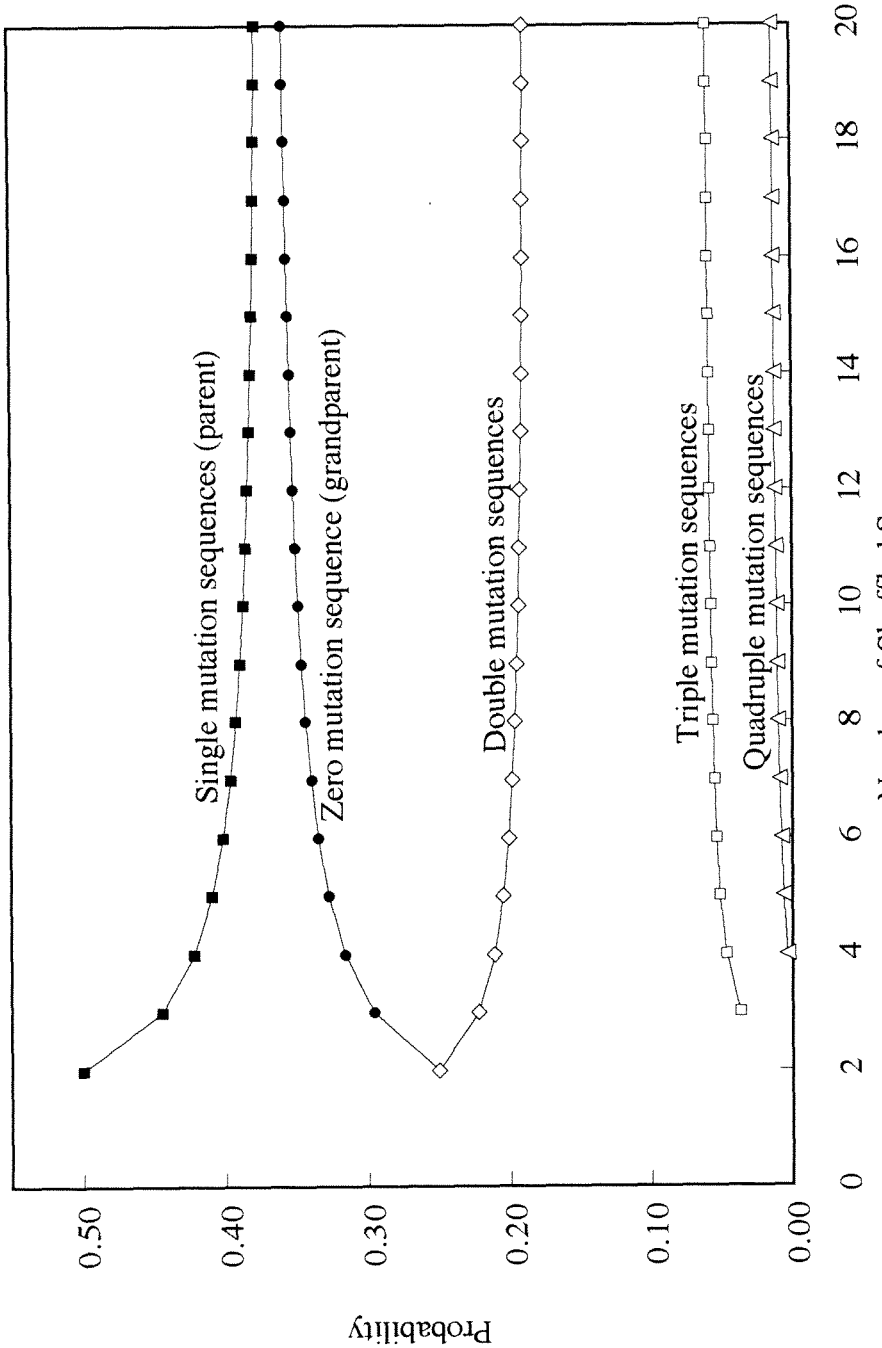



Figure 3.3: Probabilities of generating sequences containing different numbers of mutations based on shuffling a given number of equally distributed single mutation sequences. The unfilled symbols are novel variants arising as a result of the shuffling procedure.

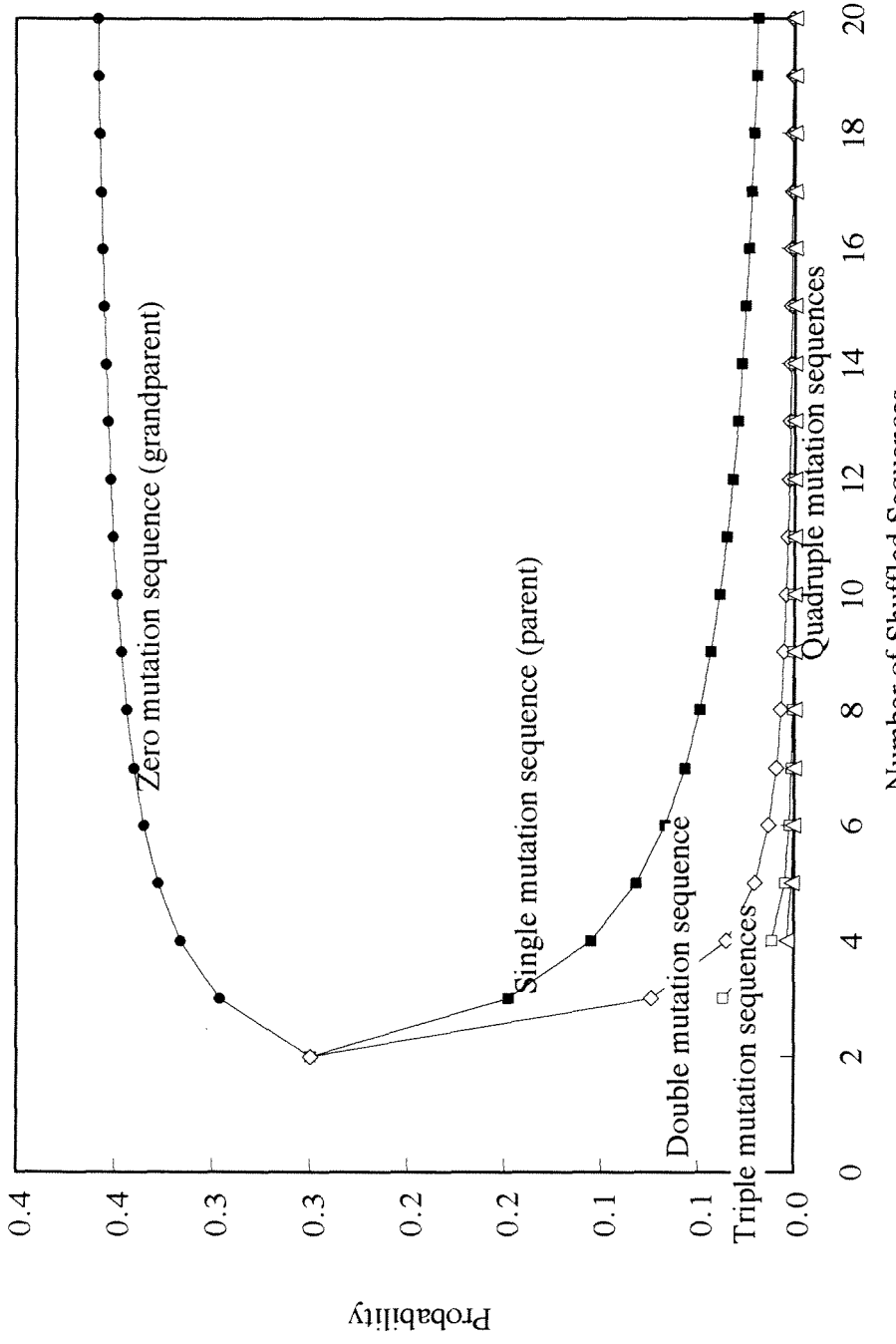



Figure 3.4: Probabilities of generating an individual sequence containing a particular number of mutations based on shuffling a given number of equally distributed single mutation sequences. The unfilled symbols are novel variants arising as a result of the shuffling procedure.

back-crossing, could remove extraneous mutations which survive the primary screen (Stemmer, 1994a).

Recombination also has the advantage of allowing parallel processing in screening. Several screens could be processed simultaneously for searching a variety of different criteria. One screen could be examining mutations which provide enhanced activity, while other screens could examine improvements in stability, solvent tolerance, protein expression or other properties. Positives could then be recombined to search for variants improved in several properties at once. Additionally, screening conditions can be difficult to optimize, especially when they are operating at or near conditions of rapid changes in enzymatic properties, such as the melting temperature. Screening for enhanced thermostability using sequential mutagenesis and screening requires altering the screening conditions after the discovery of each thermostable variant. A new temperature at which the newly found thermostable variant is inactivated must be identified. The ability to recombine mutations randomly allows for the discovery of several variants at one well-researched condition prior to combining the effects of multiple mutations and thus provides more environmental control during screening.

Recombination examines additivity between mutations more efficiently than point mutagenesis. Beginning with a pool of single mutant clones, recombination will generate approximately 75% uninteresting variants. However, approximately 20% of the clones will be double mutants. The best random point mutagenesis approach yields about 27% to 37% double mutants as a result of the Poisson distribution of sequences containing mutations (Cadwell and Joyce, 1992). Due to the degeneracy of the genetic code, approximately 40% of these are silent mutations, lowering the effective rate to that achieved by recombination. Additionally, recombination reduces the population size searched under these conditions to the likely location of the most interesting double mutants. For example, pNB esterase has approximately 9300 unique single mutants (19 times 489 amino acids). If we assume that a population consisting of the 100 most interesting clones has been

provided by random mutagenesis and screening, then randomly mutagenizing each variant to provide all the unique double mutants creates a library with 920,000 unique members. Recombination of the 100 most interesting variants requires searching a library of only 4950 to find the most interesting double mutants. Assuming the mutations are complementary, if not strictly additive, then this library is the most interesting.

*Some Examples*

When considering a single recombination event, such as when two variants are cut with a restriction enzyme and re-ligated randomly, two mutations are being sorted in the production of progeny. This is exactly analogous to the Mendelian genetics model where two parents which contain different alleles for two genes on separate chromosomes (so that they sort independently) are crossed or mated. Four unique genotypes are available to the progeny of this mating, and each occurs with equal probability, 25%. The genotype of particular interest is the double mutant, which should occur 25% of the time.

DNA shuffling can be performed on both the 5 (greater than 50% enhancement) and 15 (greater than 20% enhancement) variant populations. The probabilities for recombination with 5 or 15 parents were never considered in Mendelian genetics but are important to consider in the directed evolution of pNB esterase. Using Figure 3 and the five variant population, the probability of generating any parent sequence is 40%, the zero mutation grandparent is 33%, any double mutant is 20%, any triple mutant is 5%, any quadruple variant is 0.6% and the quintuple mutant is 0.03%. One in 3125 clones resulting from shuffling these five parent sequences should be the quintuple mutant. At the same time, Figure 4 shows the probability of finding any given mutant. For the five variant case, the probability of finding the zero mutation grandparent is 33%, a particular parent (of which there are five) is 8%, a particular double mutant (of which there are ten) is 2%, a particular triple mutant (of which there are ten) is 0.5%, a particular quadruple mutant (of which there are five) is 0.1%, and the same 0.03% for the unique quintuple mutant. These

two graphs together suggest that in searching for the quintuple mutant, the whole collection of variants has been examined. Therefore, if the most active variant arising from any combination of the mutations is not the quintuple mutant, this most active variant will be discovered by screening enough clones to ensure the screening of the quintuple mutant.

When we talk about the number of clones required for screening, it is simply the inverse of the probability. In practice, however, a certain amount of oversampling is required to ensure that a particular variant has been examined during the course of screening. In this case only the most active variant must be isolated, so to be 95% confident this variant has been examined, we must be 95% confident the most infrequently occurring variant has been examined. This requires oversampling the number of clones by 2.7 times (in the 5 variant case). In general, defining N to be the number of clones sampled and P to be the probability of generating a particular variant, then

$$(1 - P)^N < 1 - \text{confidence limit}$$

describes how the probability of not sampling a particular variant changes with increasing number of samples, and how many samples are required until the probability falls below the requirement set by the confidence limit. The oversampling is then defined by how many more samples must be screened over the theoretical minimum, or $\frac{N}{1/P}$. For example, the probability of generating the double mutant when recombining two single mutation parents is 1 in 4, suggesting that in the examination of 4 samples, the double mutant will be examined once. To be 95% confident that the double mutant has been examined requires sampling $(1 - 1/4)^N < 0.05$, (i.e., $N > 10.4$) or 11 clones. Thus the amount of oversampling is $^{10.4}/_{4}$ or 2.6. When only one clone is required with 95% confidence, the oversampling will be between 2.6 (for the shuffling of two parents) and 3.0 (for the shuffling of very large numbers of parents).

In an attempt to reduce the number of clones for screening, alternative recombination strategies were examined. An approach to producing the quintuple mutant

which requires the screening of far less is multiple step pairwise recombination. This strategy is most similar to the Mendelian approach and is outlined as follows: parent 1 and 2 are mated, as are 2 and 3, 3 and 4, and 4 and 5 (Figure 5). As each progeny is a double mutant 25% of the time, 16 clones are required to find all the double mutants. The double mutants are then similarly mated, requiring 12 clones to generate the triple mutants. Eight clones are required to generate the quadruple mutants, and four clones are required to generate the quintuple mutant. A total of 40 clones must be screened in order to generate the quintuple mutant. The cost of this method is the limited size of the space searched. Each mutation must contribute to the overall activity (not necessarily in additive fashion, but certainly in a cumulative fashion), so that the quintuple mutant is the best performer of this population. If a triple mutant is the best performer, it may not be found, since only three of the ten triple mutants will actually be examined. However, by organizing the search in this way, properties that are often improved by the introduction of independently additive mutations, such as enzyme stability, can be found by screening a minimum number of clones.

A compromise method that works well analytically but may have practical problems can be described as 'population recombination.' The idea behind this method is to shuffle all five variants and screen enough clones to see all the double mutants. From Figure 4, each double mutant occurs 2% of the time, so 50 clones are required for screening. This examines all of the pair-wise interactions between the variants and eliminates those which are not cumulative. In the case where they are all additive, the double mutant population is recombined to produce all of the triple mutants, and generating this population requires 44 clones. These would be recombined to produce all of the quadruple mutants and the quintuple mutant; generating this population requires 20 additional clones. If all of the mutations were in fact complementary, 114 clones would be needed to search the space completely. This approach is the closest to the way those that perform selection of recombination experiments operate (Stemmer, 1994b). Generally, all of the clones that

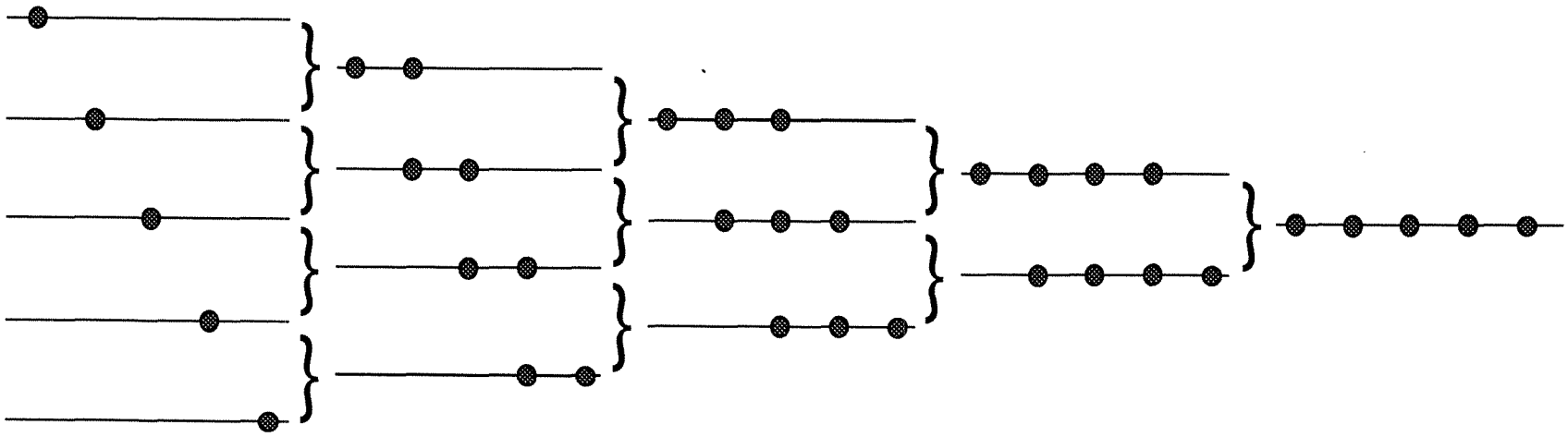




Figure 3.5: Pairwise DNA shuffling provides reduced screening requirements. By shuffling only two mutations at a time, the resulting two mutation progeny is generated in 25% of the cases. The schematic outlined here has 10 such recombination experiments for a total of 40 variants required to generate the 5 mutant sequence.

survive a particular selection criterion are recombined (often 100 clones or more serve as the parent population for the next generation) (Stemmer, 1994b).

**Results and Discussion**

From the fourth generation of pNB esterases, 15 clones demonstrating at least 20% improvement over the 3-10C4 parent were isolated. Five of these clones were at least 50% better than 3-10C4. We used the recombination technique of DNA shuffling to determine if combinations of these variants would produce variants more active than the most active variant 4-54B9 from generation four. As part of this experiment, a simple restriction and ligation recombination was performed between 4-54B9 and the remaining four clones with greater than 50% improvement. The result was clone 5-1A12 reported previously, whose activity was twice 4-54B9. Interestingly, the restriction/ligation experiment involves three mutations: I60V in one parent and K267R and L334S in the other parent, 4-54B9. The restriction/ligation occurred between the K267R and L334S mutations, and the resulting, more active enzyme contains I60V and L334S, while removing K267R. Thus in this approximate sexual recombination experiment, we saw both the joining of positive mutations and the elimination of neutral or negative mutations.

An unsuccessful attempt at DNA shuffling was initially made using the more active variants from the second generation, 2-13F3 and 2-19E10 (Moore and Arnold, 1996). The sequences of these variants were not known at the time; as it turned out, these variants shared a single mutation (M358V) responsible for the increased activity. The shuffling of DNA sequences with themselves has been reported (Stemmer, 1994). The similar sequences should not affect the shuffling procedure; they should only affect the screening results. The three-part DNA shuffling procedure was carried out as described in the Materials and Methods section. First, the digestion times with DNase I were optimized, as seen in Figure 6. The DNA fragment begins as a 2 kb band. Incubation with DNase I for increasing times produces fragments of ever-shorter average length. The digestion

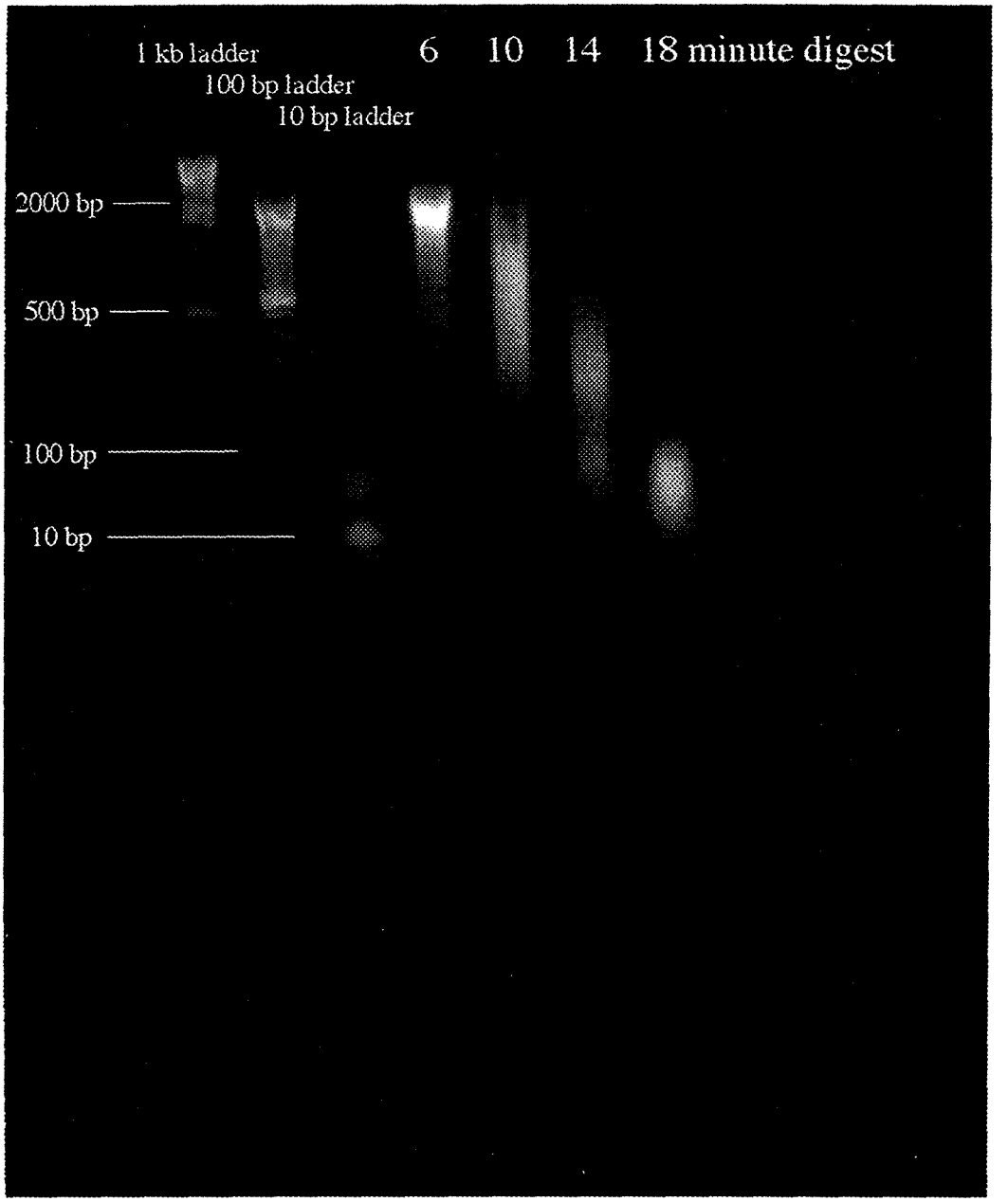



Figure 3.6: Gel electrophoresis of DNA samples removed from a DNase I digest reaction at various times. The initial 2 kb band is the predominant product of the 6 min. digest. After incubating with 0.015 units of DNase I at 25 °C for 18 min., 2-3 $\mu$g of 2 kb starting material has been digested to fragments less than 100 bp.

times were then utilized to produce the maximum amount of fragments in a size range of interest, such as the 75 to 200 base pair fragments isolated in Figure 7. Altogether, fragment sizes in ranges from 50 to 200, 100 to 300, 300 to 600, and 800 to 1200 base pairs were isolated for reassembly. The PCR reassembly of the fragments isolated resulted in either DNA fragments of increasing length, as seen in Figure 8a, or in DNA assemblies so large that they do not enter the gel, as seen in Figure 8b. Figure 8b shows the final result of continuing the reassembly process from Figure 8a. Most reassemblies were stopped when they surpassed the 2 kb size of the original DNA. The last step of this DNA shuffling technique involves using these reassembly products as templates for PCR reactions. An aliquot of each diluted 1:20, 1:40 and 1:100 with PCR amplification solution and amplified for 25 PCR cycles resulted in an amplification of exactly the image in Figure 8a or b. Restriction of this result with the Xba I and Bam HI enzymes produced no identifiable band useful for cloning. Additional experiments altering the level of $MgCl_2$ from 0.5 mM to 2.5 mM had no positive effect in the generation of product bands.

Similar results were initially seen when shuffling of the fourth generation variants was attempted. Figure 9 demonstrates the purification of 125 to 350 base pair fragments from a DNase I digest of a pool of the five best variants. Similar purification of a pool of the best 15 variants was also performed. The reassembly of three different DNA concentrations of each of these small fragment pools is shown in Figure 10a. Starting from the 125 to 350 base pair size, the reassembling of DNA fragments to well over the 2 kb gene size can clearly be seen. Figure 10b shows the result of using increasing cycles of these reassemblies as template for the final PCR step. Rather than producing a DNA product of 2 kb, the final PCR step simply amplifies the smear evident in Figure 10a.

The solution to performing DNA shuffling on this gene turned out not to be related to the size of DNA fragments used, the concentration of DNA fragments, the concentration of $MgCl_2$ or any of the thermocycler conditions, but, rather, the use of alternate primers listed in the Materials and Methods section. (The positions of the primers are shown in the

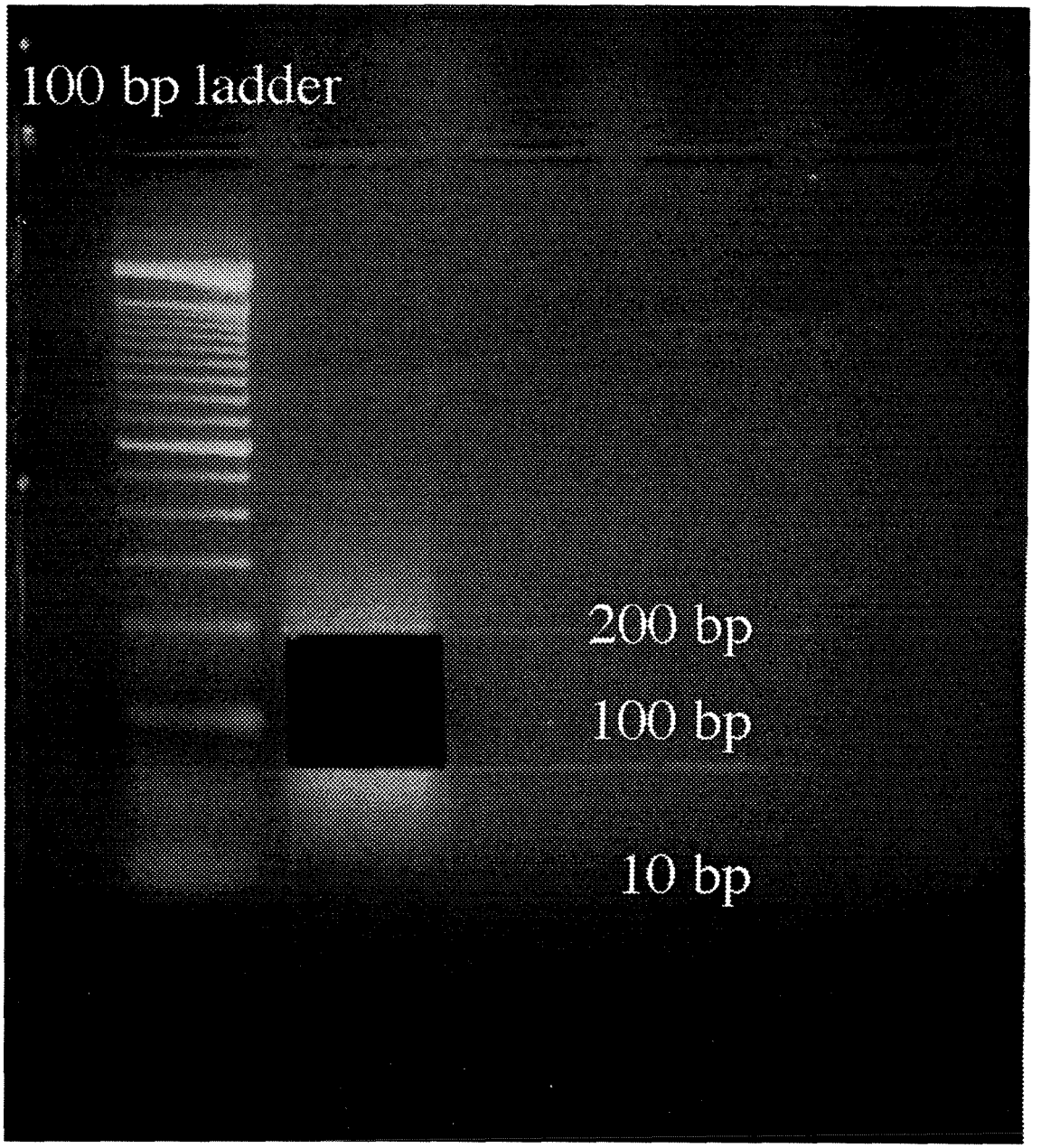



Figure 3.7: Gel electrophoresis of a DNase I digest reaction of second generation variants. DNA fragments between 75 and 200 bp have been isolated.

8a)


8b)

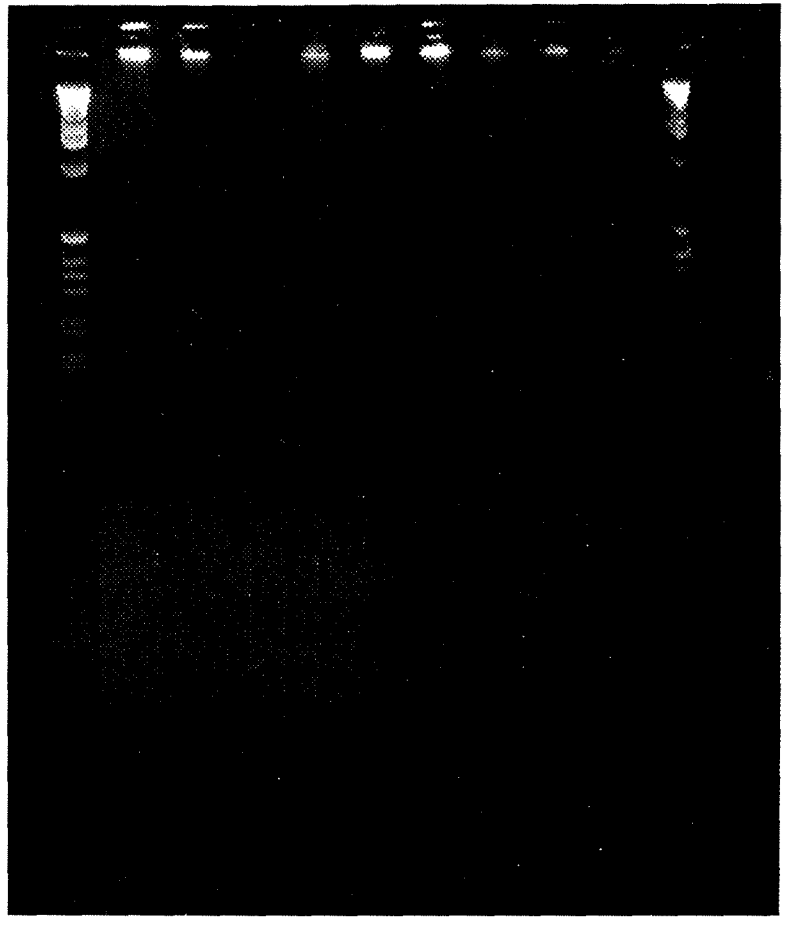


Figure 3.8: Gel electrophoresis of DNA samples resulting from reassembly reactions. 5a shows typical reassembly results from various concentrations of DNA fragments in both the 5 and 15 variant pools. 5b demonstrates common, although less frequent reassembly results from similar DNA fragment pools.

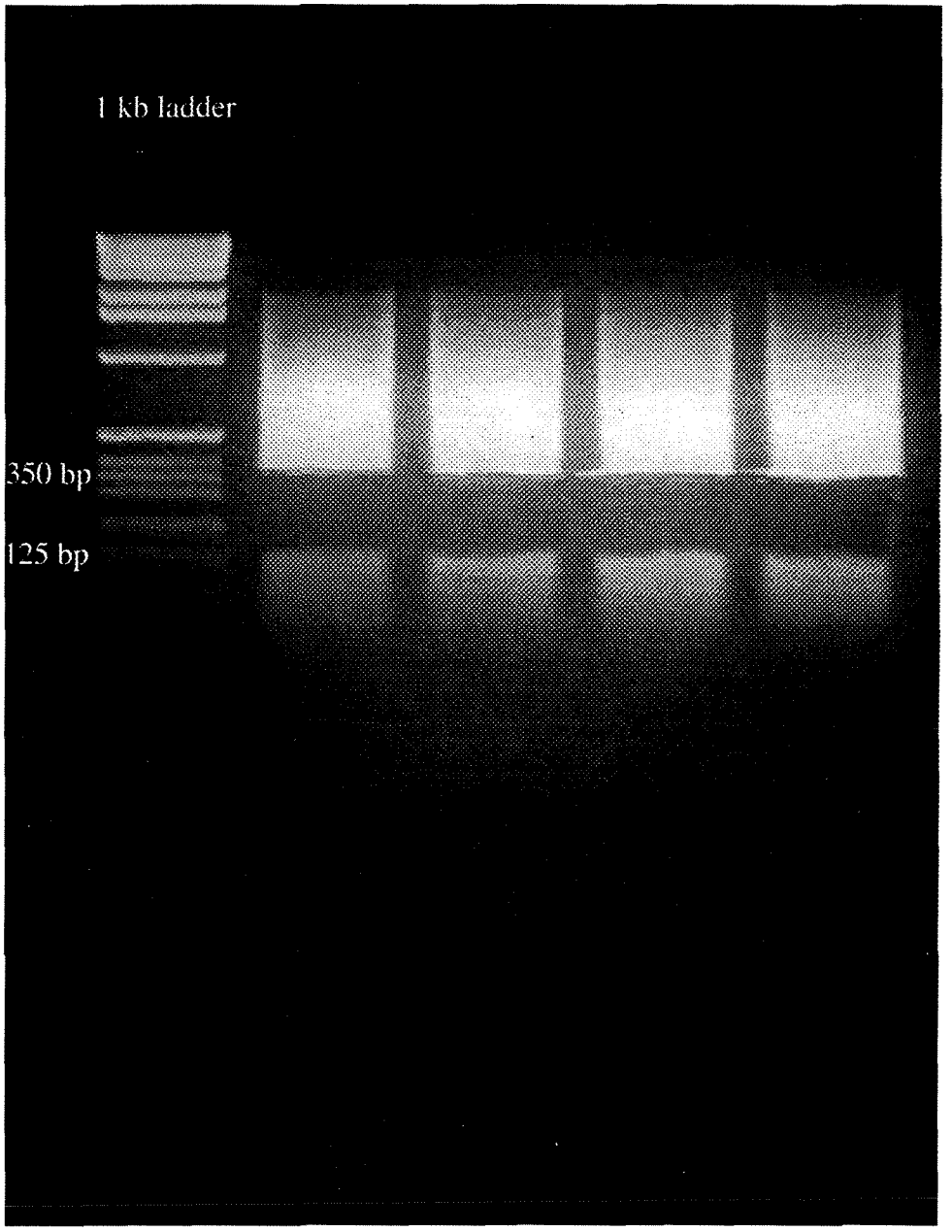



Figure 3.9: Gel electrophoresis of a DNase I digest reaction of the five best fourth generation variants. DNA fragments from 125 to 350 bp were isolated.

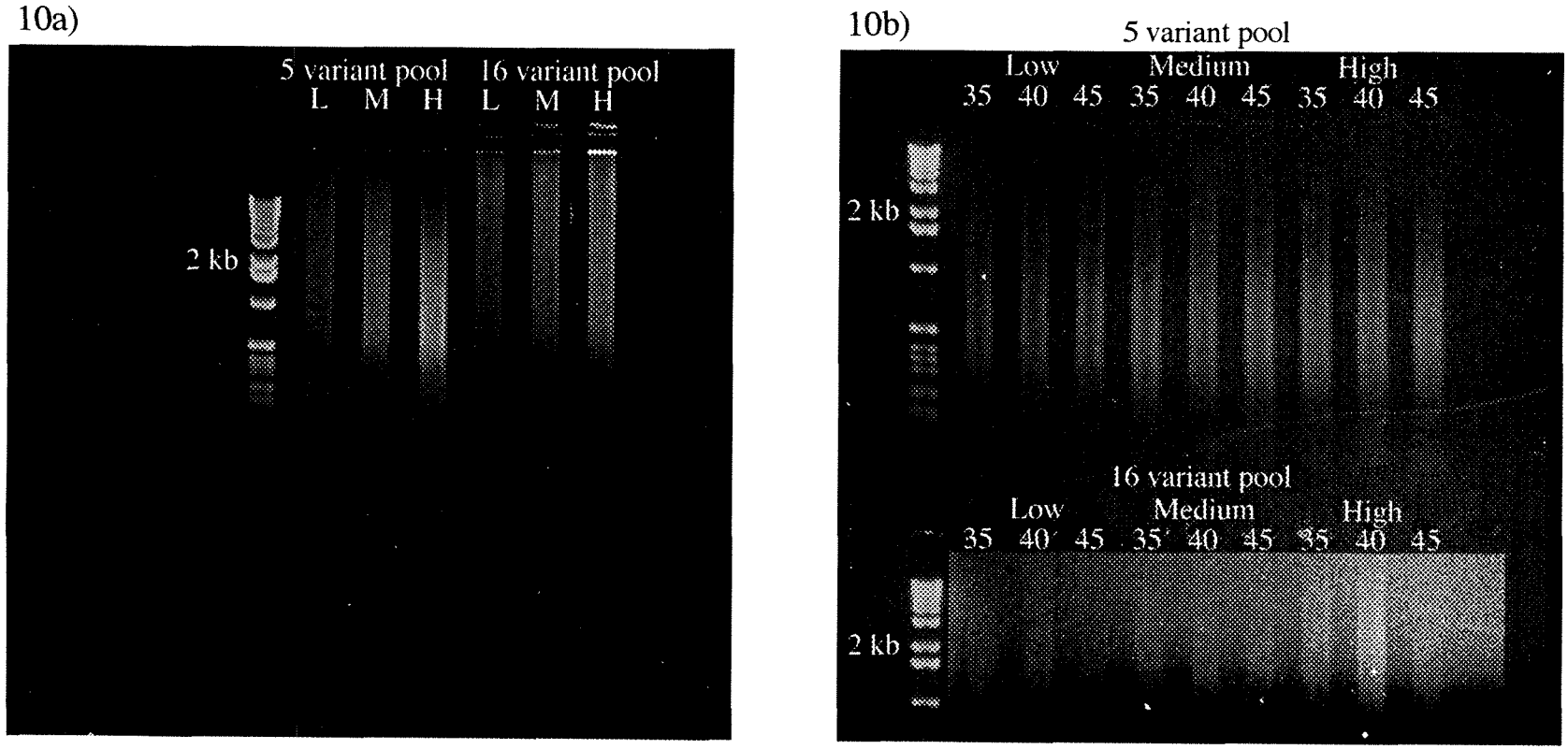



Figure 3.10a: The 45 cycle reassembly product of the 125 to 350 bp digest for the 5 and 15 variant pools using low, medium and high concentrations of DNA as indicated. These concentrations are the result of precipitating the fragment DNA and resuspending it in 30 $\mu$L of buffer. Of this solution 1 $\mu$L, 4 $\mu$L and 10 $\mu$L were brought to a final volume of 20 $\mu$L for reassembly, and correspond to low, medium and high concentrations of DNA. Samples of the reassembly were also taken at 35 and 40 cycles. Figure 10b: Amplification of the reassembly products from cycles 35, 40 and 45 from the low, medium and high concentration of DNA from each variant pool. The 2 kb band of the molecular weight marker has been labeled to indicate the size of the expected product.

partial plasmid DNA map in Figure 11.) Figure 12 demonstrates the DNase I digest of similar DNA fragments amplified using primers located slightly outside the original primers used. The first noticeable difference in this DNA shuffling as compared to the previous ones is the appearance of discrete bands in the DNase I digest (Figure 12). Every shuffling experiment that worked demonstrated this feature, while every experiment that did not work lacked it. This feature has been observed in the Stemmer laboratory as well (Crameri-A, personal communication). DNA fragments in the 200 to 350 base pair range for both the 5 variant and 15 variant pools were isolated and used for reassembly. The result of this experiment is the lower half of the gel in Figure 13. Again, reassembly is seen to sizes well past the 2 kb gene size expected. Using dilutions of this DNA as template for the subsequent DNA amplification step resulted in the appearance of faint 2 kb product bands in the 1:40 dilution and strong 2 kb product in the 1:20 dilution, as shown in the top half of Figure 13. The primers seem to be the single biggest factor determining the performance of DNA shuffling. It is unknown why the original set did not lead to a successful shuffling experiment, or why the new primers did.

The 300 to 600 base pair fragments from the DNase I digest were reassembled and amplified using Taq DNA polymerase. The 2 kb DNA bands formed through the DNA shuffling experiments were restricted by Xba I and Bam HI and ligated into the expression vector. From the recombination pool of the best five variants in generation four (of which 4-54B9 was the best variant), over 500 colonies were examined. The data from one 96-well plate are shown in Figure 14. The most striking result was that 90% of the fifth generation variants screened were inactive. The most active clones from this screening were retested against 4-54B9 and were shown to be similar in activity to 4-54B9. These results, and in particular the high number of inactive variants, do not correspond well with the predicted results of Figure 3. At the very least, a significant fraction of the fifth generation variants should be identical to the third generation "grandparent" 3-10C4 and each of the fourth generation parents. Problems with the cloning step were ruled out when

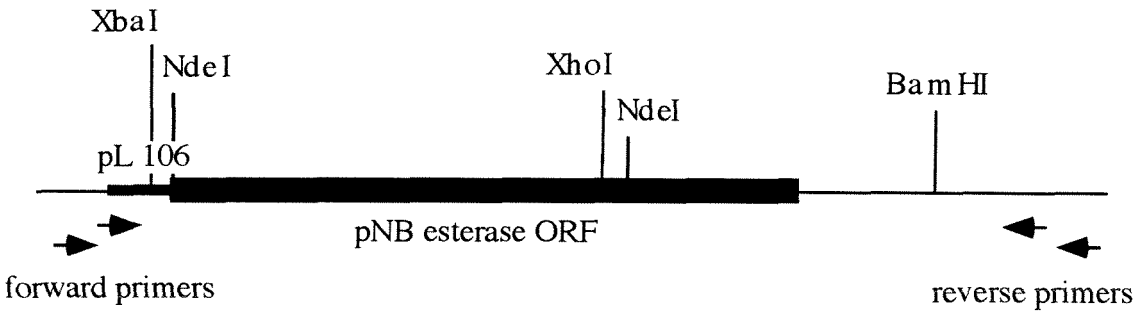



Figure 3.11: The portion of the plasmid map containing the pNB esterase gene. The thick line indicates pNB esterase open reading frame, the medium line is the pL106 promoter region, and the thin line indicates the continuation of plasmid DNA. Restriction sites are as marked. The location of the forward and reverse primers used in the PCR procedures are indicated by small arrows. The inner forward and reverse primers represent the original primers; the outer primers represent the new, alternate primers.

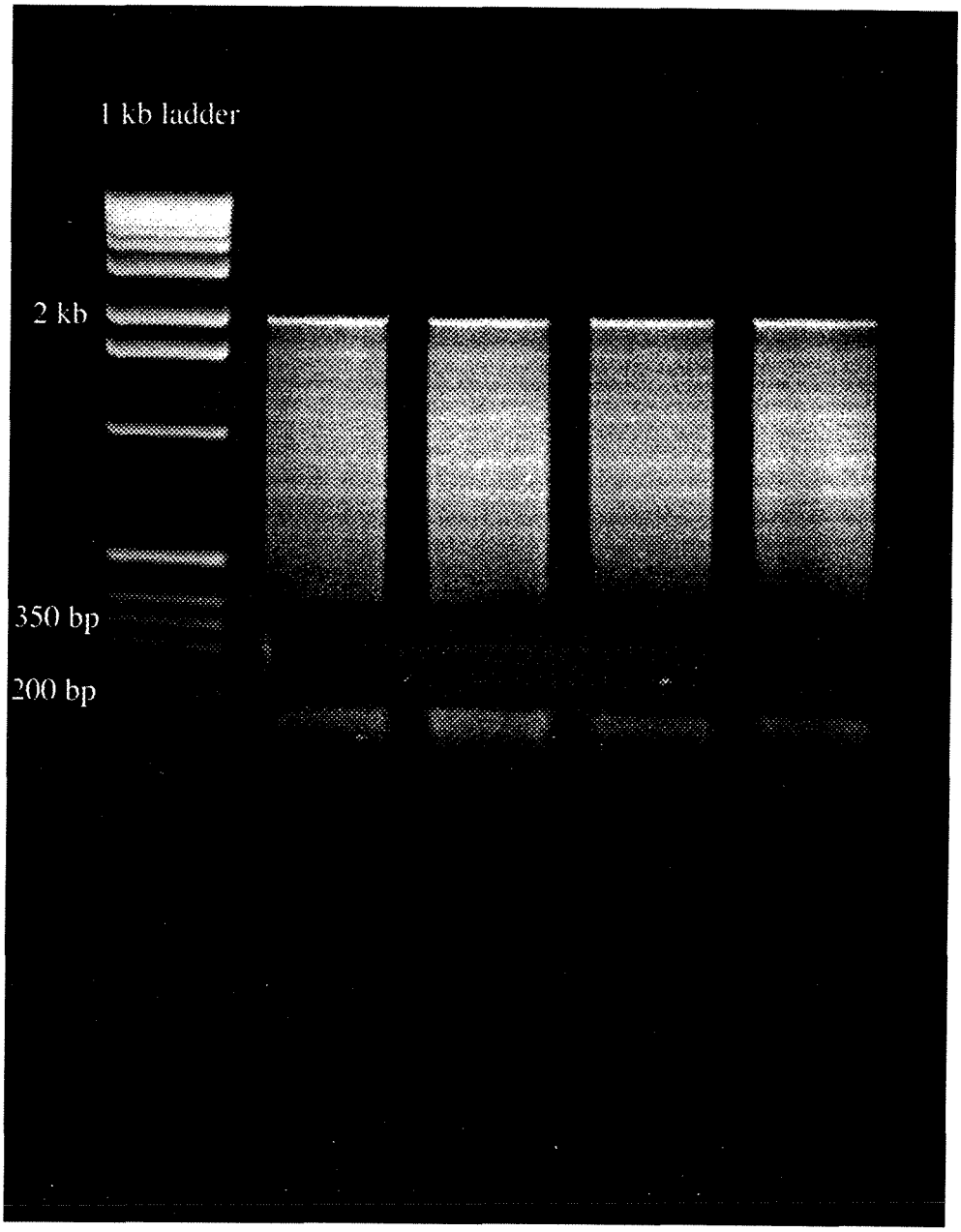



Figure 3.12: Gel electrophoresis of a DNase I digest from the 5 variant pool. DNA fragments from 200 to 350 bp were isolated. Note the striations in the digest.

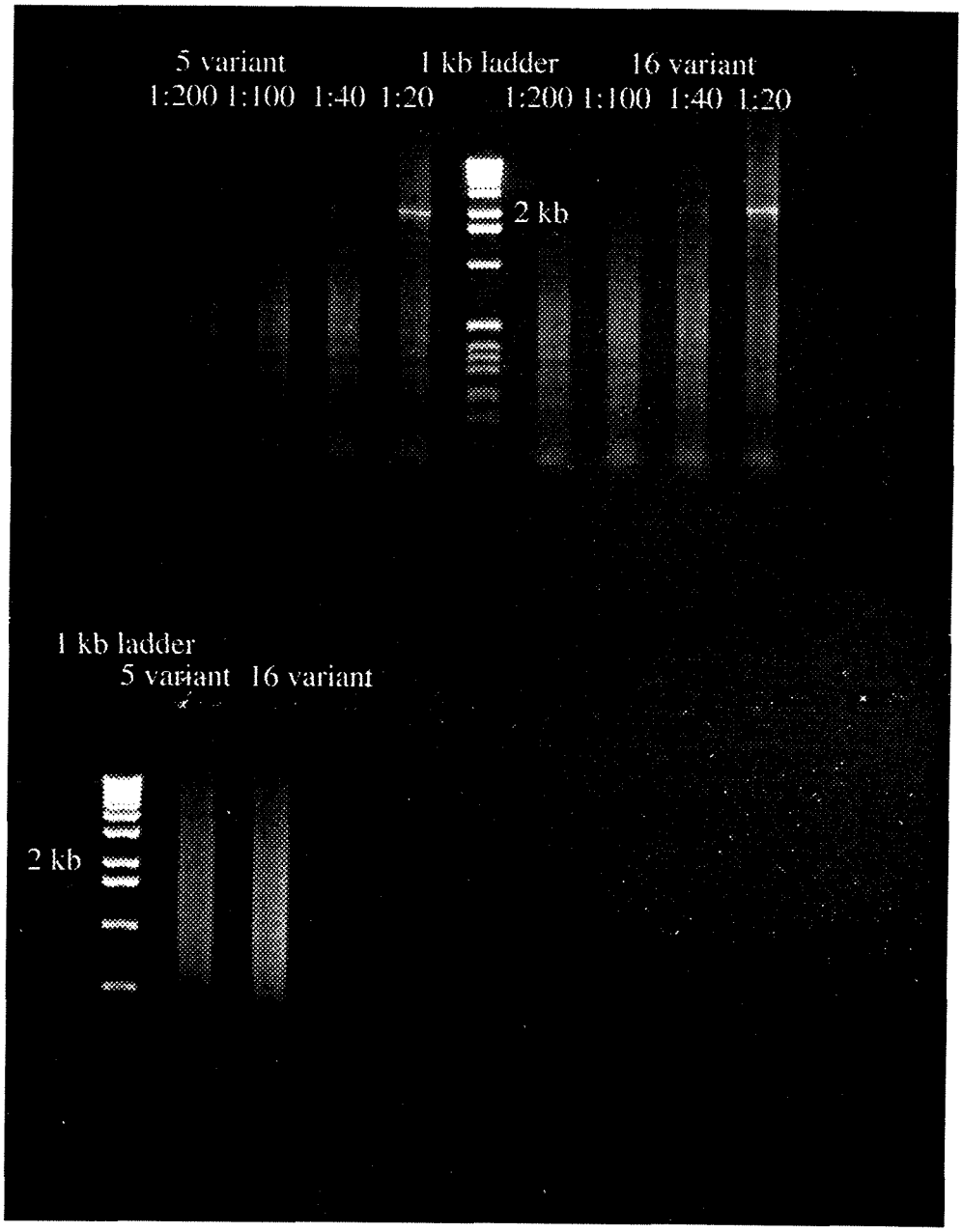



Figure 3.13: The lower half of this gel shows the result of the reassembly reaction for the 200 to 350 bp digest product for both the 5 and 15 variant pools. The upper half shows the amplification of this reassembly diluted 1:200 to 1:20 as indicated.

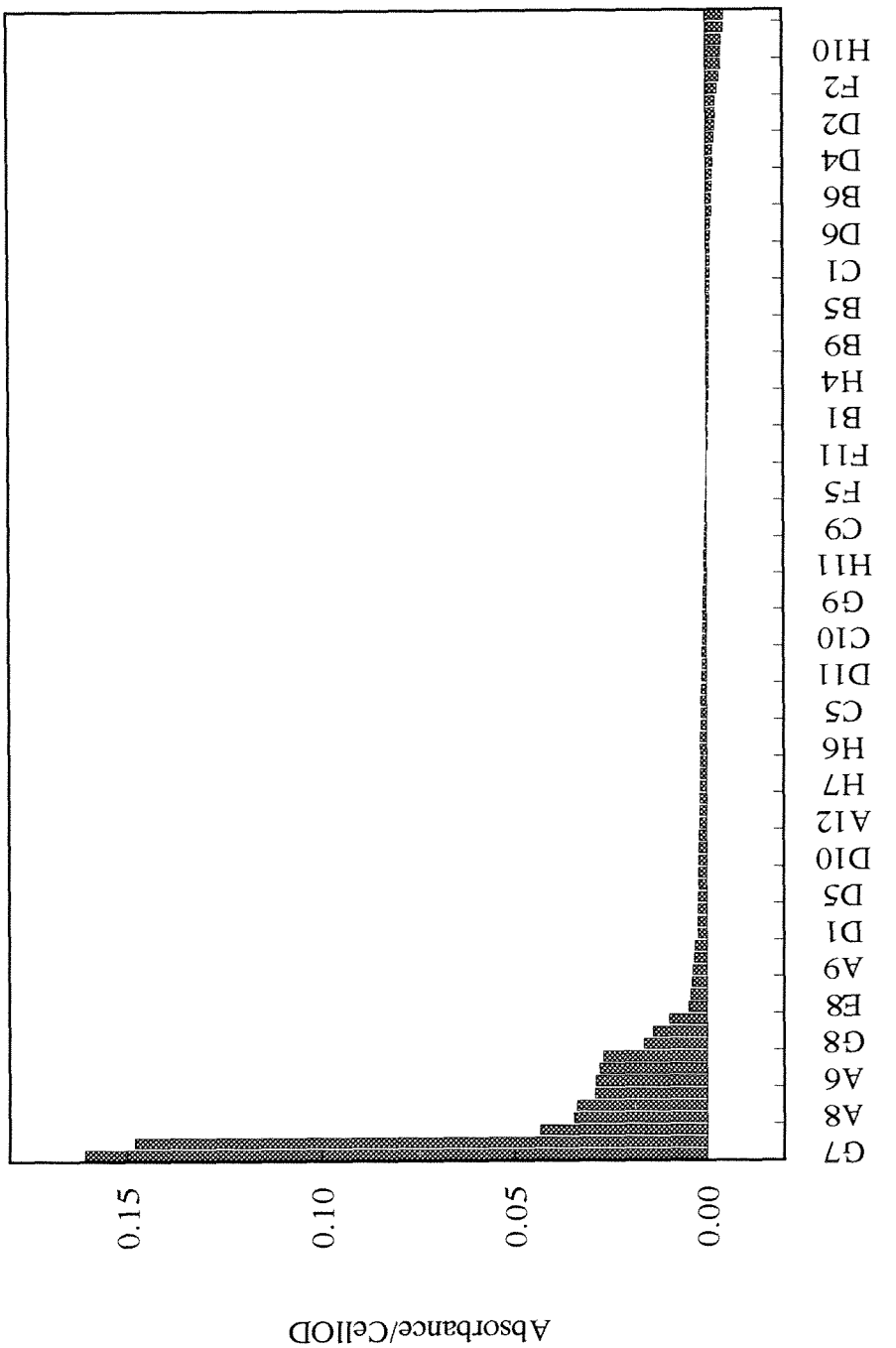


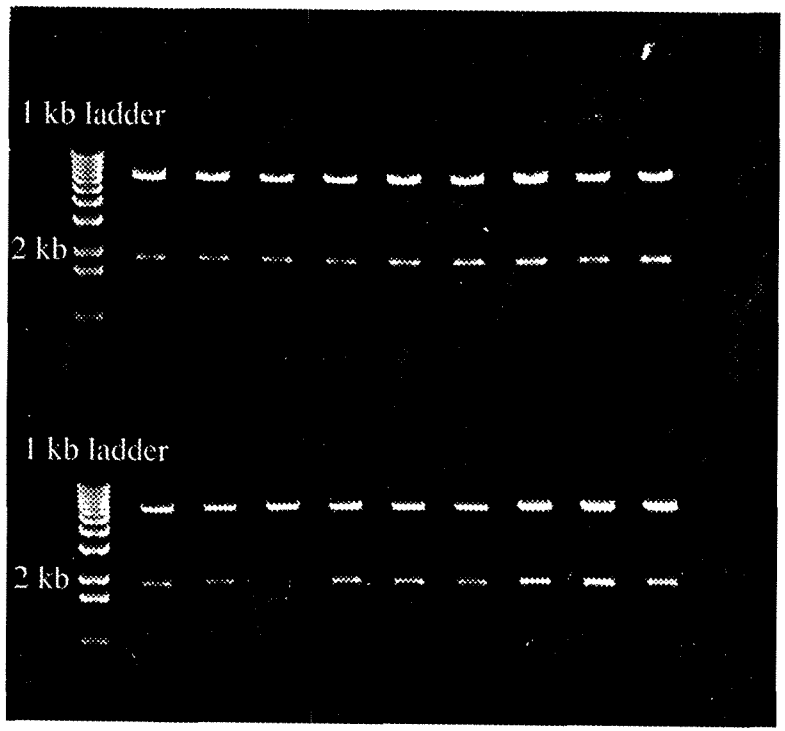

Figure 14: Screening assay data from DNA shuffling clones using the Taq polymerase only.

Approximately 90% of the clones screened appear inactive.

18 random clones were grown in 3 mL liquid culture and the plasmid DNA isolated. This plasmid DNA was restricted with the Xba I and Bam HI (Figure 15a) and with Nde I and Bam HI (Figure 15b). The digest using Xba I and Bam HI (Figure 15a) confirms that all 18 clones have DNA inserts, and that 17 of 18 contain inserts of the correct 1.8 kb size. The pNB esterase gene contains two Nde I sites, one 50 bp inside the Xba I site and the other 1.1 kb from the first and 0.7 kb from the Bam HI site (Figure 11). Of 18 clones, 16 demonstrate the 1.1 kb and the 0.7 kb bands as a result of this digest. The 1.8 kb band seen in varying intensities is an incomplete digest product. The high proportion of correct inserts seems to rule out difficulty in the cloning step and suggests that a relatively high rate of mutation accompanies the recombination process under the conditions used (300-600 bp fragments and the use of Taq DNA polymerase for reassembly and amplification). This procedure has been reported to introduce random mutations with error rates as high as 0.7% per base; the rate of mutation is believed to be dependent on the size of the digested fragments and the Taq DNA polymerase used to reassemble the small fragments (Stemmer, 1994). Previously reported shuffling experiments also indicate the presence of point mutagenesis in progeny (Lorimer and Pastan, 1995, Stemmer, 1994b).

In an effort to reduce the rate of background random mutagenesis, a proofreading polymerase enzyme (Pwo - Boerhinger) was added to the reassembly mix. The total number of units of polymerase activity was kept constant and different dilutions of Taq to Pwo polymerases were added. Figure 16a shows the digest and the resulting reassembly using Taq only, 20:1, 10:1 and 1:1 Taq to Pwo ratios. Figure 16b shows the 2 kb DNA product resulting from the final PCR amplification using 1:200, 1:100, 1:40, and 1:20 dilutions of the reassembled product in Figure 16a as template. The PCR reactions generating the 2 kb product for each Taq:Pwo ratio were pooled, purified, restricted, ligated into the expression vector and screened for activity. The effect of the ratio of Taq to Pwo polymerases is summarized in the following table (Table 3.1).

15a)




15b)

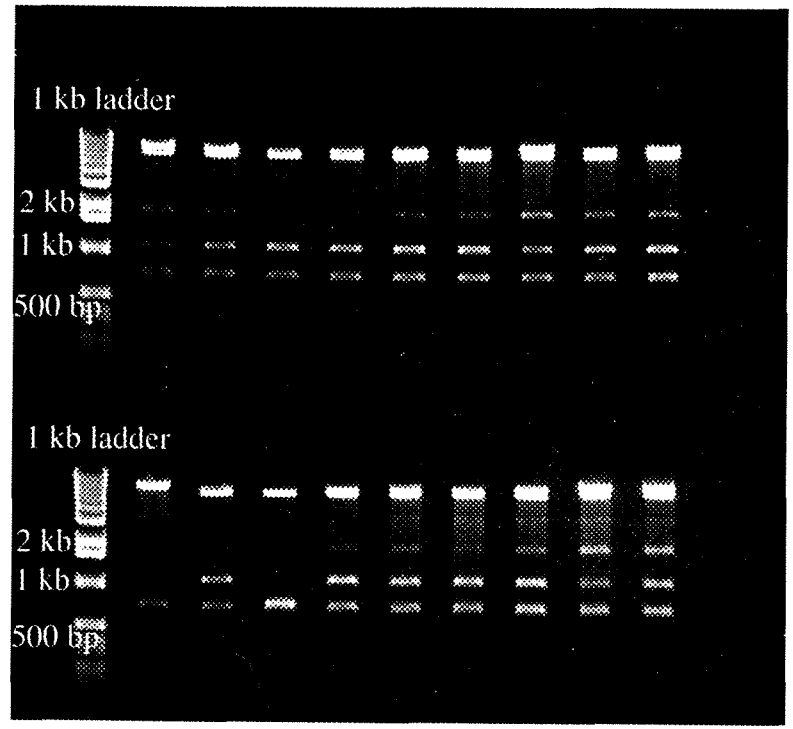
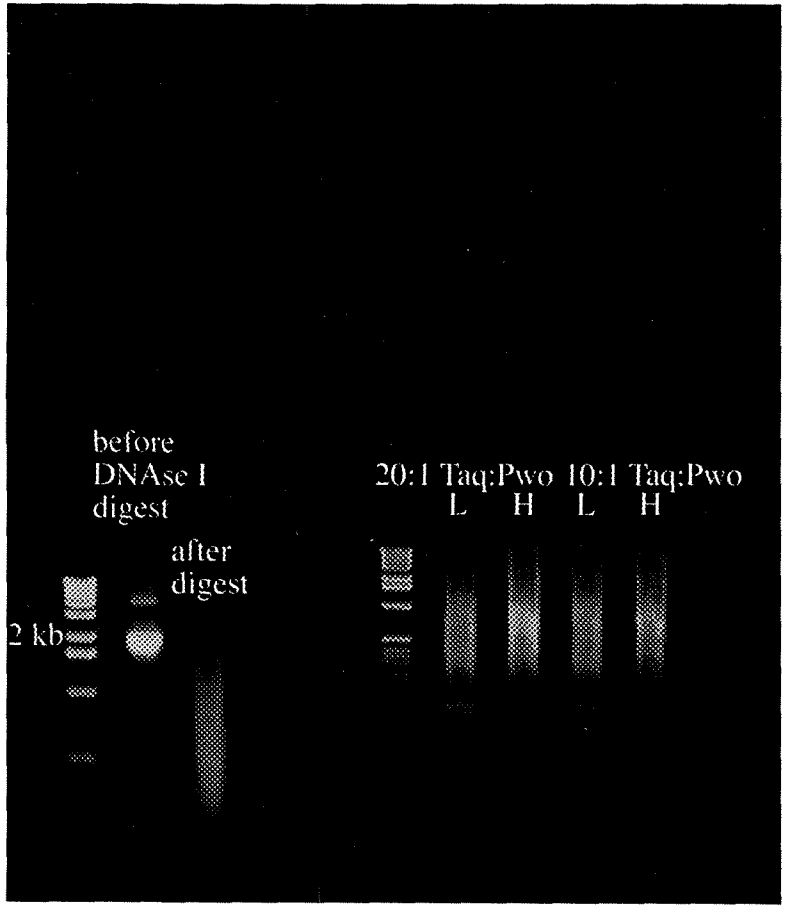



Figure 3.15: Restriction digests of 18 random clones. The DNA samples in Figure 15a were cut with Xba I and Bam HI, the two restriction enzymes used for cloning, and a 1.8 kb band is the expected result. The DNA samples in Figure 15b were cut with Nde I and Bam HI, and two fragments 0.7 kb and 1.1 kb are expected. The 1.8 kb product is a partial digest product and is the sum of the two smaller fragments.

16a)




16b)

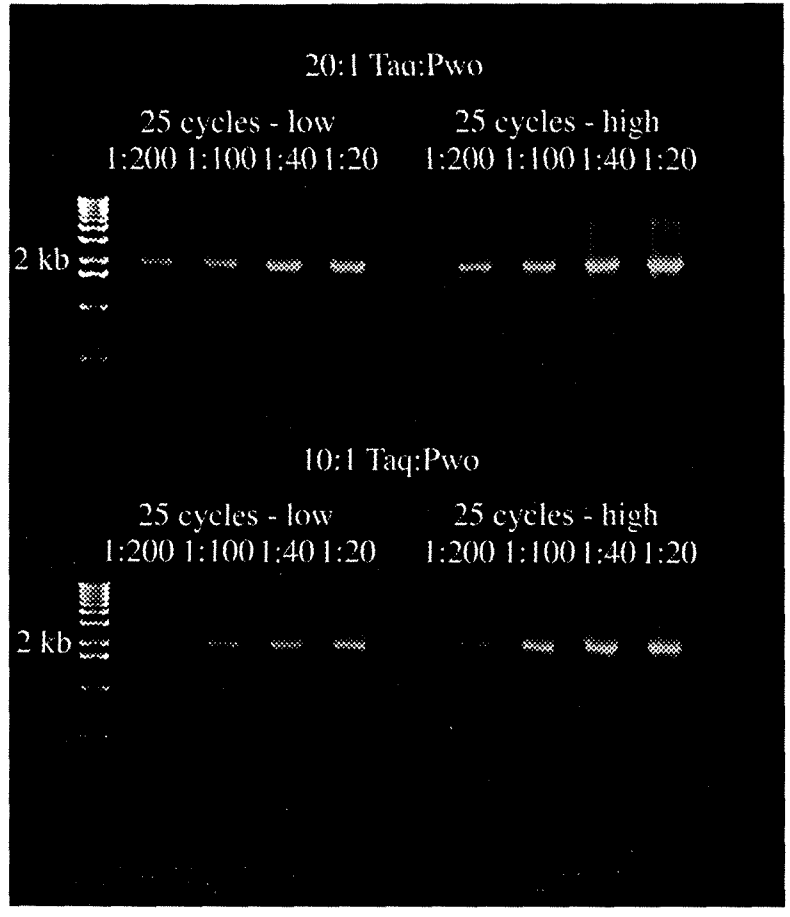



Figure 3.16a: Before and after DNase I digestion, and a 25 cycle reassembly of digested fragments using different dilutions of Taq and Pwo polymerases. Figure 16b: Amplification of the 25 cycle reassembly product using the dilutions noted.

| Taq:Pwo ratio | % variants showing no activity |
|---|---|
| Taq only | 90% |
| 20:1 | 45% |
| 10:1 | 50% |
| 1:1 | 20% |
| Pwo only (Jin et al., 1996) | 20% |

The conditions used in the reassembly process were those associated with the Taq polymerase. According to the manufacturer, the Pwo polymerase prefers $MgSO_4$ to $MgCl_2$ and does not perform well under the conditions that are optimal for the Taq polymerase. The Pwo polymerase was used in the absence of Taq under the conditions optimal for Taq, and no evidence of reassembly was observed. Not only was no reassembly observed, but the concentrated small DNA fragments to be reassembled were completely gone, suggesting that the proofreading activity of the Pwo enzyme had degraded the DNA fragments. Further experiments using only the Pwo enzyme on this variant pool and under optimal conditions for the Pwo polymerase have been shown to produce the results included in Table 1 (Jin, Moore, and Arnold, 1996).

Variants demonstrating apparent activity greater than 4-54B9 were rescreened. Five variants more active than 4-54B9 were found as a result of shuffling the five best variants from generation four and screening 440 colonies. Of the 440 colonies, 200 were inactive. The number of colonies screened could be significantly reduced using the lowest mutagenic condition described here (primarily the use of 1:1 Taq:Pwo DNA polymerases). The activity of the five variants and the fifth generation variant produced by ligation shuffling are shown compared to 4-54B9 in Figure 17. The ligation of variant 4-54B9 and 4-38B9 displays twice the activity of 4-54B9. All but one of the screened variants display comparable activity. The variant 5-31D4 demonstrates a 3.5-fold activity increase over 4-54B9 for an approximate 100-fold increase in total activity over the wild type pNB esterase.

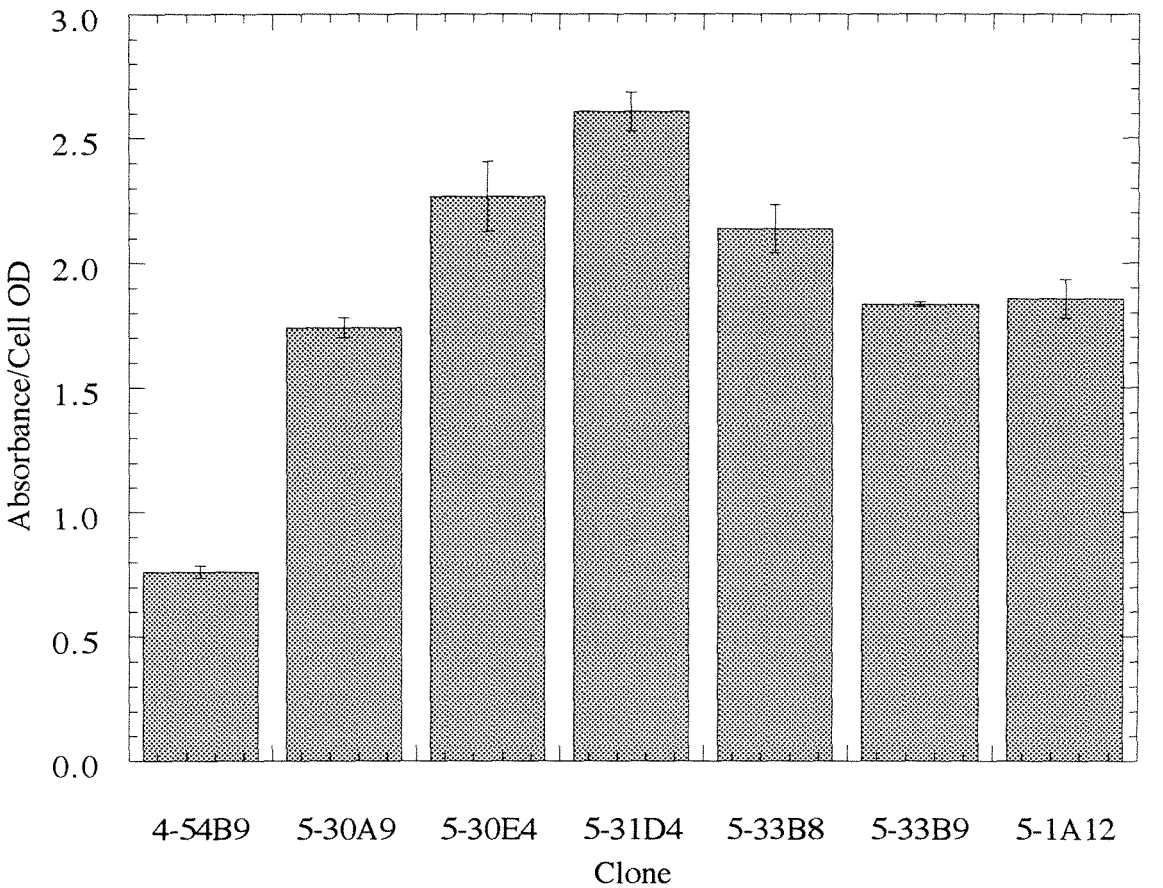



Figure 3.17: Activity of five positive variants as a result of DNA shuffling under the screening assay conditions of 0.8 mM LCN-pNP and 25% DMF. Clone 4-54B9 is the most active variant in the shuffled pool, and 5-1A12 was found by restriction/ ligation of 4-54B9 and 4-38B9.

The shuffling was done in the absence of any sequence information. The sequences obtained for the fourth generation variants which were shuffled show that two of the variants, 4-38B9 and 4-73B4, are identical (Moore and Arnold, 1996). Further, 4-54B9 and 4-43E7 share different mutations in a codon for the same amino acid. This reduces the number of likely positions for recombination from five to three. At the DNA level, these three positions are at base 181, 1003 or 1004 (2 mutations in same codon), and 1031. Table 3.2 illustrates the mutations found in 5-1A12, the variant generated by restriction/ligation and the two most active variants generated by DNA shuffling. All three contain the mutations at position 181 and 1004, which combines the Ile 60 to Val substitution from 3-38B9 with the Leu 334 to Ser from 4-54B9, while none contain the mutation at position 1031. The variant 5-30E4 appears 25% more active than 5-1A12, and this may be due to an increase in specific activity as a result of the new translated mutation at position 1364, producing a Val substitution for Ala at amino acid position 454, or an increase in total protein produced by the bacteria as a result of any of the changes at the DNA level. This variant also contains a silent mutation at position 255 from 4-53D5, and therefore, clearly demonstrates that at least two recombination events have occurred within this sequence. The variant 5-31D4 is 50% more active than 5-1A12 and contains 5 new mutations, of which 2 are translated (the C to T mutation at position 769 translates to Leu 256 replaced with Phe, and the A to T at position 1354 translates to Asn 451 changed to Tyr). Again the increase in activity could be the result of an increase in specific activity or total protein produced. An updated lineage tree reflecting these new fifth generation sequences is shown in Figure 3.18, and the approximate locations of the new substitutions are indicated in Figure 3.19.

Screening 500 variants from the similar shuffling of the pool of 15 variants failed to provide any significantly positive variants. Figure 3 suggests that 20% of these 500 variants should be double mutants, but that each double mutant should be represented 1.5 times in 1000 clones. Since the best 5 variants are a subset of the best 15, the positives

| | 4th Generation Variants | | | | 5th Generation Variants | | | |
|---|---|---|---|---|---|---|---|---|
| DNA Pos. | 4-38B9/73B4 | 4-43E7 | 4-53D5 | 4-54B9 | 5-1A12 | 5-30E4 | 5-31D4 | Amino Acid Substitution |
| 102 | | | | T → C | | | | |
| 181 | A ⇒ G | | | | A ⇒ G | A ⇒ G | A ⇒ G | I 60 V |
| 255 | | | T → C | | | T → C | | |
| 283 | | A ⇒ G | | | | | | S 94 G |
| 360 | | | | | | | C → T | |
| 720 | | | T → C | | | | | |
| 769 | | | | | | | C ⇒ T | L 256 F |
| 803 | | | | A ⇒ G | | | | K 267 R |
| 864 | | | | | | | A → G | |
| 903 | | | | | | | T → C | |
| 1003 | | T ⇒ G | | | | | | L 334 V |
| 1004 | | | | T ⇒ C | T ⇒ C | T ⇒ C | T ⇒ C | L 334 S |
| 1031 | | | C ⇒ T | | | | | A 343 V |
| 1354 | | | | | | | A ⇒ T | N 451 Y |
| 1364 | | | | | | C ⇒ T | | A 454 V |

Table 3.2: The novel mutations present in the fourth and fifth generation variants. The reverse type arrows indicate those mutations which are translated into amino acid substitutions, with the final column indicating the substitution. The DNA positions identified from the fourth generation giving rise to enhanced activity and available for recombination are boxed.

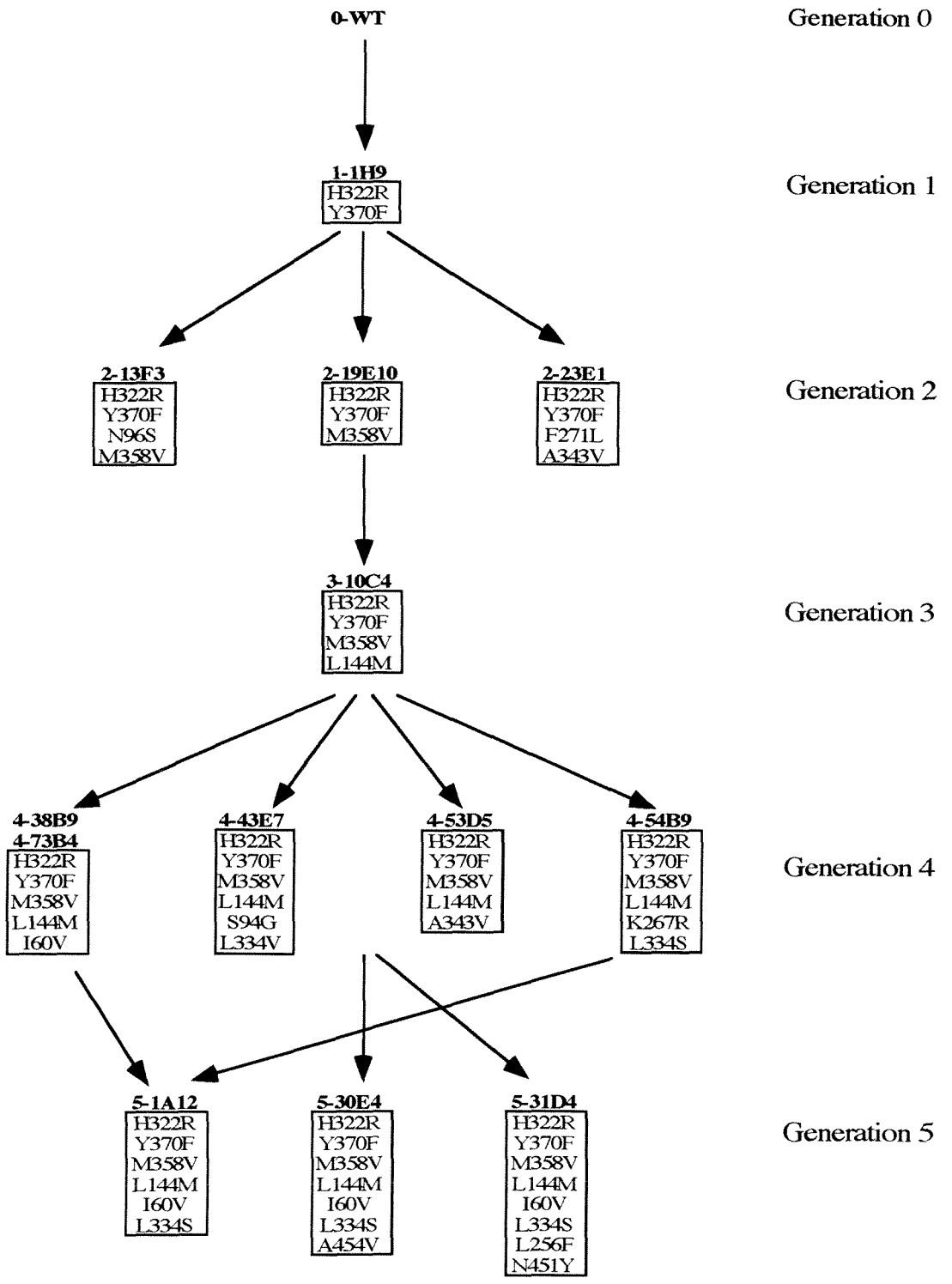



Figure 3.18: Tree showing amino acid substitution progression in pNB esterase variants, including the fifth generation variants created by recombination.

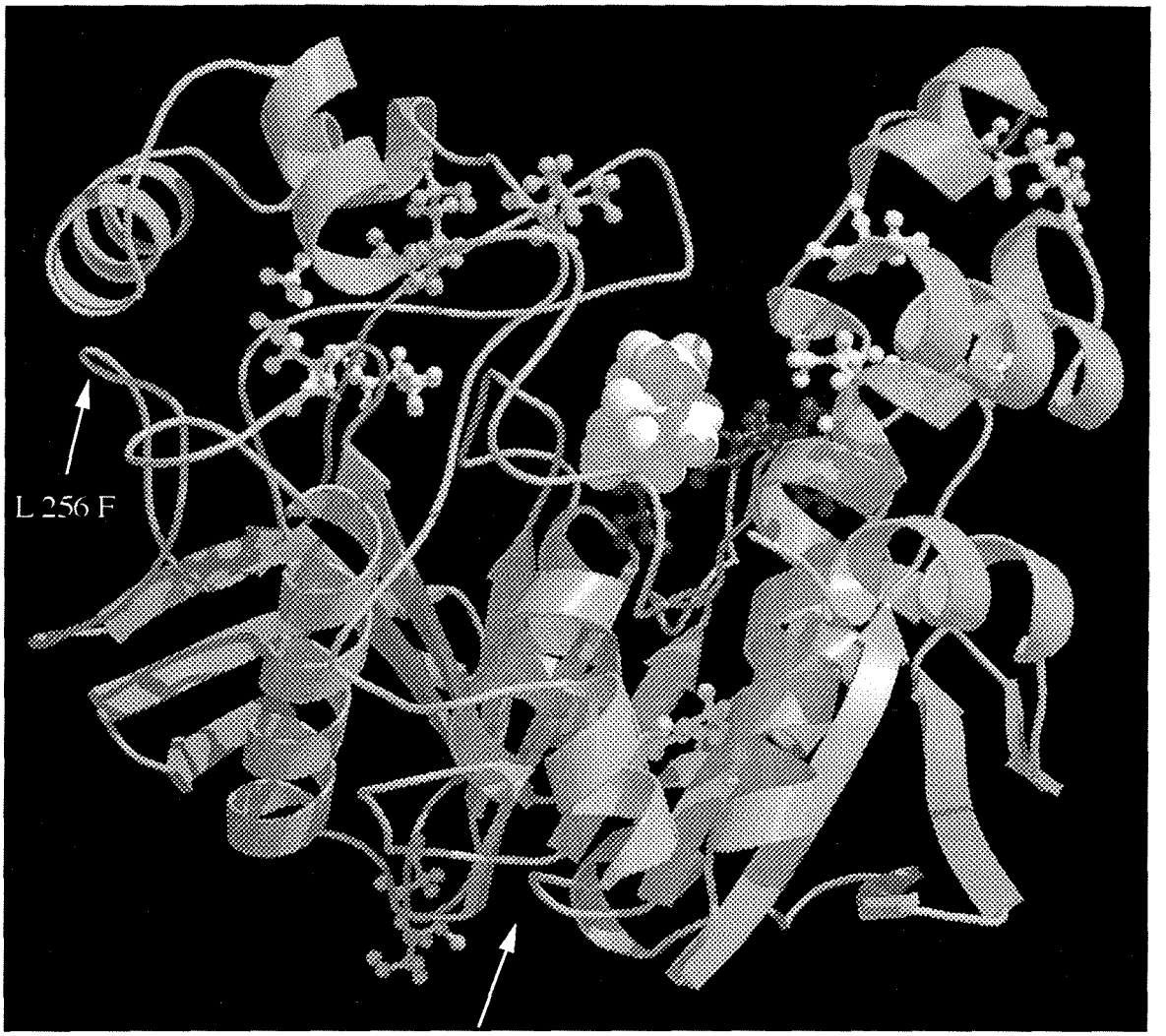



Figure 3.19: Approximate locations of new amino acid substitutions found in variants 5-30E4 and 5-31E4.

found during screening of the 5 variant pool are present in the 15 variant pool. It is therefore likely that insufficient numbers have been screened to find clones significantly more active than 4-54B9. This demonstrates that shuffling this relatively larger number of sequences may be disadvantageous when screening is required to find positives.

Recently, this work has been continued by others in our lab. The original five variant pool was reduced to the four unique variants and was reshuffled and rescreened, with results very similar to those presented here. Six clones were identified, four of which were approximately two-fold more active than 4-54B9, and one of which was four-fold more active. These six clones were then shuffled together, resulting in two clones with more than a 5.5-fold increase in activity over 4-54B9.

## Conclusion

Recombination is a powerful natural tool used to propagate positive traits and eliminate negative traits in progeny. This natural process has been imitated through the DNA shuffling to produce new pNB esterase variants more active than any previously discovered. These new variants combine those positive mutations and eliminate the negative or neutral mutations present in the parent population, just as in the natural world. Unlike the natural case, however, the best progeny generated in these experiments resulted from the recombination of DNA from five parents. The number of progeny one must examine in order to find more "fit" progeny increases dramatically with an increasing number of parents. In fact, simple calculations suggest that increasing the number of parents increases the likelihood that offspring will resemble parents and grandparents. This was born out experimentally by the DNA shuffling of a 5 and a 15 variant pool. The best progeny appeared to be 3.5 times more active than the most active parent enzyme, giving this new pNB esterase variant more than a 100-fold increase in activity over the original wild type enzyme.

**Materials and Methods**

*DNA shuffling*

The DNA shuffling method outlined here follows the procedure outlined by Stemmer (Stemmer, 1994a). The genes coding for the variants from the second and fourth generation were amplified individually by PCR in an MJ Research (Watertown, Mass.) PTC-200 thermocycler to a final volume of 100 (for 16 variants) to 300 $\mu$L (for 5 variants). Initially, the primers used for amplification were those listed previously. Alternatively, new primers (5' to 3') ATACAGATAACCATCTGCGGT (forward primer) and ATTTCCTTACGCGAAATACGG (reverse primer) were used. The PCR products were then pooled, providing a final volume of 1.5 to 1.6 mL. The pooled volume was divided into five equal volume aliquots and purified using the Promega (Madison, WI) Wizard PCR preps kit. The resulting DNA was eluted in 500 $\mu$L water. 96 $\mu$L of this solution was combined with 12 $\mu$L digest buffer (500 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$) and 12 $\mu$L of DNase I (diluted 1 $\mu$L in 750 $\mu$L water) (RNase-free, Boerhinger Mannheim, Indianapolis, IN). The digest was allowed to proceed for 4 minutes, after which time 4 $\mu$L were removed and quickly loaded onto an agarose gel for electrophoresis (LKB GNA-100 gel box and power supply, Pharmacia, Piscataway, NJ). The remaining sample was rapidly frozen in a dry ice/acetone bath to prevent further digestion. The agarose gel was examined utilizing ethidium bromide staining and UV radiation. If the digest did not produce DNA smearing through the 100 to 200 bp region as indicated by the 10, 100 and/or 1 kb bp markers (Boerhinger), the sample was thawed and allowed to digest for another 2 to 4 minutes. This freezing procedure was repeated until DNA smearing of the appropriate size was observed. At this point, the entire reaction was loaded onto a new agarose gel. The sample was run until good definition was observed between the different bands in the DNA markers. An incision was made in the gel at the location of the marker of the smallest DNA size of interest. Pre-cut DE-81 filter paper (Whatman, Maidstone, England) was inserted into the slit, and the gel underwent further electrophoresis. The gel

was run until the DNA marker equivalent in size to the largest DNA size of interest reached the filter paper in the gel. The filter paper was removed from the gel, placed in 0.5 mL 1.2 M NaCl, and finely ground using a pipette tip. The resulting slurry was then spun in a Beckman E centrifuge with a XX rotor for 2 minutes. The supernatant was re-centrifuged through a Millipore "red dot" filter to remove all traces of DE-81 paper. The DNA was then precipitated with 2 volumes of ethanol, rinsed with 1 volume 70% ethanol, 1 volume of 100% ethanol and allowed to air dry. The DNA pellet was resuspended in 30 $\mu$L of PCR buffer containing either Taq (Promega) or Taq and Pwo (Boerhinger) polymerases and reassembled without DNA primers in the thermo-cycler for 25-45 cycles using the following cycle conditions: 94 °C for 30 sec., then repeated cycles of 94 °C for 30 sec., 48 °C for 30 sec., and 72 °C for 30 sec. + 1 sec/cycle. Dilutions of 1:20, 1:40 and 1:100 of the PCR product into fresh PCR solution with DNA primers were then amplified for 20 to 25 cycles, using repeated cycle conditions of 94 °C for 30 sec., 48 °C for 30 sec., and 72 °C for 1 min. These cycles were followed by a 10 min. 72 °C step. The PCR product was purified with Promega's PCR Wizard Prep kit, and digested with Xba I and Bam HI. The digest was run on an agarose gel, and the band of the correct size was cut out and purified using the Qiagen (Chatsworth, CA) Qiaex gel purification kit.

*Cloning*

Previously prepared plasmid DNA was restricted with Xba I and Bam HI, and the fragment lacking the pNB esterase gene purified. This fragment was ligated (T4 DNA ligase, Boerhinger) with the pNB esterase gene prepared by DNA shuffling. The ligation product was transformed using the standard $CaCl_2$ procedure and transformants were selected on LB plates containing 20 mg/mL tetracycline at 30 °C.

*Plasmid isolation*

Plasmid isolates were performed using the Qiagen plasmid mini-prep kit.

*Screening*

Screening was performed as described previously. Colonies were induced at 40 ℃ for eight hours and then picked and resuspended in 0.1 M Tris-HCl, pH 7.0 in individual wells of 96 well plates. A 20 $\mu$L aliquot was placed in a new 96 well plate, 200 $\mu$L of 0.8 mM LCN-pNP, 25% DMF, 0.1 M Tris-HCl solution was added, and the absorbance at 405 nm was recorded. The rate of absorbance change was normalized by cell density (absorbance of cell resuspension at 620 nm) and compared to the parent values. Those variants which outscored the parents were streaked on LB tetracycline plates and reassayed.

## References

Arnold, F. H. 1993. Engineering proteins for nonnatural environments. FASEB J. 7:744-749.

Berkenpas, M.B., Lawrence, D. A., and Ginsburg, D. 1995. Molecular evolution of plasminogen-activator inhibitor-1 functional stability. EMBO J. 14:2969-2977.

Brock, T. D., Madigan, M. T., Martinko, J. M., and Parker, J. 1994. Biology Of Microorganisms. Prentice-Hall, Inc., Englewood Cliffs, NJ. 909 pp.

Chen, K. Q. and Arnold, F. H. 1993. Tuning the activity of an enzyme for unusual environments - sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc. Natl. Acad. Sci. 90:5618-5622.

Crameri, A. 1995. Personal Communication.

Crow, J. F. 1988. The importance of recombination. *In* The Evolution Of Sex: An Examination Of Current Ideas. Sinauer Associates, Inc., Sunderland, Mass. 56-73.

Gulick, A. M. and Fahl, W. E. 1995. Forced evolution of glutathione-s-transferase to create a more efficient drug detoxication enzyme. Proc. Natl. Acad. Sci. 92:8140-8144.

Hames, B. D. and Glover, D. M. 1990. Gene Rearrangement. *In* Frontiers In Molecular Biology. Oxford University Press, New York. 1-154.

Light, J. and Lerner, R. A. 1995. Random mutagenesis of staphylococcal nuclease and phage display selection. Bio. & Med. Chem. 3:955-967.

Lorimer, I. A. J. and Pastan, I. 1995. Random recombination of antibody single-chain fv sequences after fragmentation with DNase I in the presence of $Mn^{2+}$. Nucl. Acids Res. 23:3067-3068.

Maynard Smith, J. 1988. The evolution of recombination. *In* The Evolution Of Sex: An Examination Of Current Ideas. Sinauer Associates, Inc., Sunderland, Mass. 106-125.

Moore, J. C. and Arnold, F. H. 1996. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nat. Biotech. 14:458-467.

Muller, H. J. 1932. Some genetic aspects of sex. Amer. Nature. 66:118-138.

Okada, Y., Yoshigi, N., Sahara, H., and Koshino, S. 1995. Increase in thermostability of recombinant barley ß-amylase by random mutagenesis. Biosci. Biotech. & Biochem. 59:1152-1153.

Rawn, J. D. 1989. Biochemistry. Neil Patterson Publishers, Burlington, NC. 125-129 pp.

Rellos, P. and Scopes, R. K. 1994. Polymerase chain reaction-based random mutagenesis - production and characterization of thermostable mutants of *Zymomonas mobilis* alcohol dehydrogenase-2. Prot. Expr. Purif. 5:270-277.

Stemmer, W. P. C. 1994. DNA shuffling by random fragmentation and reassembly - in-vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. 91:10747-10751.

Stemmer, W. P. C. 1994. Rapid evolution of a protein in-vitro by DNA shuffling. Nature. 370:389-391.

Stryer, L. 1988. Biochemistry. W. H. Freeman, New York. 1089 pp.

Yabuta, M., Onaimiura, S., and Ohsuye, K. 1995. Isolation and characterization of urea-resistant *Staphylococcus aureus* V8 protease derivatives. J. Ferm. Bioeng. 80:237-243.

# Chapter 4

# Local Fitness Profiles

## Introduction

Protein sequence space is a concept introduced to discuss evolutionary pathways from one set of protein sequences to another (Maynard Smith, 1970). It is constructed by examining a protein of length N and the $20^N$ different protein sequences that can be created. A multi-dimensional space is created to house all the sequences, and each sequence is positioned so that its nearest neighbor is a single amino acid substitution distant. Evolution can then be envisioned as a random walk through this space with natural selection acting on the resulting sequences. If a new sequence is evolutionarily more 'fit' than the old one, natural selection forces the fixation of the mutation(s) which generated the new fitness. A fitness landscape is the surface created by evaluating or assigning fitness values to different sequences in protein sequence space. Studies have been done to model the different possible behaviors of this fitness landscape space (Kauffman and Levin, 1987, Macken et al., 1991) and to find optimal methods of searching for macromolecules containing desired fitnesses (Kauffman and Macready, 1995, Perelson and Macken, 1995).

The fitness landscape of protein sequence space is impossibly large, too large to construct completely or to work with experimentally. A simplification of this space is the *local* fitness landscape, which deals only with the immediate (single substitution) neighbors of a given sequence. For instance, all the sequences containing single amino acid substitutions (19xN of them) from a given sequence and each of those sequences' associated fitnesses constructs the local fitness landscape around that sequence. Random mutagenesis and screening processes usually represent single steps of a directed evolutionary walk through a series of local landscapes. In this format, recombination (as described in the previous chapter) can be viewed as a method to take information from multiple steps of different walks and combine this information to make a longer walk in a particular direction.

Local fitness landscapes can be sampled and analyzed; everyone who performs directed evolution experiments probes some aspect of the local fitness landscape. A method that aids in the visualization of the local fitness landscape is construction of a *fitness profile.* This is an ordered plot of the measured fitness of a series of clones, rather than a plot ordered by sequence in the local sequence space. The fitness profile's primary purpose is to locate the fitness of a parent sequence relative to the neighboring sequences.

The fitness profiles pertaining to thermostability, catalase activity and peroxidase activity around a thermostable catalase were examined and the results discussed in this context (Trakulnaleamsai et al., 1995). Measurements of the different enzymatic properties were converted to rates of reaction in the cases of activity, and rates of enzyme deactivation in the case of stability. The natural logarithm of these rates is proportional to the free energy of activation (minus the natural logarithm for rates of deactivation), and the fitness profiles were compared at this level.

In the study of fitness profiles of a thermostable catalase, only single steps of random walks were examined. No effort was made to evaluate how the local fitness landscape might change from one local sequence space to the next. Some of the interesting questions not examined are: How do these adjacent fitness landscapes differ? How fast does the number of more fit sequences dwindle as we move from generation to generation? Do different types of landscapes warrant different search strategies? Here we attempt to address some of these issues using data compiled through four rounds of random mutagenesis and screening on pNB esterase, two rounds of DNA shuffling and screening on pNB esterase, and a round of random mutagenesis and screening on subtilisin BPN'.

## Results and Discussion

*pNB Esterase*

Figure 1 demonstrates an activity profile (fitness profile where fitness is defined as activity) generated in the first generation screening of pNB esterase clones on p-nitrophenyl acetate (pNPA) after normalization against wild type activity. Approximately half of the population is inactive on this substrate in 20% dimethylformamide (DMF), while approximately three percent demonstrate more activity than the wild type enzyme. The large number of inactive clones suggests a high mutagenic rate (Moore and Arnold, 1996); this is corroborated by the single clone (1-1H9) sequenced from this generation, which contains six DNA base mutations in an 1800 bp region. Using this one mutation per 300 bp as an estimate of the mutation rate, the 1500 bp pNB esterase sequence corresponding to the open reading frame might contain five substitutions per gene. The Poisson distribution has been shown to accurately predict the distribution of mutations in sequences generated by error-prone PCR (Cadwell and Joyce, 1992), and Figure 2 is a histogram showing the probability that a sequence containing a given number of mutations will be generated. Based on the estimation of the mutation rate, this plot illustrates that fully half of the sequences contain five mutations or more. Small changes in the estimate do not alter this overall profile. Based on the assumption that a single mutation which decreases enzymatic activity will decrease activity regardless of supplemental mutations in the same sequence, we expect the region of essentially inactive clones in Figure 1 to be heavily populated with the 50% of the sequences containing five or more mutations. Figure 3 shows the same data plotted against ln (activity), which is proportional to the free energy of activation, for each clone. Here we can more clearly see that the wild type enzyme is very near the top of this profile, suggesting that the introduction of multiple mutations reduces the enzyme's fitness on the local fitness landscape with respect to activity on this substrate. The few clones demonstrating higher activity than wild type indicate that the enzyme is not the most active

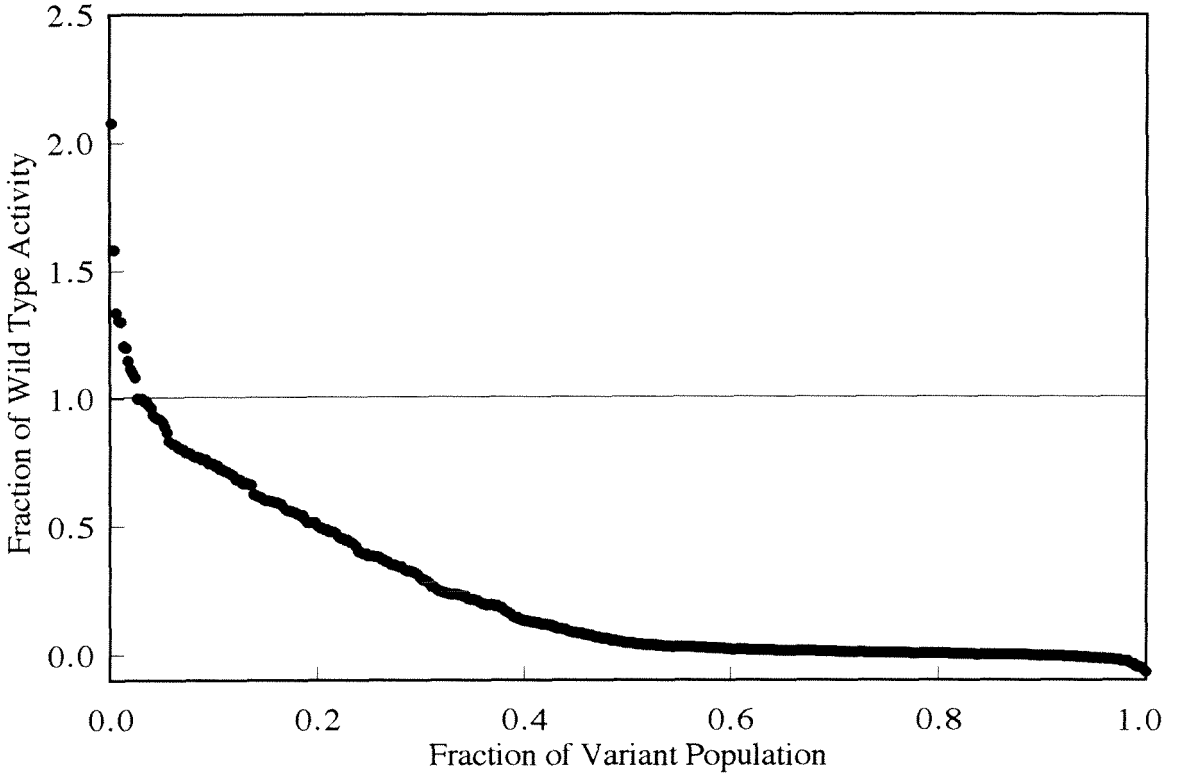



Figure 4.1: Activity profile of first generation pNB esterase variants. A total of 453 variants screened on pNPA in 20% DMF are represented.

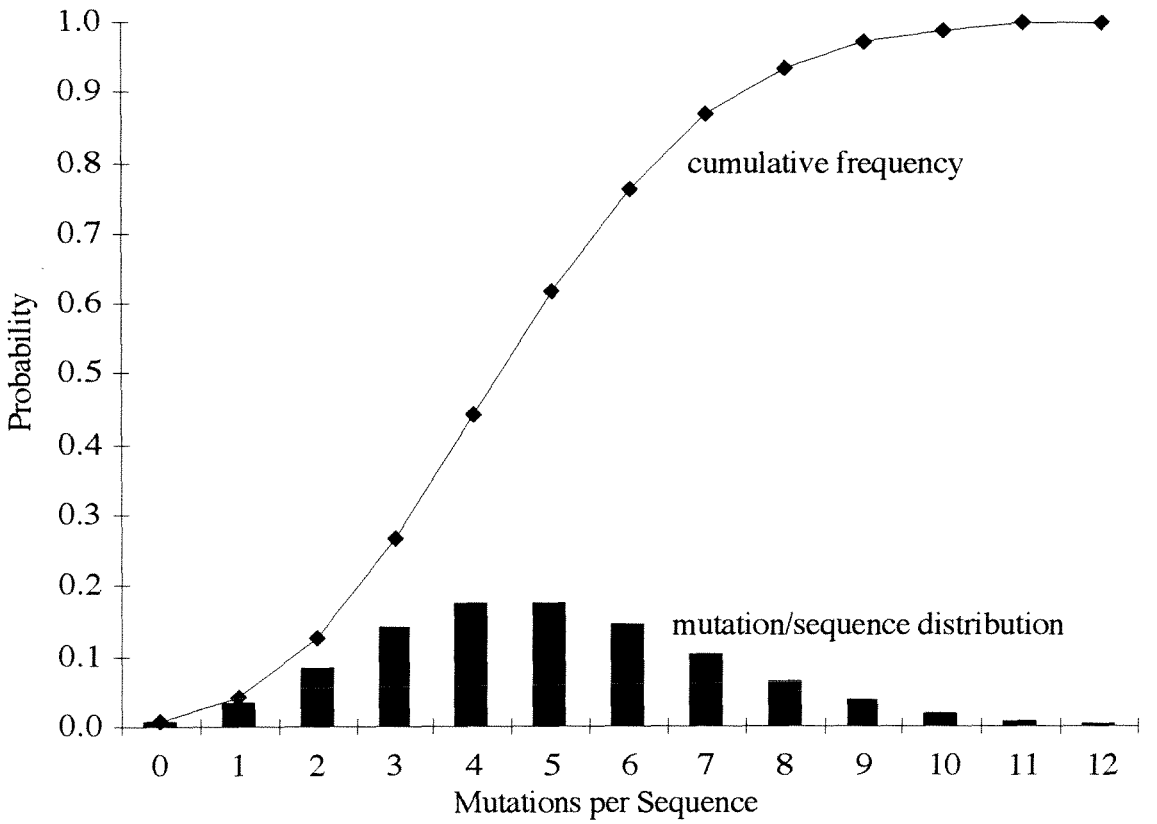



Figure 4.2: Calculated distribution of the number of mutations per sequence based on the Poisson distribution and the population of sequences averaging five mutations per sequence. The cumulative frequency curve shows the sum of the probabilites of the mutation per sequence distribution as a function of increasing mutations per sequence.

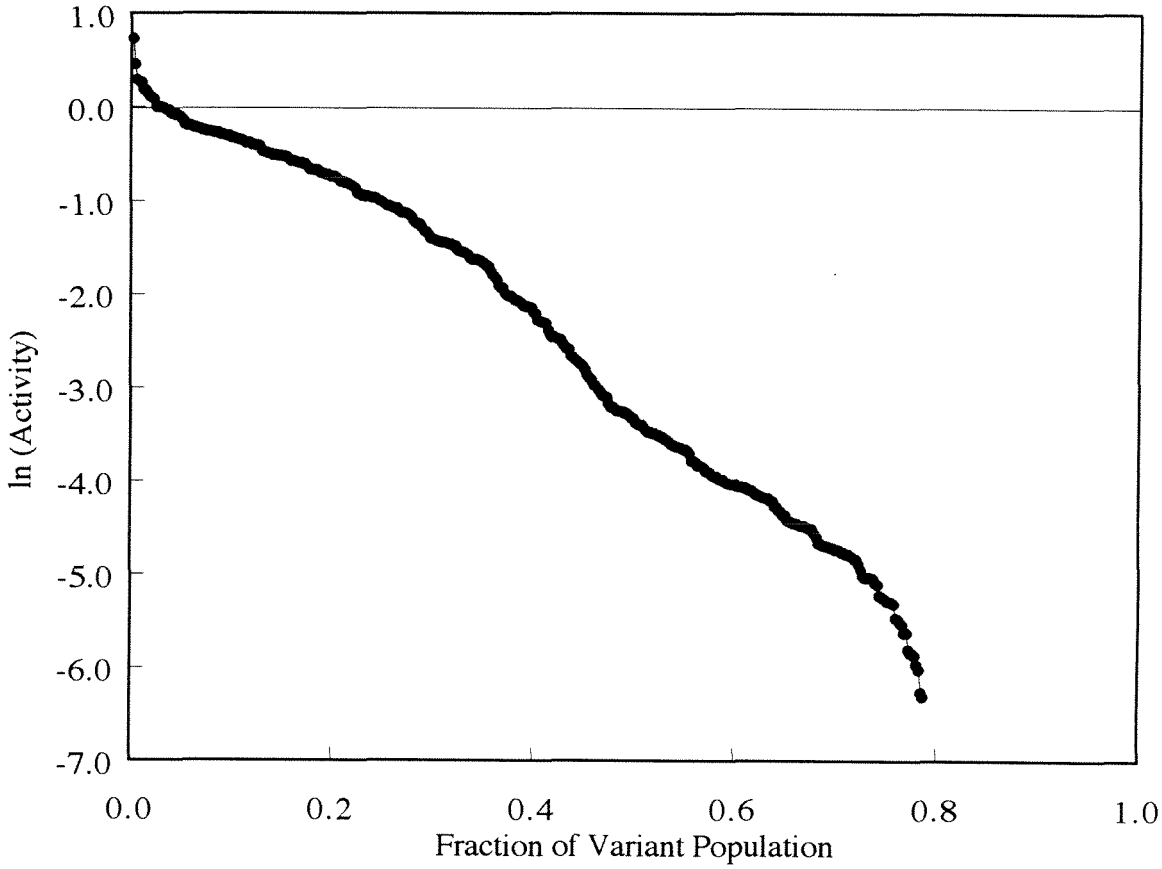



Figure 4.3: Free energy of activation profile for the first generation pNB esterase clones. Variants exhibiting zero or less activity are not shown.

enzyme on this substrate. The wild type enzyme's natural substrates are not known, but they are not likely to include pNPA. The clone demonstrating the highest activity against pNPA (1-1H9) was also the best clone tested against the p-nitrobenzyl loracarbef nucleus substrate (LCN-pNB). It was therefore used to parent the next generation.

The second generation clones screened against pNPA yield the profile in Figure 4. Approximately one-third of the clones show no measurable activity on this substrate (background hydrolysis rates for this substrate under the screening conditions are between 10 and 15-fold lower than the wild type activity). A reduction in the estimate of the number of mutations per sequence to 3.6 mutations per 1500 bp in this generation (based on three sequenced variants) corresponds to the expectation of 30% of the population containing five mutations or more (Figure 5). The free energy of activation plot (Figure 6) again suggests that mutations in the 1-1H9 parent reduce activity toward pNPA, and shows that 1-1H9 is still not the best enzyme in terms of activity on this substrate. Additionally, only approximately 1% of the clones appear more active than 1-1H9, fewer than the 3% in the first generation.

In the middle of screening the second generation mutation library, the screening substrate was switched from pNPA to LCN-pNP. The clones screened on LCN-pNP yield the activity profile shown in Figure 7 and the free energy of activation profile in Figure 8. We expect the profiles to be similar in shape and appearance to the pNPA profile in Figures 4 and 6 as a result of screening the same genetic pool. While the activity profile is similar, some differences exist. The activity profile (Figure 7) does not come to zero as in Figure 4 but does demonstrate a flattening at a position near the activity detection limit. Calling these activities zero then corresponds to a population of which one-third is approximately inactive. Based on the estimated mutation rate, the distribution in Figure 5 predicts that approximately 3% of the clones should be mutation-free (i.e., 1-1H9), 10% should be single mutation sequences, and 18% should be double mutation sequences. These

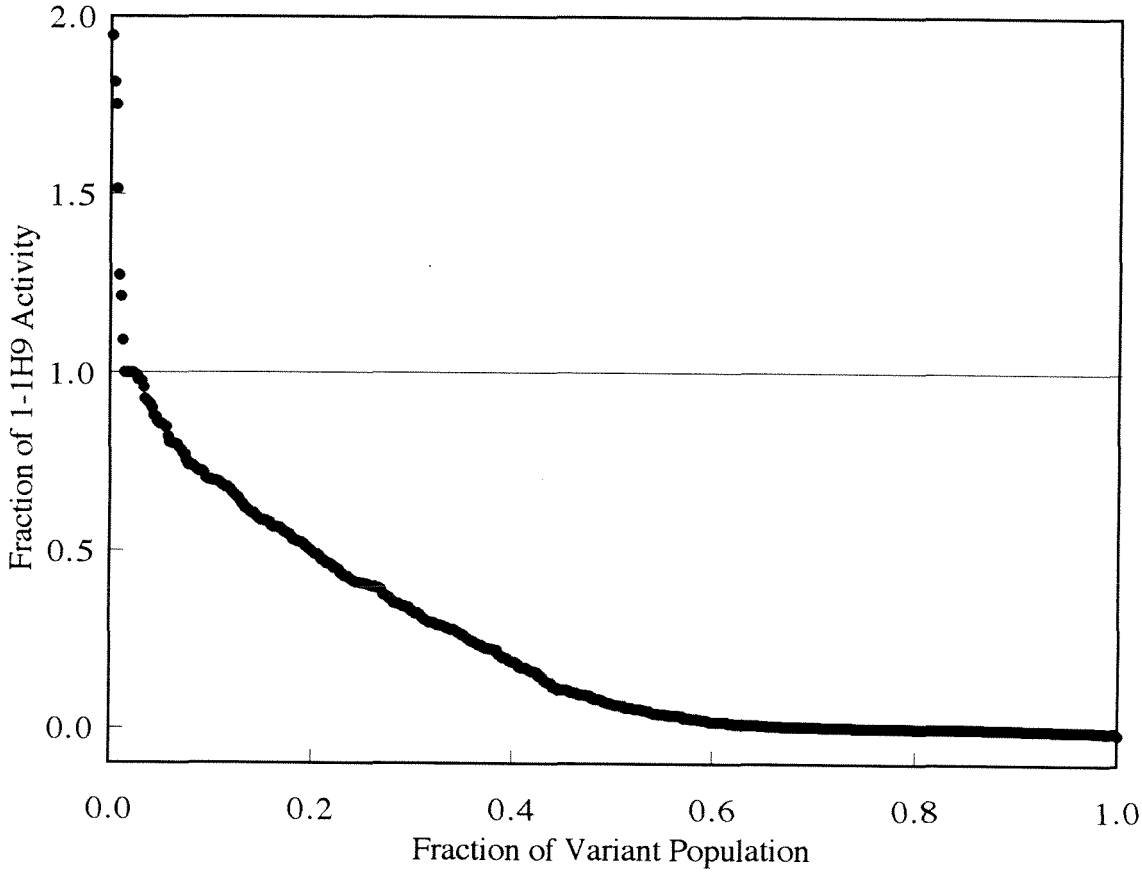



Figure 4.4: Activity profile of second generation pNB esterase variants. A total of 567 variants screened on pNPA in 20% DMF are represented.

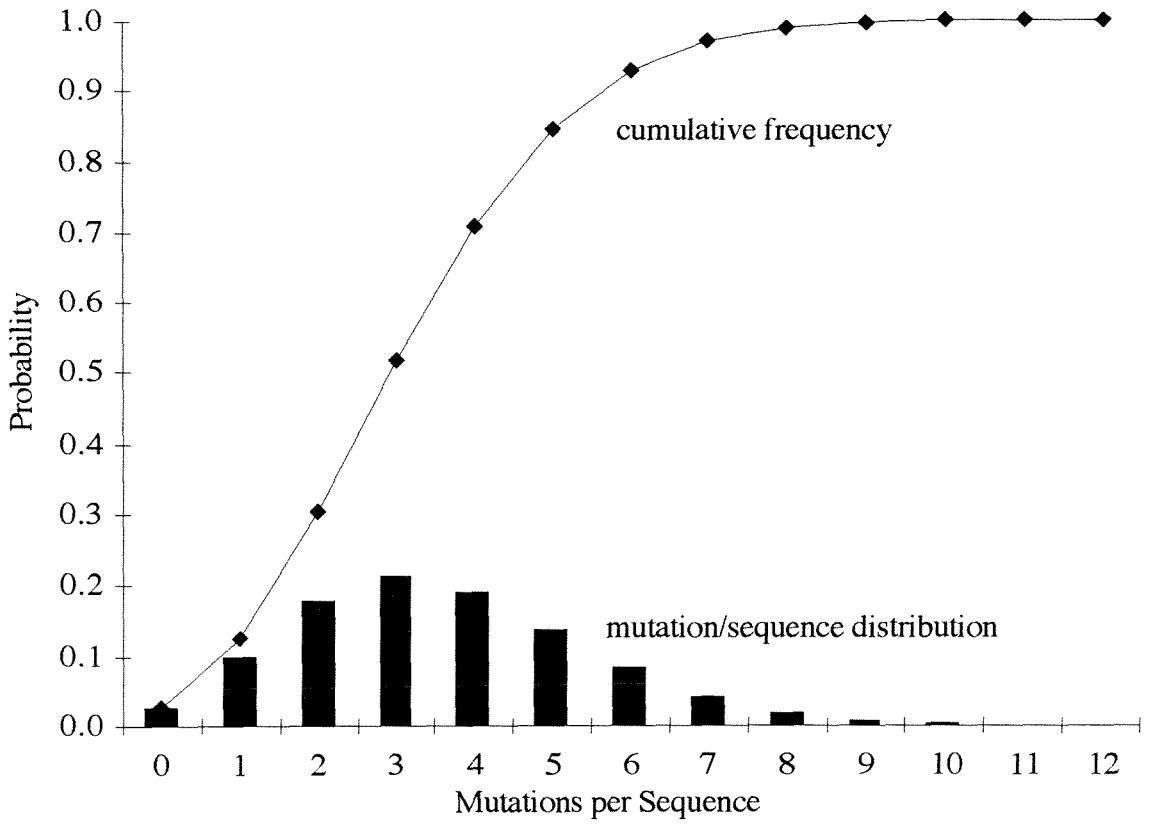



Figure 4.5: Calculated distribution of the number of mutations per sequence based on the Poisson distribution and the population of sequences averaging 3.6 mutations per sequence. The cumulative frequency curve shows the sum of the probabilites of the mutation per sequence distribution as a function of increasing mutations per sequence.

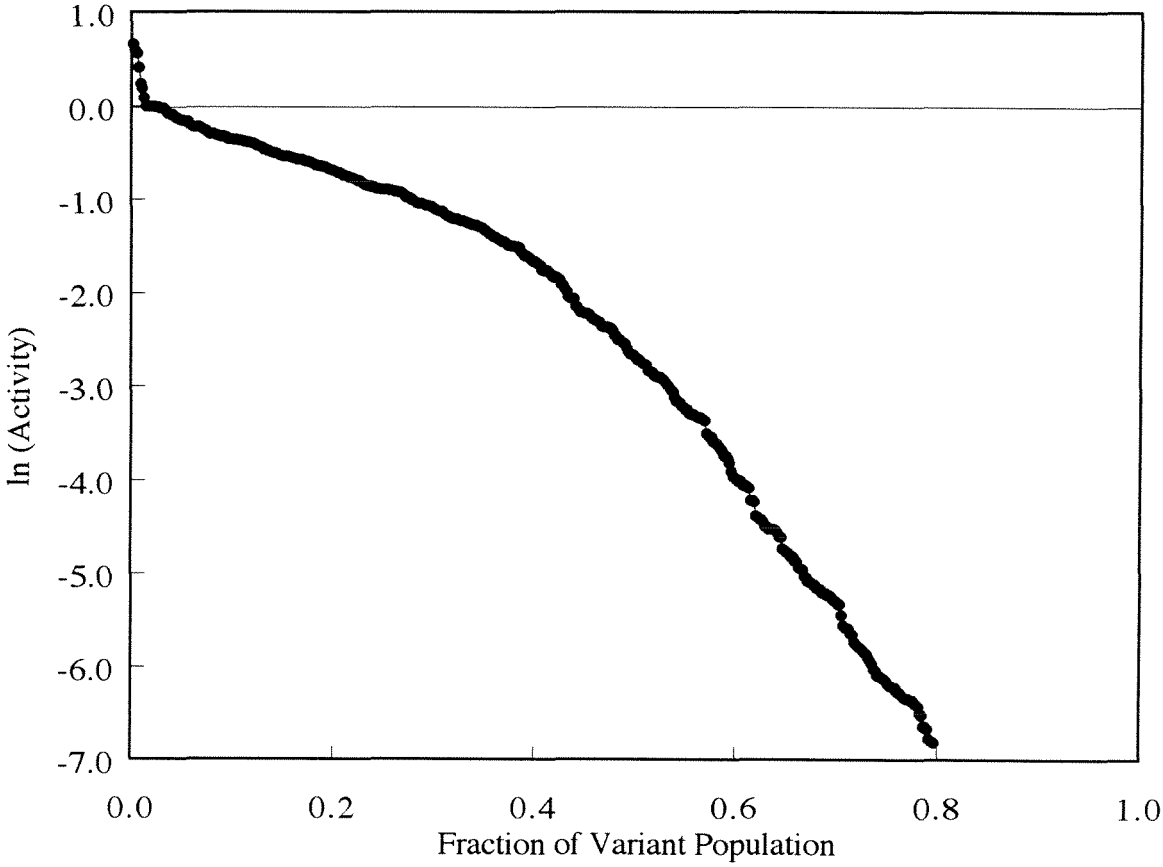



Figure 4.6: Free energy of activation profile for the second generation pNB esterase clones screened on pNPA. Variants exhibiting zero or less activity are not shown.

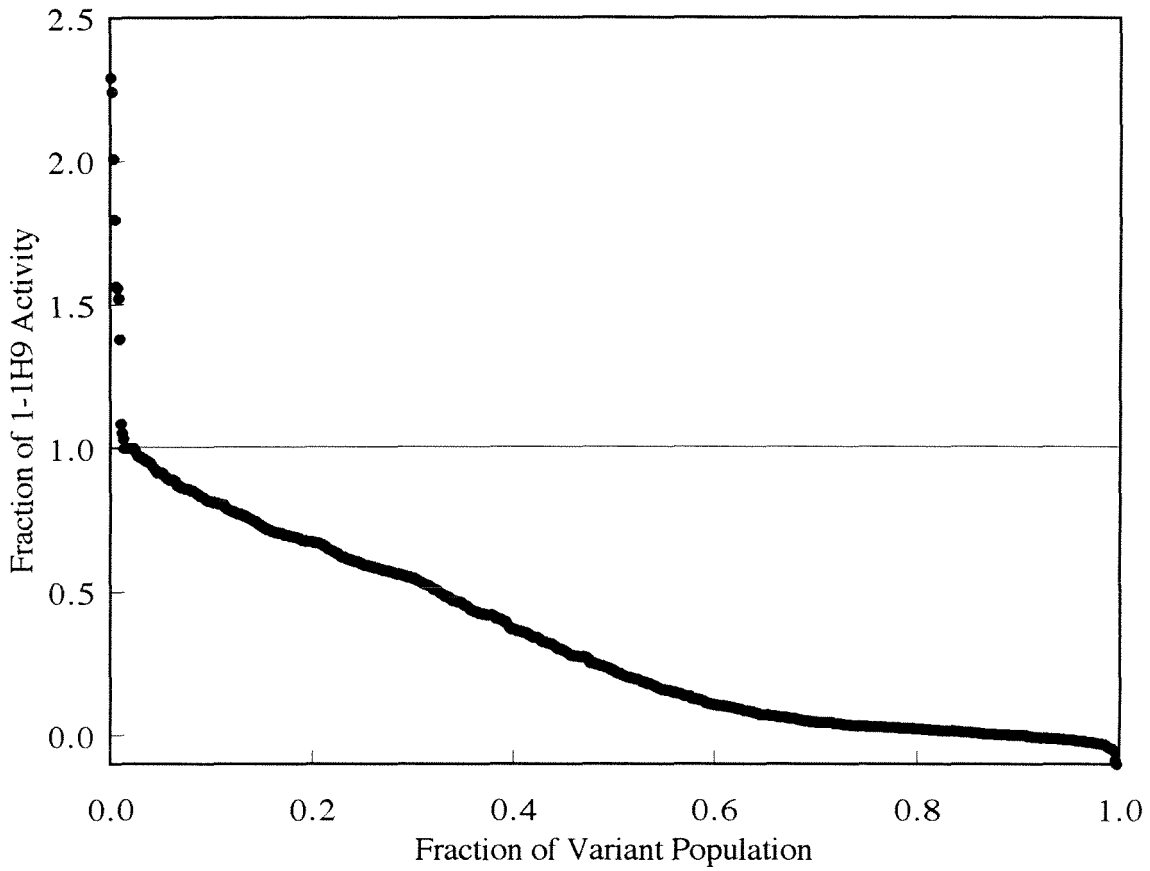



Figure 4.7: Activity profile of second generation pNB esterase variants. A total of 855 variants screened on LCN-pNP in 15% DMF are represented.

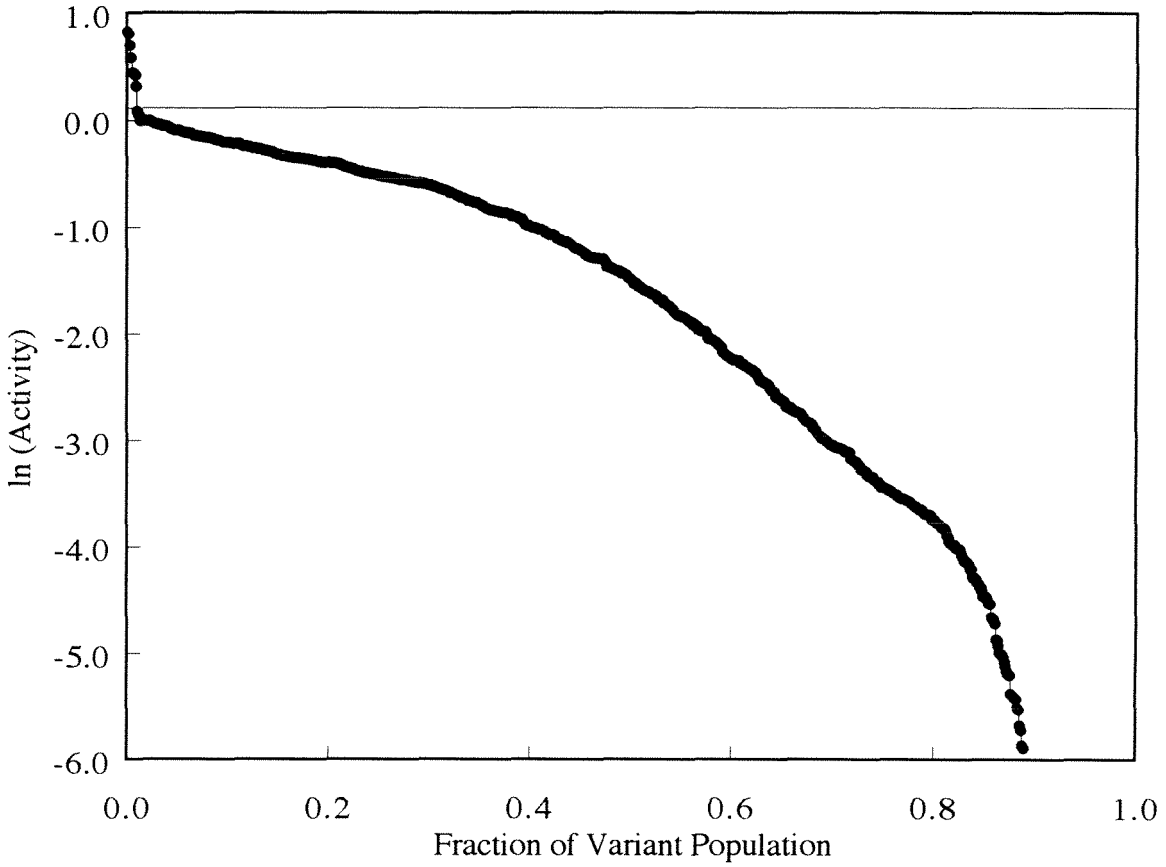



Figure 4.8: Free energy of activation profile for the second generation pNB esterase clones screened on LCN-pNP. Variants exhibiting zero or less activity are not shown.

sequences are less likely to contain activity-destroying mutations and are more likely to display neutral or partially decreased activities.

The third and fourth generations were screened exclusively on LCN-pNP and the resulting activity profiles are shown in Figures 9 and 10. Generations 3 and 4 both yield approximately 20-25% inactive clones. Figures 11 and 12 are the free energy of activation profiles. The sequence information for both the third and fourth generations suggests an average of two mutations per sequence (sequence data from one third generation and five fourth generation positive clones), and the Poisson distribution predicts the distribution in Figure 13. The calculated distribution suggests that the inactive sequences are predominantly of four mutations and higher with the three mutation sequences present at some lower frequency. Figures 10 and 12 also show the locations of the parents from each previous generation on the distributions. The location of the parents, especially on the activity profile in Figure 10, demonstrate the improvement from generation to generation. The increase in activity is close to a doubling with each generation. Thus, if a 16-fold increase in activity is required, having 8-fold activity is not half-way to 16-fold activity; in evolutionary terms, it is one step short of completion. The ever-increasing activity from generation to generation demonstrates the progression of fitness from one local landscape to the next.

All of the free energy of activation profiles demonstrate a behavior (in shape and position of the parent within the profile) similar to the free energy of deactivation profile for thermostability of a thermostable catalase (Trakulnaleamsai et al., 1995). If we assume general additivity for the positive mutations, then the fitness landscape becomes a single smooth peak, and the position of the parent in the fitness profile is a measure of how close the parent is to the top of this peak. In the final generation of asexual evolution (Generation 4, Figure 10), we see that as many as 4% of the population demonstrate more activity than the 3-10C4 parent, indicating that a number of positive steps toward the peak exist. This

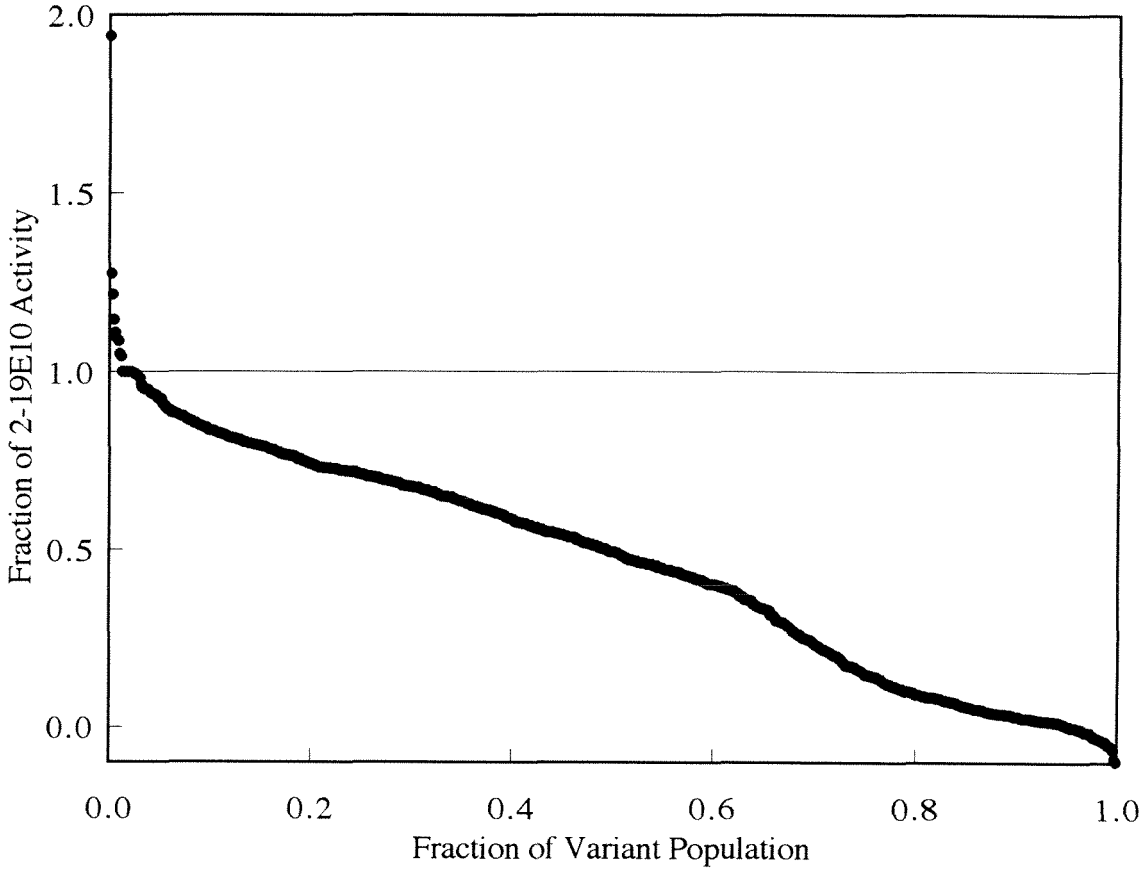



Figure 4.9: Activity profile of third generation pNB esterase variants. A total of 856 variants screened on LCN-pNP in 20% DMF are represented.

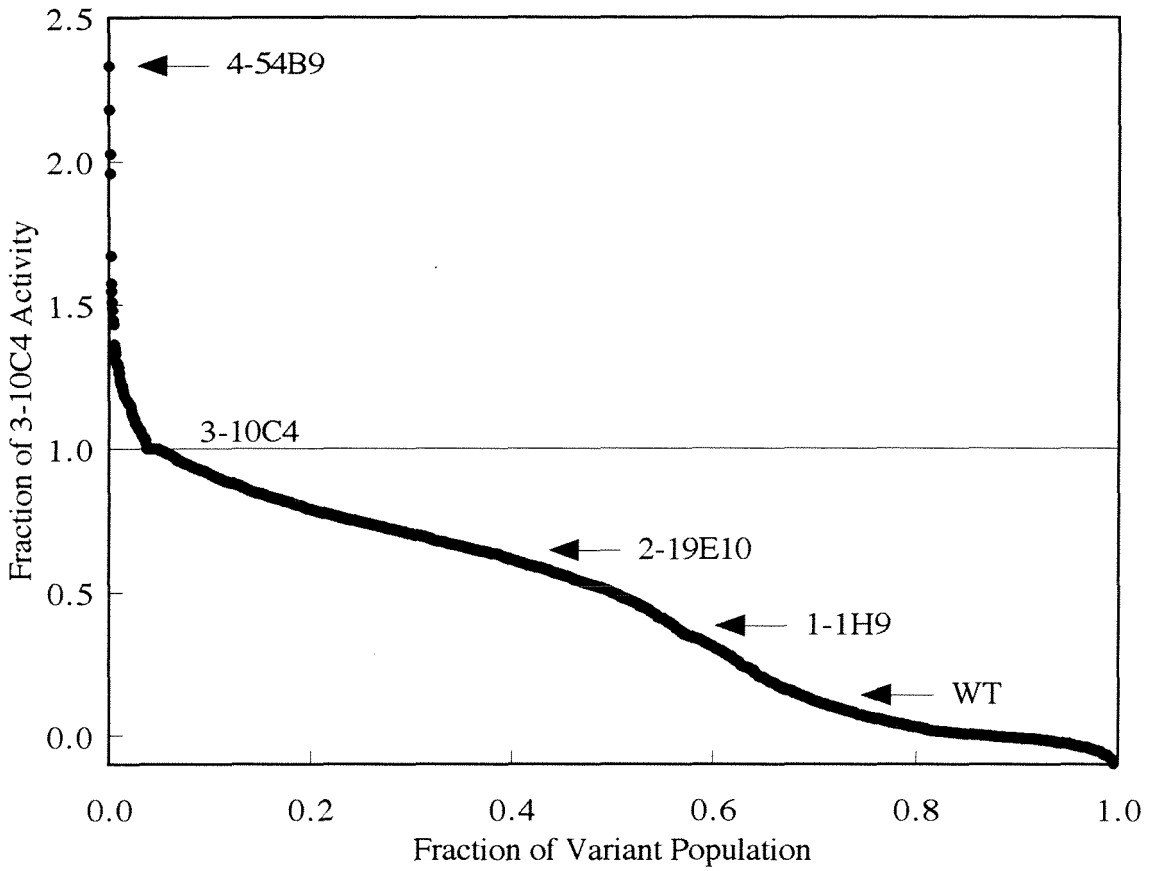



Figure 4.10: Activity profile of fourth generation pNB esterase variants. A total of 2375 variants screened on LCN-pNP in 20% DMF are represented. The parents of previous generations are also indicated.

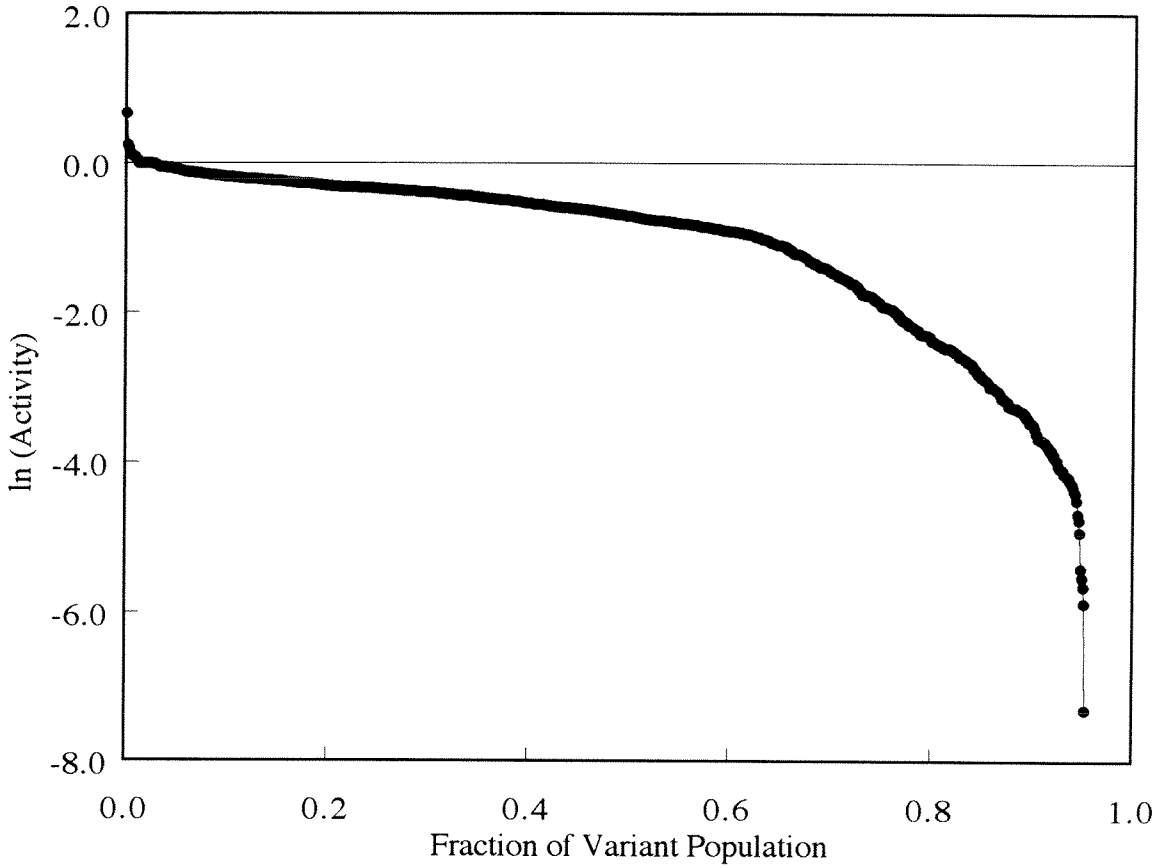



Figure 4.11: Free energy of activation profile for the third generation pNB esterase clones screened on LCN-pNP. Variants exhibiting zero or less activity are not shown.

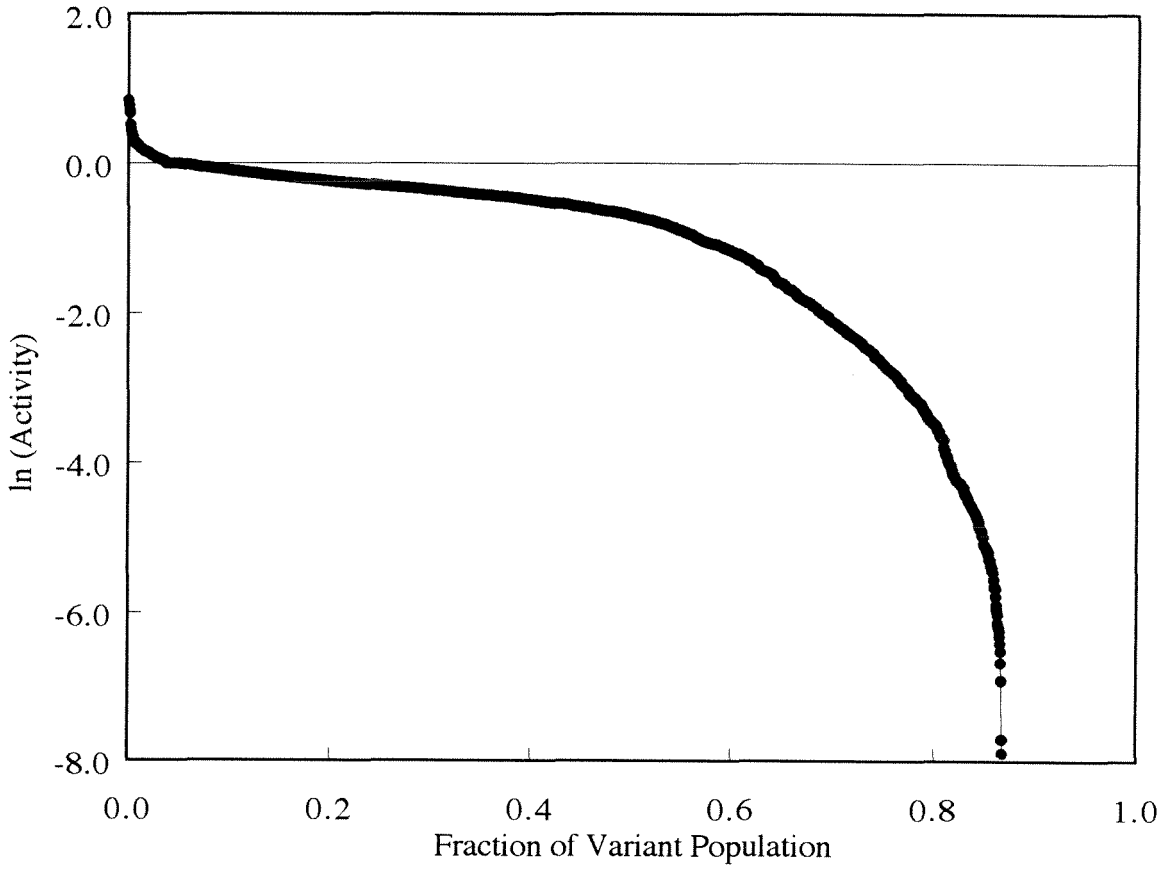



Figure 4.12: Free energy of activation profile for the fourth generation pNB esterase clones screened on LCN-pNP. Variants exhibiting zero or less activity are not shown.

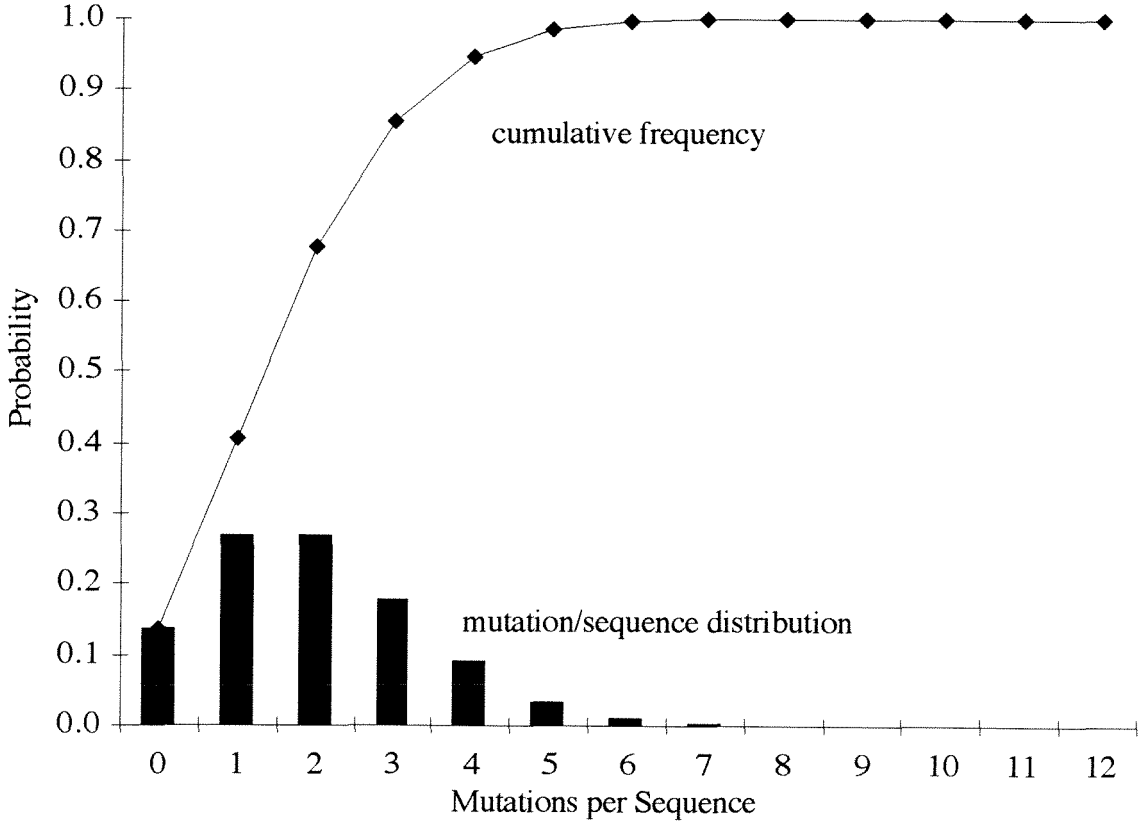



Figure 4.13: Calculated distribution of the number of mutations per sequence based on the Poisson distribution and the population of sequences averaging two mutations per sequence. The cumulative frequency curve shows the sum of the probabilites of the mutation per sequence distribution as a function of increasing mutations per sequence.

corresponds well to our impression that the catalytic efficiency on this substrate and under these conditions is on the order $10^2$ or $10^3$, still orders of magnitude from diffusion limitation ($\sim 10^8$). One method to test this additivity assumption is through the use of recombination.

The fifth generation of more active clones was the result of just such a recombination. The best five clones from generation four were recombined and screened to produce the activity profile in Figure 14. The activities were normalized to the activity of the most active parent, 4-54B9. Approximately 20% of these clones are inactive, similar to the third and fourth generation (Figures 9 and 10). For the recombination process to generate inactive variants requires that some combinations of mutations existing in highly active enzymes produce an inactive variant. This explanation for the generation of inactive clones seems less likely than simply attributing the inactive clones to the mutagenesis rate associated with DNA shuffling. The free energy of activation profile (Figure 15) is different from those of the third and fourth generation (Figures 11 and 12) in the region pertaining to clones with higher free energy than the parent. This region does not contain an abrupt transition to clones with higher activity, but simply extends the trend seen below the 4-54B9 free energy level. This highlights the fact that, given a pool of positive clones whose properties are additive, recombination is more effective at combining these mutations than random mutagenesis is at creating them. As a result, a significant increase in the number of more active variants as a percentage of the population is seen.

The previous activity profiles screened on LCN-pNP (Figures 7, 9, 10, and 14) were divided into the following groups: clones containing zero or less activity, clones containing from zero to 50% of the parent activity, clones containing from half to 100% of the parent activity, clones containing from 100% to 150% of the parent activity, and so on. Counting the number of clones in each category, a distribution was obtained. The midpoint of each activity category was used to represent the activity of all the clones in that category,

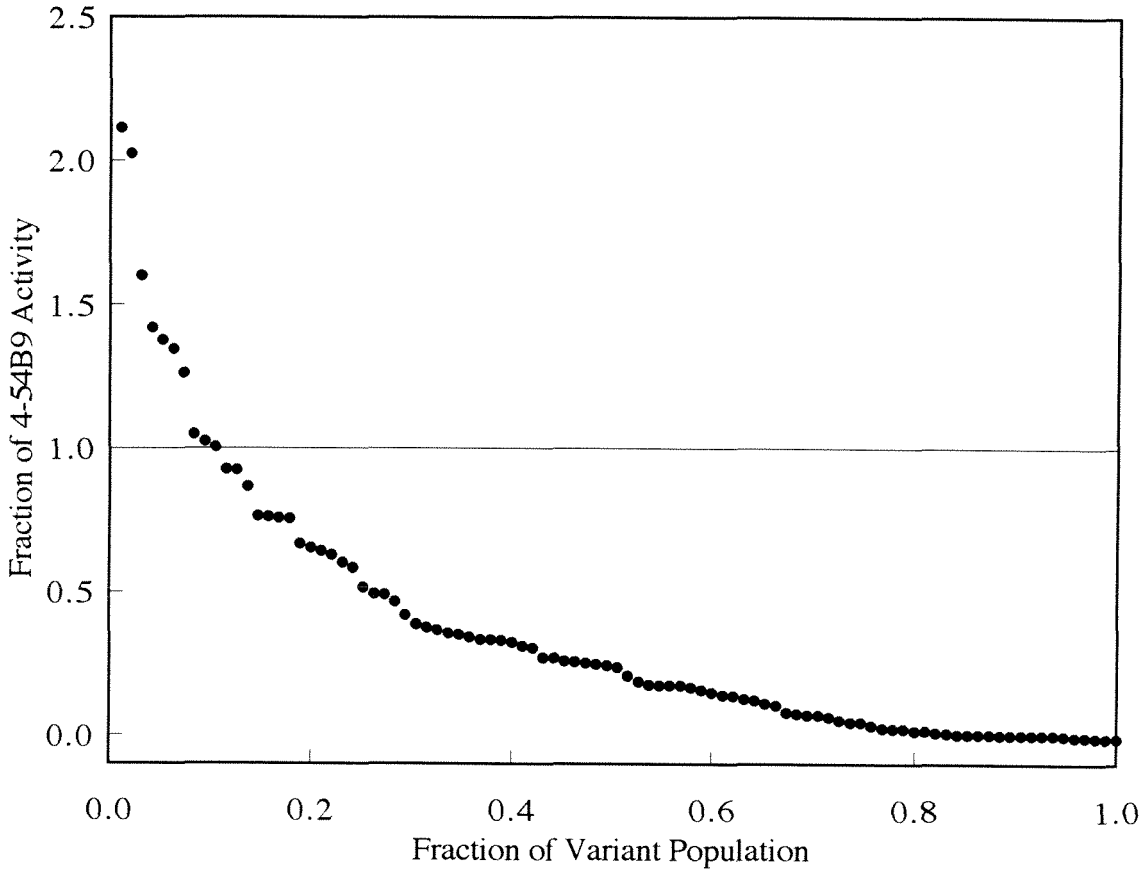



Figure 4.14: Activity profile of fifth generation pNB esterase variants. A total of 95 variants screened on LCN-pNP in 25% DMF are represented.

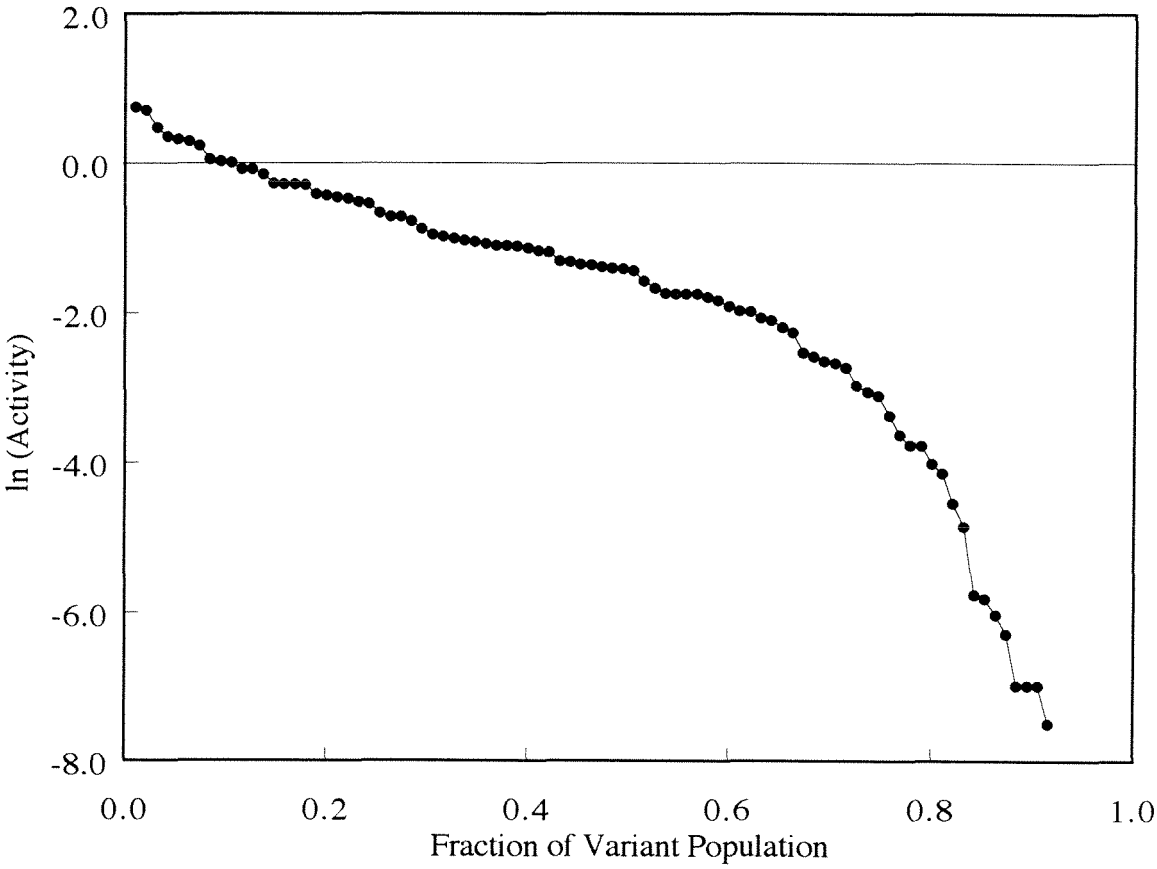



Figure 4.15: Free energy of activation profile for the fifth generation pNB esterase clones screened on LCN-pNP. Variants exhibiting zero or less activity are not shown.

with the exception that the clones with less than zero activity were represented by zero. The activities were then multiplied by the parent activity to generate real values for activity (not scaled as a percentage of the parent). Finally, the real activities were rescaled to the value of the parent of the fourth generation. The resulting plot (Figure 16) demonstrates the movement of the asexually-evolved populations (generations 2, 3, and 4) towards higher activity with increasing generation. The highest data point of each of these generations is an approximation of the clone used to parent the next generation. It can be seen that the next generation's population contains only a small percent of clones with greater values. In contrast, the sexual population (generation 5) recombines the information contained in the several positives to greatly increase the percentage of the population demonstrating higher activity.

Additional experiments performed by others in our group demonstrate the mutagenesis associated with DNA shuffling (Jin, Moore, and Arnold, to be published). Individual sequences were shuffled with themselves to produce asexual (single parent) populations by DNA shuffling. The activity profiles generated are shown in Figure 17 and are similar to the types of activity profiles discussed here for error-prone PCR. The profiles for shuffling 4-53D5 with itself and 4-54B9 with itself demonstrate few inactive clones, similar to those profiles of generations containing only a few mutations per sequence (Figures 9 and 10). The two profiles of 4-43E7 contain a slight increase in the number of inactive clones. Finally, the profiles of 4-38B9, due to the tremendously high fraction of inactive clones, suggest either a very high rate of mutagenesis, a problem in the sequence as related to how DNA shuffling generates full length sequences, or a problem in the cloning of the resulting products into the expression vector.

Repeating the DNA shuffling which produced the fifth generation results (Figures 14 and 15) and screening ten times as many colonies produced the profile shown in Figure 18 (Jin, Moore and Arnold, to be published). This experiment displays a decreased

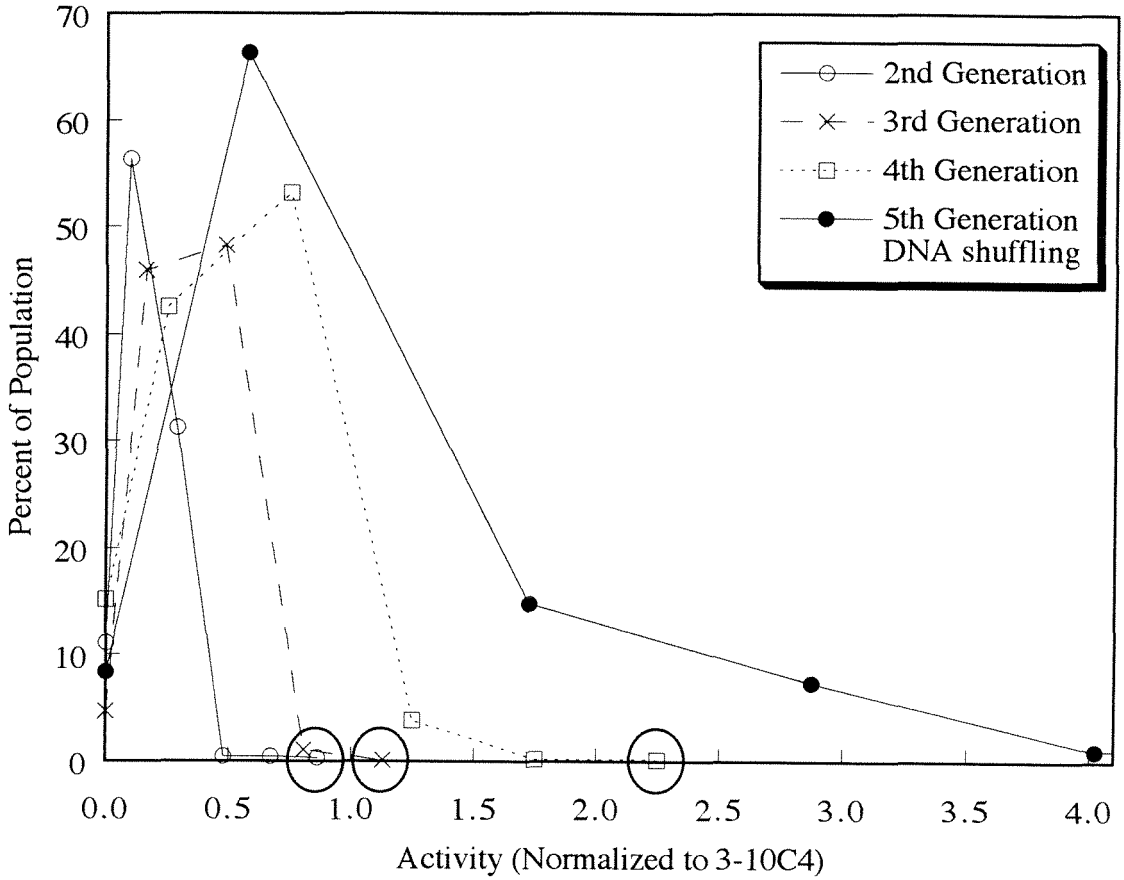



Figure 4.16: The progression of increased activity of the variant populations in each generation. The histograms were created by scaling the activity of each variant in each activity profile (Figures 4.7, 4.10, 4.11, and 4.15) to 3-10C4, then sub-dividing the activity profiles into groups and counting the number of clones in each group. The midpoint of the group was used as a representative value for activity. The group containing the clone from each generation used to parent the next generation is circled.

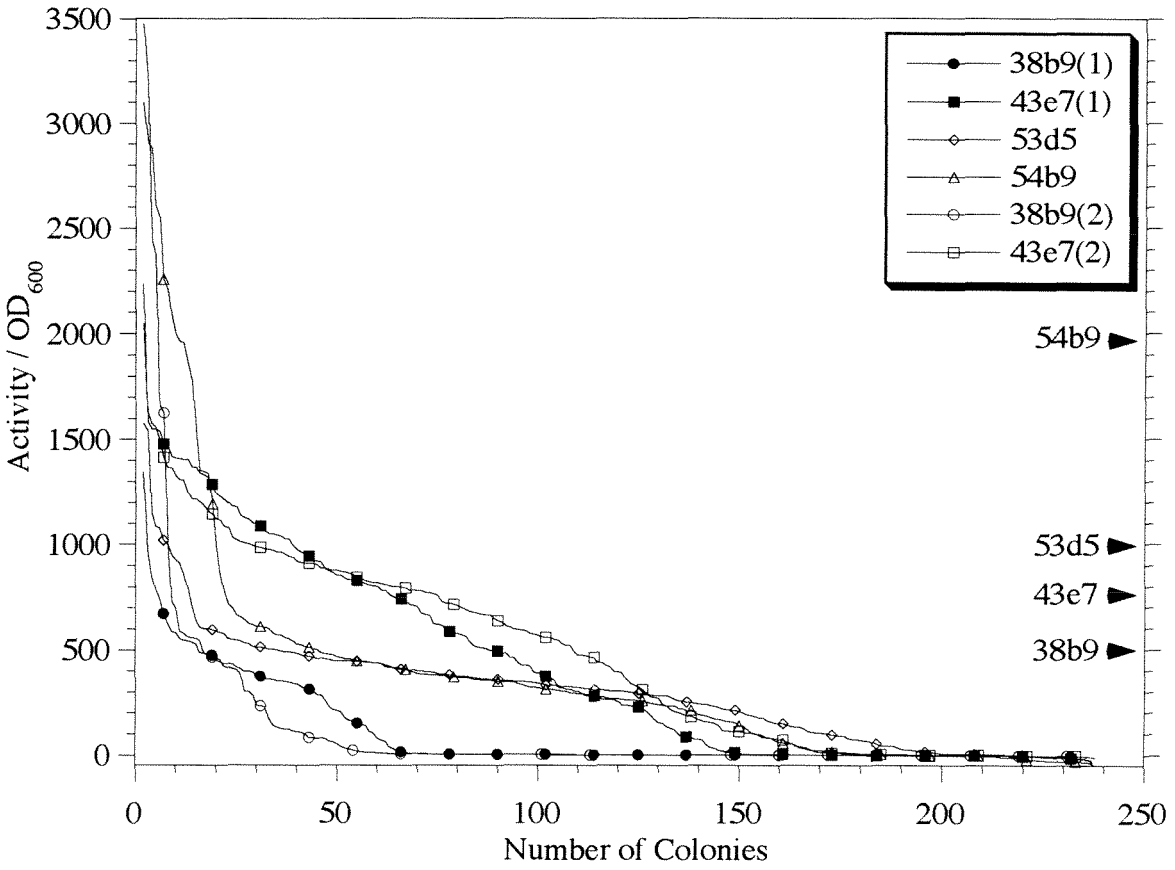



Figure 4.17: Activity profiles generated by screening 236 colonies of pNB esterase variants created by DNA shuffling each parent sequence with itself. The activities of the individual parents are indicated.

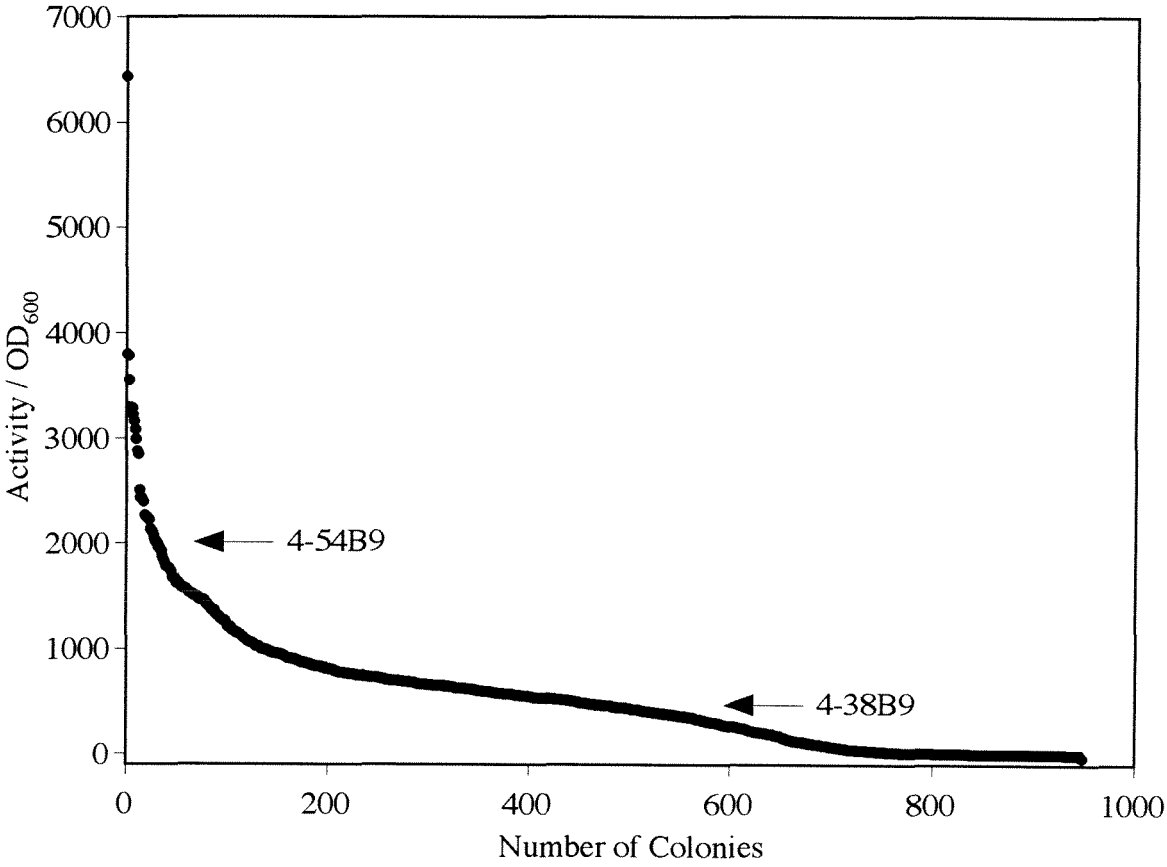



Figure 4.18: Activity profile generated by screening 950 clones created by DNA shuffling the four parents 4-38B9, 4-43E7, 4-53D5, and 4-54B9. The most active (4-54B9) and the least active (4-38B9) parent activities are indicated.

frequency of positive variants and an increased number of inactive variants as compared to the previous fifth generation activity profile (Figure 14). One of the positive variants, however, appears significantly more active than the positives from the previous experiment as a result of screening significantly more clones.

Shuffling the clones exhibiting substantially more activity than 4-54B9 from the population shown in Figure 18 together, a new population was generated. In a separate experiment, the best clone generated by each of the asexual shufflings of individual parent sequences with themselves were shuffled together to generate a new population. The comparison of these two new populations is shown in the activity profiles in Figure 19. Again the profiles look similar to all of the previously described shuffling profiles. The two populations being shuffled are distinctly different in the types of mutations the sequences should contain. In the two rounds of sexual recombination case which started with four unique sequences, the best sequences after the initial round of shuffling should contain a high number of similar sequences as a result of combining positive mutations and eliminating neutral or deleterious ones. Shuffling these new sequences in a second recombination should generate an increased fitness in the overall population. In the case of an initial round of asexual shuffling, introduction of new mutations creates increased diversity in the positives selected. When these positives are then shuffled sexually, the increased diversity increases the number of different progeny produced, and when screening a limited number of clones, this decreases the likelihood of finding the extremely active positives. This is reflected in the small decreased fitness of this population as compared to the two rounds of recombination. The increased initial diversity should still be at least partially reflected in the positives selected, and as a result we would expect that a second round of sexual shuffling on these positives would provide populations more fit than the two sexual recombination case.

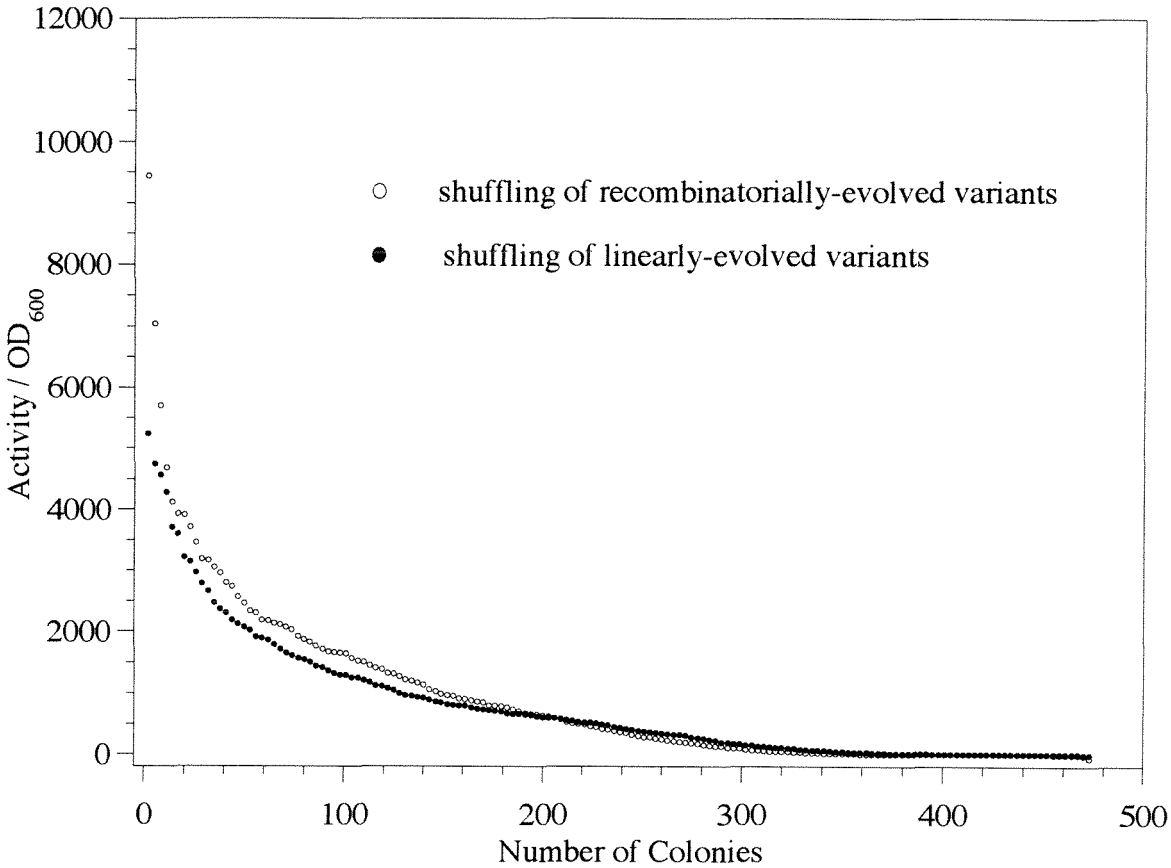



Figure 4.19: Activity profiles generated by screening 475 pNB esterase clones created by DNA shuffling the positives from the Figure 4.17 (multiple parent pools) or Figure 4.16 (single parent pools).

Because some of the screening during the second generation was performed on two substrates, the degree of correlation between enzymatic activity on these two substrates can be evaluated. Figure 20 demonstrates 95 randomly chosen clones and their activity on each substrate, as normalized by one clone demonstrating wild type behavior on both substrates. This figure demonstrates a reasonably strong correlation between activity on pNPA and LCN-pNP, even though the two substrates seem quite different. The pNPA substrate contains only acetate as the carboxyl-containing group, while LCN-pNP contains a much larger heterocyclic ring structure consisting of a four- and a six-membered ring. The diagonal line drawn represents a perfect correlation of activities on the two substrates and is a good representation of the data. The clone circled demonstrates an enzyme that has reduced its ability to recognize and cleave the pNPA substrate in favor of the LCN-pNP substrate. Similar behavior has been observed for the correlation between catalase and peroxidase activities (Trakulnaleamsai et al., 1995) and demonstrates an altered catalytic specificity.

At the end of the fourth generation screening, the clones testing as more active than 3-10C4 were measured against the substrate for which activity was ultimately desired, LCN-pNB. The correlation between LCN-pNP and LCN-pNB from the selected clones is seen in Figure 21 (Chapter 2, Figure 7). The diagonal line indicates the trend expected if the correlation of activity between the two substrates were exact, and the data are represented by this line reasonably well. In Figure 20, the correlation between pNPA and LCN-pNP was examined, and a clone demonstrating a new specificity was circled. A clone demonstrating the dramatic difference in Figure 20 is not evident in Figure 21. This could be due to the non-random clones tested on LCN-pNB, the small number of clones, or the ability of LCN-pNP to effectively mimic LCN-pNB.

The fourth generation was screened in 5% and 20% DMF. The kinetic data on the variants isolated from generations one through four and the wild type enzyme indicate a

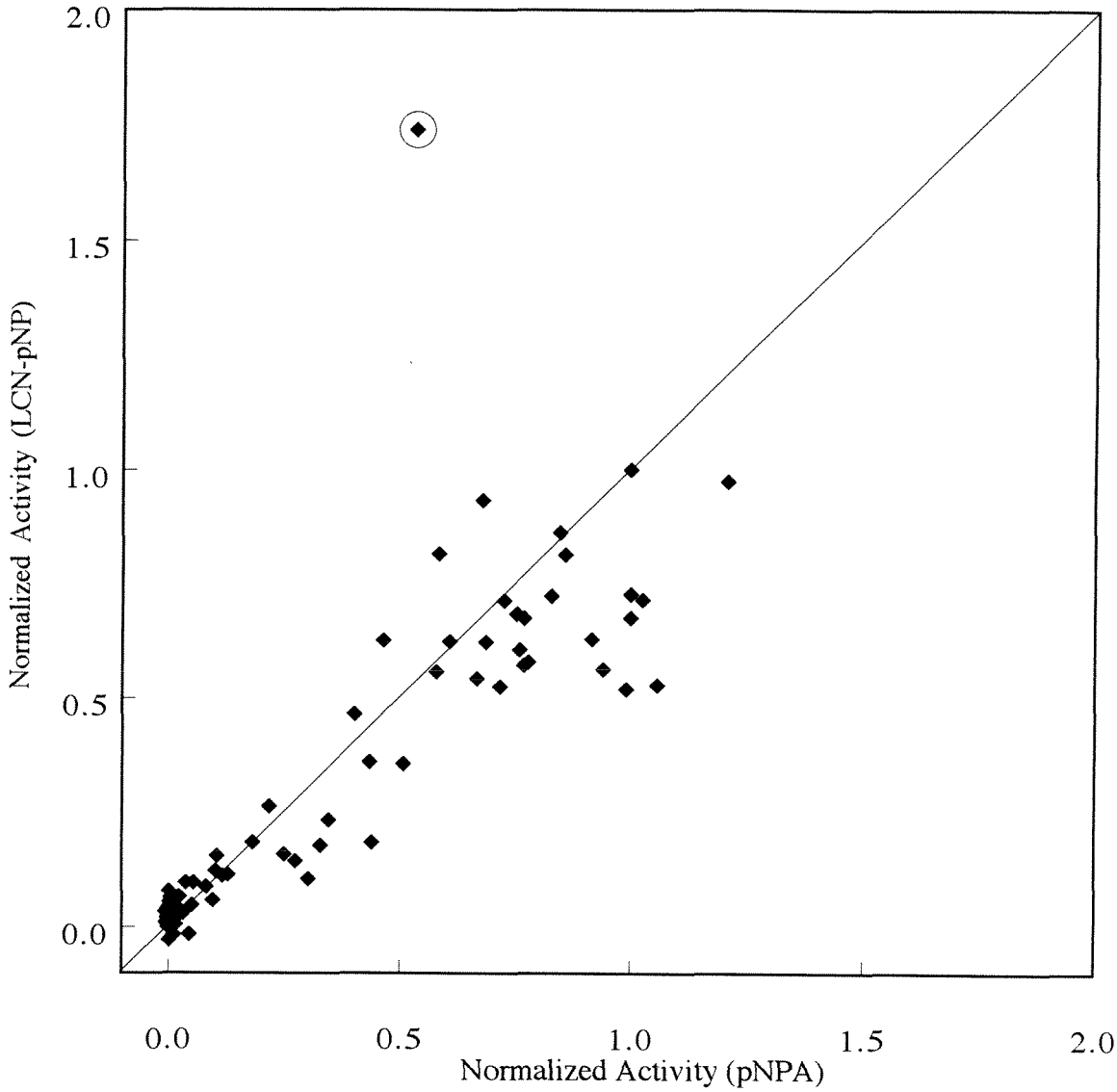



Figure 4.20: Correlation of esterase activity between pNPA and LCN-pNP. The activity of 95 randomly selected variants from the second generation were normalized to a clone exhibiting wild type behavior on both substrates. The diagonal line represents a perfect correlation between the two activities. The circled variant is one that represents a significant change in catalytic specificity.

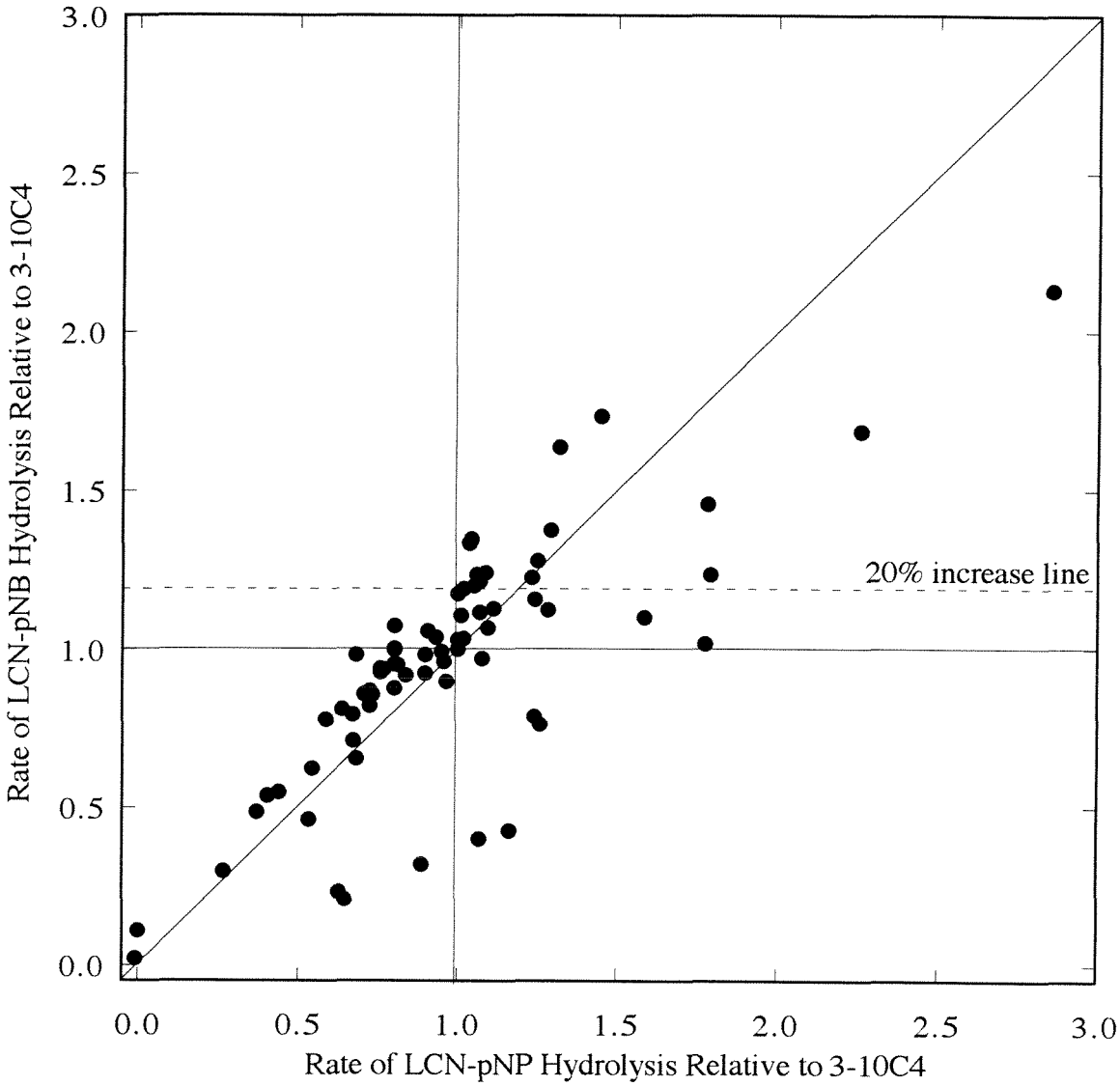



Figure 4.21: Hydrolysis rates of fourth generation pNB esterase variants between the screening and actual substrates. The rates are normalized to the third generation variant 3-10C4. After an 8 hour induction period, whole cell screening assays were performed at 25 °C in a 0.1 mM Tris-HCl pH 7.0, 15% dimethylformamide reaction solution containing 0.8 mM of either p-nitrophenyl loracarbef nucleus (LCN-pNP) or p-nitro-benzyl loracarbef nucleus (LCN-pNB).

shift in DMF's effect on enzymatic activity (Moore and Arnold, 1996). At low concentrations of DMF, the enzymes tested suffer from dramatic increases in $K_M$, while at high concentrations of DMF, $k_{cat}$ is more greatly affected. Figure 22 demonstrates the correlation between activities in 5% DMF and 20% DMF. The correlation between activities in the two environments is strong, and suggests that clones found to be more active in one environment should be equally more active in the second environment. Thus one might be able to identify several clones at 5% DMF and use recombination to accumulate several mutations into a single sequence in order to obtain higher activity in 20% DMF. This correlation is reflected in the study of enhanced protease activity in DMF by subtilisin BPN' (Chen and Arnold, 1992). Positives with increased activity were identified in 30% DMF and combined into a single sequence which was greatly improved in all DMF concentrations.

*Subtilisin BPN'*

The gene for subtilisin BPN' was subjected to error-prone PCR, and the resulting enzyme library was screened for increased stability to hydrogen peroxide ($H_2O_2$) at low pH (pH 4.5). Residual activity of approximately 2500 clones was measured as the ratio of activity in culture supernatants after a two hour incubation in 3% peroxide to activity after the same incubation in the absence of peroxide. Little to no decrease in activity was seen in the absence of peroxide over the incubation period (data not shown).

Figure 23 demonstrates the precision of the screening assay carried out only on wild type clones. Wild type generates an activity ratio of 0.025 with a standard deviation of 0.001 under these conditions. This corresponds to a 15% residual activity (due to a six-fold dilution in the residual activity plate - see Materials and Methods). The high level of reproducibility demonstrated in this assay is a direct result of normalizing activity after peroxide treatment with the non-peroxide treated sample. This removes variability

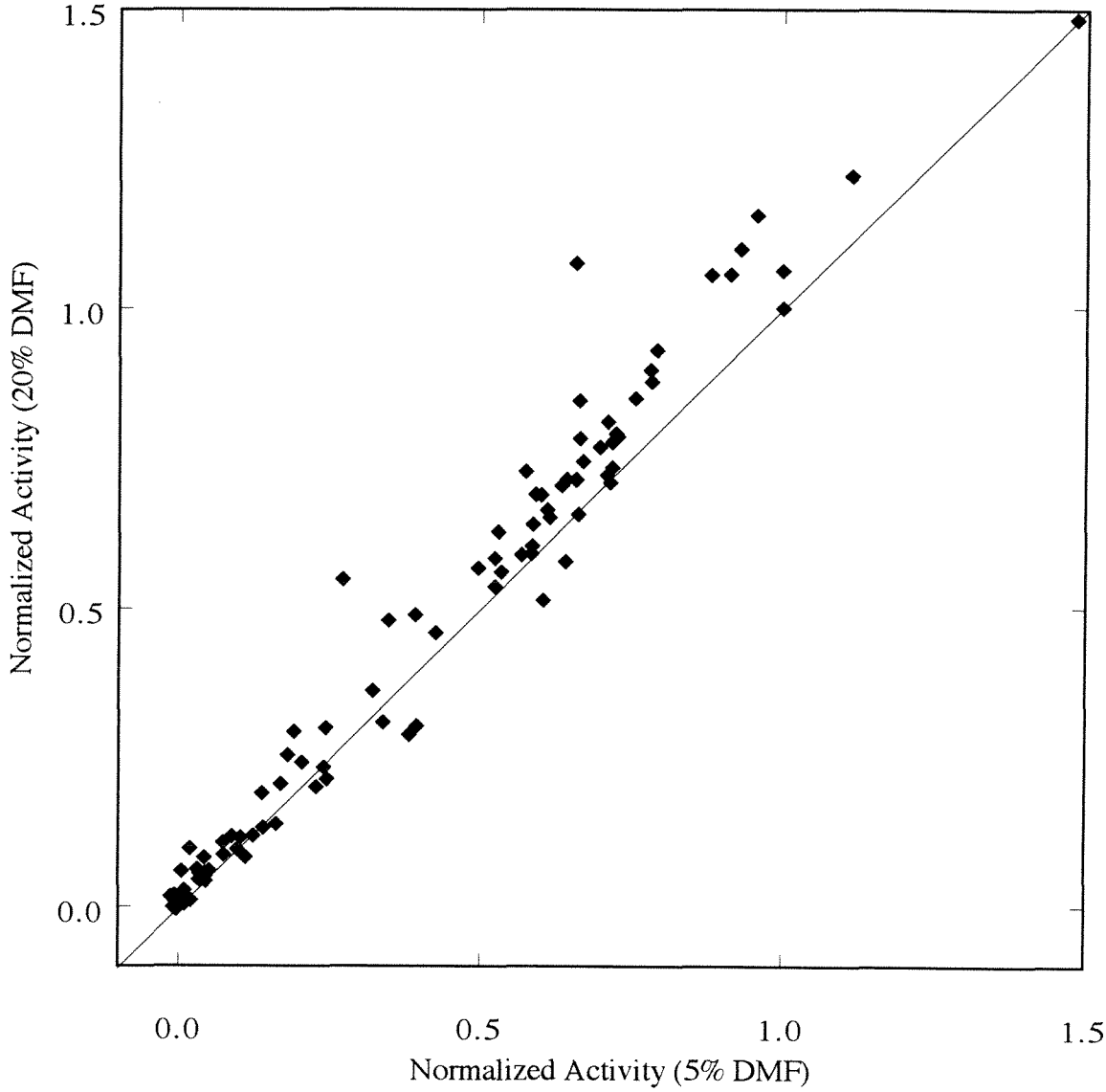



Figure 4.22: Correlation of pNB esterase activity on LCN-pNP between 5% DMF and 20% DMF environments based on 95 random clones from the fourth generation screening. The diagonal line represents a perfect correlation between the activities in the two environments.

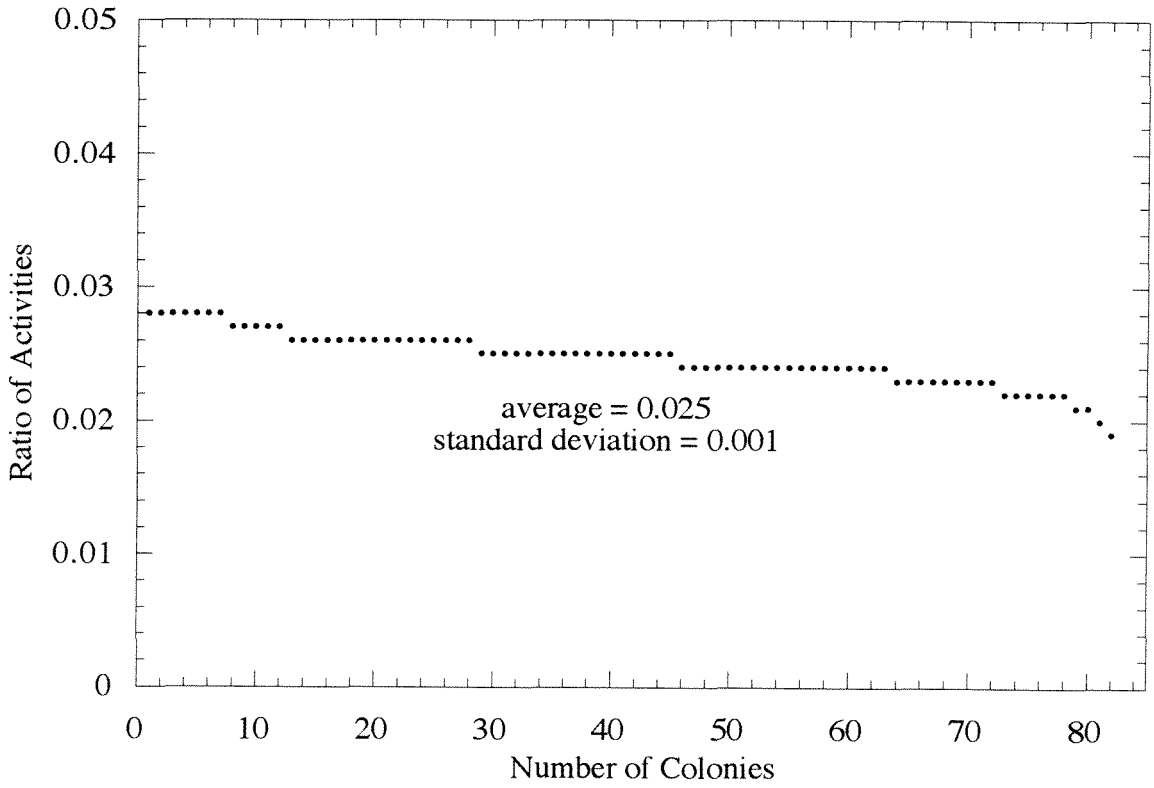



Figure 4.23: The results of screening 83 wild type subtilisin BPN' clones in the residual activity assay. The assay is performed by dividing the enzymatic activity present after a 2 hr hydrogen peroxide treatment at pH 4.5 by the initial activity. As a result of dilution effects, the ratio calculated is six-fold less than the actual residual activity.

associated with differences in expression levels, specific activities, or any culture-based differences. This level of reproducibility was not available in the pNB esterase screening, because changes in overall activity, rather than residual activity, were targeted.

Figure 24 shows the stability relative to the wild type subtilisin BPN' of the 18 positives isolated and retested after screening 2500 clones. One clone, 23H5, reproducibly demonstrates almost six times the stability of wild type. However, its activity before peroxide treatment is about seven-fold less than wild type. For the majority of positive clones, the total protease activity after peroxide treatment was between one and three times wild type activity under the same conditions.

Figure 25 demonstrates the activity profile of 940 subtilisin BPN' variants relative to wild type activity before treatment with peroxide at low pH. As much as a 2.5-fold increase in activity was observed, with approximately 8% of the population demonstrating more activity than wild type. These data are not normalized to cell density as was the case for pNB esterase; some of these positives near the wild type activity may be the result of differences in the bacteria concentration. Approximately 0.5% demonstrate significantly more activity (greater than 1.5 times) than the wild type sequence. Comparing the high fraction of inactive clones with the pNB esterase data suggests multiple mutations per sequence (9 mutations per sequence are suggested by the protocol used).

Because of the activity ratio used for assaying stability, clones demonstrating less than 10% of the wild type activity in Figure 25 were removed (497 clones) from the population as this data often gave extremely large residual activities as a result of division by extremely small activity values. Figure 26 represents the residual activity profile for 443 subtilisin BPN' clones after 2 hours in 3% hydrogen peroxide at low pH (pH 4.5). The wild type appears at the midpoint of the gently sloping transition between the number of clones more stable and those that are unstable. This profile is dramatically different than the pNB esterase and the subtilisin BPN' activity profiles. The location of wild type within

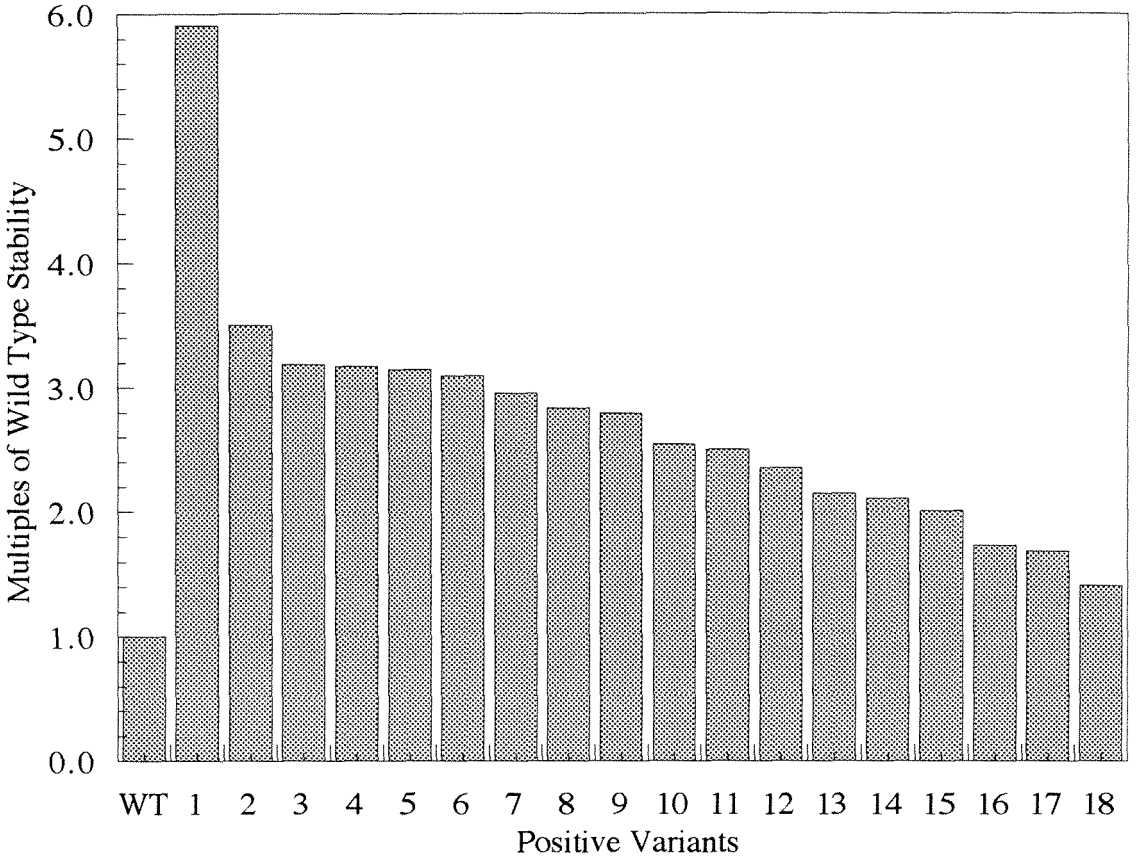



Figure 4.24: The stability of positive variants isolated as compared to wild type subtilisin BPN'.

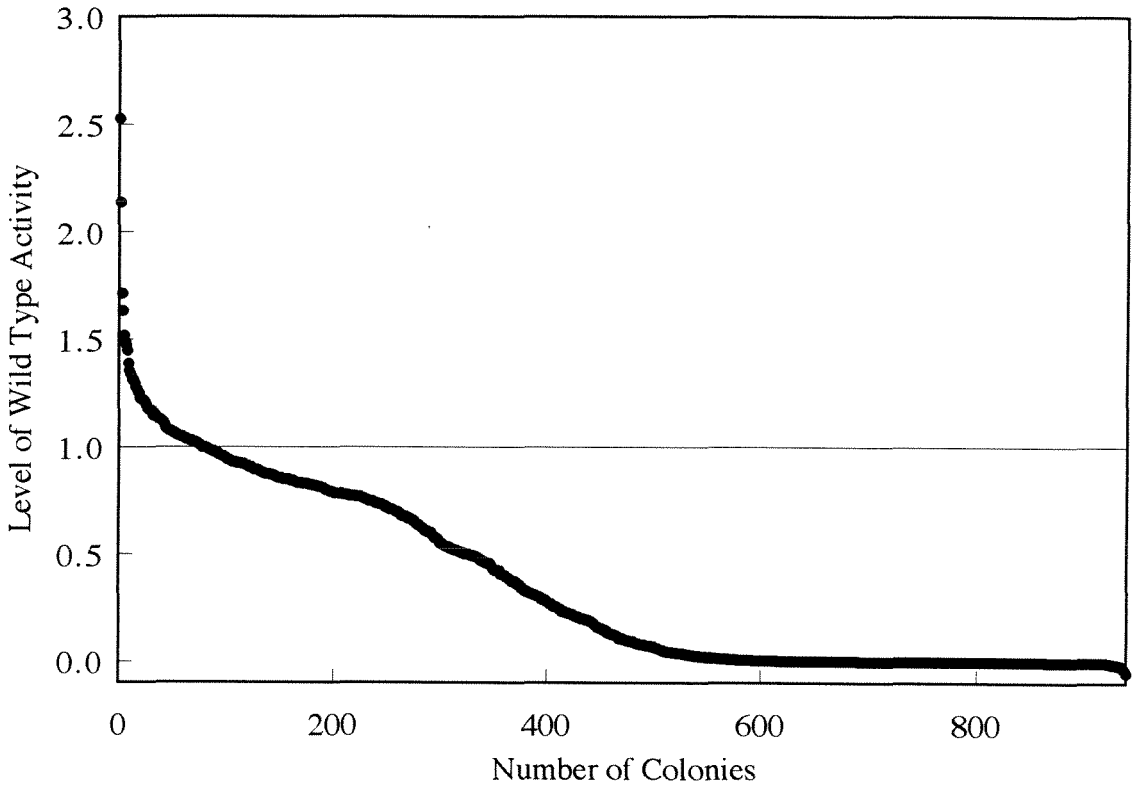



Figure 4.25: Activity profile of subtilisin BPN' variants. A total of 940 clones were screened on the peptide substrate succinyl-AAPF-pNA in tris buffer, pH 8.6.

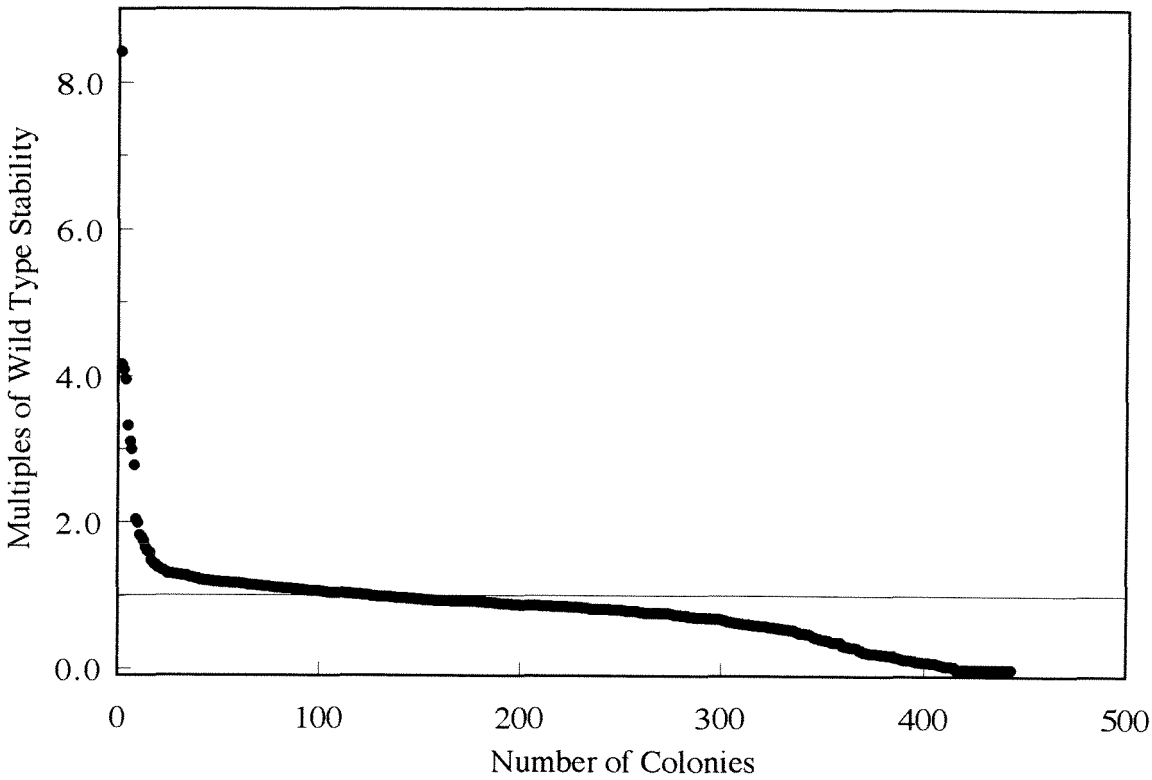



Figure 4.26: Stability profile of subtilisin BPN' variants to 3% hydrogen peroxide and pH 4.5. The 454 clones presented are those containing enough activity to be accurately measured from 900 total clones.

the profile suggests that subtilisin BPN's stability to peroxide has not been evolved in nature. This implies that either stability of subtilisin BPN is not an important issue in nature, or that stability to peroxide is not significantly correlated with stability toward the enzyme's natural environment. Fully 13% of the 940 clones screened demonstrate more stability to peroxide than wild type, and 1.5% demonstrate significant (greater than 1.5 times) stability as compared to wild type. This large number of positive clones in the local fitness landscape suggests that recombination might be a very successful technique to taking longer walks in productive directions.

The calculation -ln (-ln (residual activity)) is used to generate values proportional to the free energy of activation. This profile is seen in Figure 27, and again demonstrates a long plateau containing the wild type-like stability, despite the high mutation rate. Similar plots were generated for thermostability in a thermostable catalase (Trakulnaleamsai et al., 1995). That data showed that the wild type was one of the most thermostable enzymes of a population (similar to the pNB esterase plots). These data clearly indicate room for improvement in subtilisin BPN' stability to peroxide and are more similar to the catalase activity of the thermostable catalase.

Others have shown that methionine 222 is a residue responsible for peroxide sensitivity (Estell et al., 1985). The mutations at methionine 222 responsible for increased stability to peroxide are cysteine, serine and alanine. The ATG codon responsible for producing methionine at residue 222, when changed by single DNA base changes, will generate only isoleucine, leucine, valine, threonine, lysine and arginine. These mutations are not among those already known to improve subtilisin stability to peroxide. Our data, especially the number and varied stability of the positive clones, suggest that several mutations exist which improve stability to peroxide.

Figure 28 shows the lack of correlation between the activity and stability measurements for subtilisin BPN. Because stability is independent of expression levels,

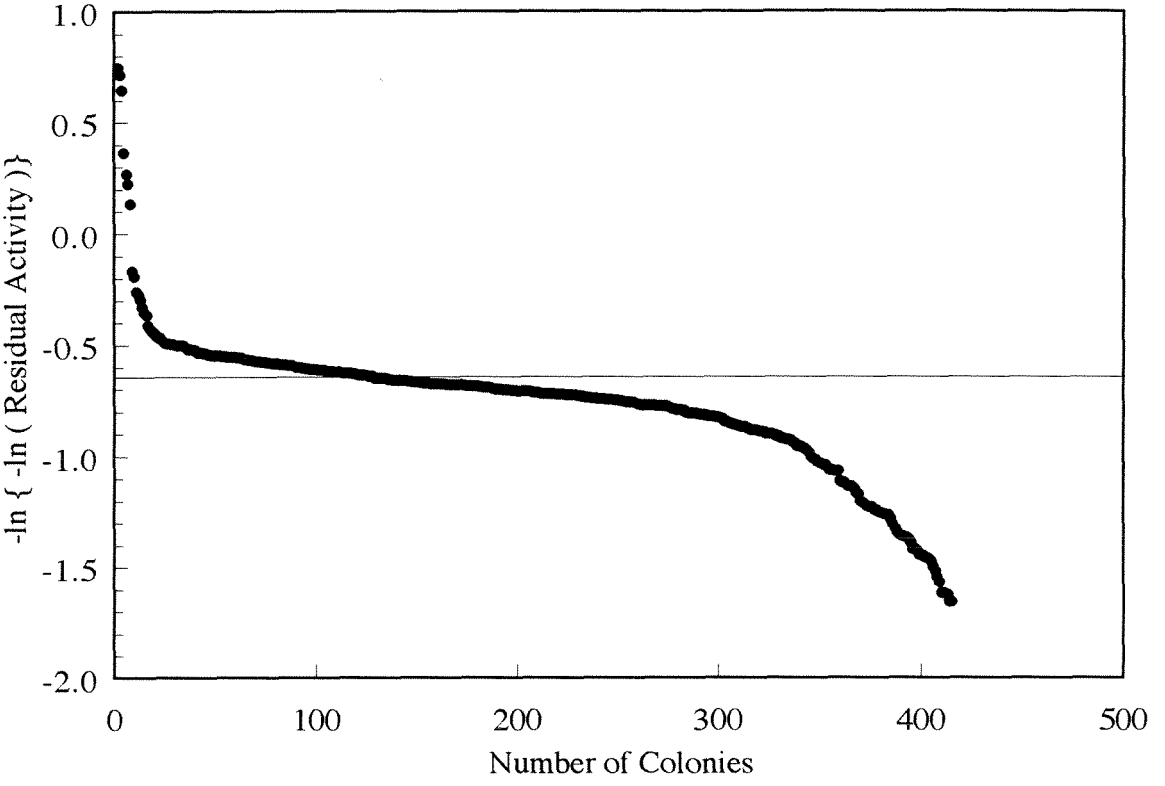



Figure 4.27: Stability profile of subtilisin BPN' variants to 3% hydrogen peroxide and pH 4.5. The 443 clones presented are those containing enough activity to be accurately measured from 940 total clones.

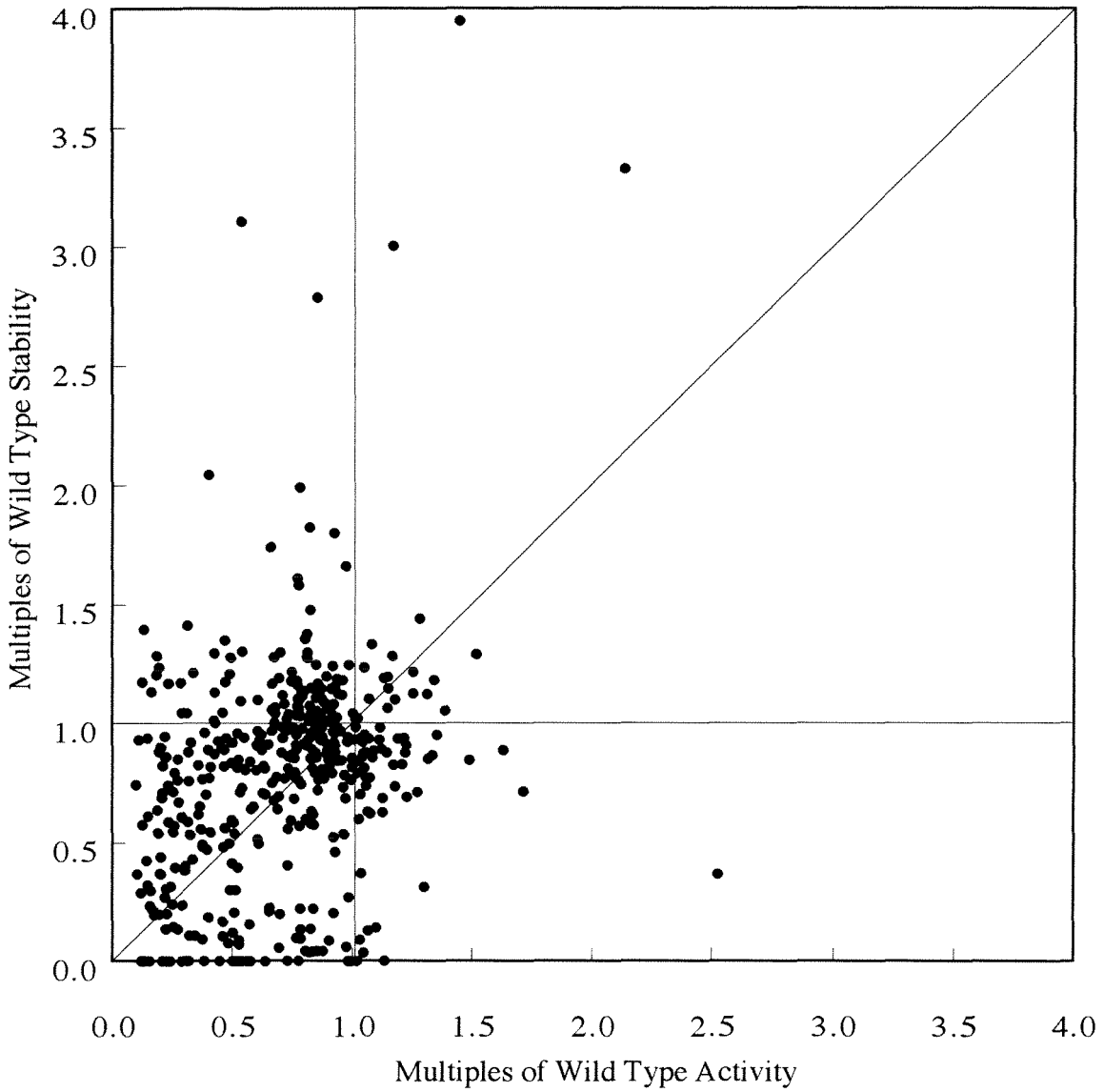



Figure 4.28: Activity versus stability for 443 subtilisin BPN' clones. The horizontal line indicates the trend expected if activity differences were a result of changes in expression levels. The diagonal line indicates the trend expected if a perfect correlation existed between stability and activity.

the horizontal line indicates the stability expected if the variation in activity was expression dependent. The diagonal line indicates the trend expected if activity and stability were highly correlated. From this plot, we can see that the enzymatic properties of stability and activity for subtilisin BPN' are decoupled. This indicates that if the two properties were evolved independently from one another, the independent solutions to increased stability and increased activity could be combined with the expectation of generating a single enzyme both more active and more stable.

## Conclusion

The ability to produce macromolecules meeting design criteria is an important goal of directed evolution. Here we have shown the activity and stability fitness profiles of two enzymes and compared their free energy of activation profiles. The activity profiles of pNB esterase show the enzyme to be improvable on the ester substrates tested, even though the variants improving activity are not frequent. The different locations of wild type on the activity and stability profiles of subtilisin BPN' shows the increased frequency of positives when screening for stability toward hydrogen peroxide at low pH. The progression of population fitness as an enzyme sequence walks from one local landscape to the next was observed through the multiple generations of pNB esterase. This was compared to the population fitness produced by DNA shuffling, which appeared to move a population as a whole to regions of more fit local sequence space.

The measurements of activities on various ester substrates showed the esterase activity to be reasonably well correlated. Enzyme variant activities in DMF environments are highly correlated and suggest that screening for increased activity in DMF can be done at many different DMF concentrations. Activity and stability to peroxide at low pH are not correlated for subtilisin BPN', and this decoupling suggests that the two properties can be evolved independently from each other.

**Materials and Methods**

The mature sequence of subtilisin BPN', as defined by the restriction enzymes BsiWI and BamHI, was chosen for random mutagenesis in an effort to increase this protease's stability to hydrogen peroxide ($H_2O_2$) at low pH. Using conditions which generate a nominal 1% error rate and 3:1 transition:transversion bias (Leung et al., 1989), a DNA library was created, passaged through *E. coli* and screened in *Bacillus*. The screen developed to measure stability in $H_2O_2$ at low pH was as follows. Plasmid DNA from *E. coli* was isolated and used to transform *Bacillus*. This transformation was plated onto LB plates in an overlay of soft agar and allowed to grow for 1 hour. The plates were then overlaid again with LB plus kanamycin (50 $\mu$g/mL) as selective media and incubated overnight at 37 °C. The resulting colonies were individually picked and resuspended in LB plus kanamycin into a unique well of a 96-well plate. These were grown at 37 °C overnight to allow cells to reach stationary phase. These plates were spun at maximum speed in a Beckman tabletop centrifuge. Two 20 $\mu$L aliquots of culture supernatant were removed to new 96-well plates. 180 $\mu$L of a substrate solution consisting of 0.1 M Tris-HCl pH 8.6, 10 mM $CaCl_2$ and 1 mg/mL Suc-Ala-Ala-Pro-Phe-pNA was added to the first plate, and the absorbance read for two minutes at 410 nm. To the second plate 80 $\mu$L of 0.1 M acetate buffer (pH 4.5) containing 3% $H_2O_2$ was added and allowed to incubate at room temperature for 2 hours. 20 $\mu$L of 2.5 M Tris-HCl pH 8.6 was added to raise the pH. 20 $\mu$L was removed, combined with the same 180 $\mu$L of substrate solution used in the first plate, and the absorbance read for five minutes. The ratio of these two activities was used as a measure of stability to $H_2O_2$ and low pH.

## References

Cadwell, C. R. and Joyce, G. F. 1992. Randomization of genes by PCR mutagenesis. PCR Methods and Appl. 2:28-33.

Chen, K. and Arnold, F. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc. Natl. Acad. Sci. USA 90:5618-5622.

Estell, D. A., Graycar, T. P., and Wells, J. A. 1985. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. J. Biol. Chem. 260:6518-6521.

Kauffman, S. A. and Levin, S. 1987. Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Bio. 128:11-45.

Kauffman, S. A. and Macready, W. G. 1995. Search strategies for applied molecular evolution. J. Theor. Bio. 173:427-440.

Leung, D. W., Chen, E., and Goeddel, D. V. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1:11-15.

Macken, C. A., Hagan, P. S., and Perelson, A. S. 1991. Evolutionary walks on rugged landscapes. Siam J. Applied Math. 51:799-827.

Maynard Smith, J. 1970. Natural selection and the concept of a protein space. Nature. 255:563-564.

Moore, J. C. and Arnold, F. H. 1996. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nature Biotech. 14:458-467.

Perelson, A. S. and Macken, C. A. 1995. Protein evolution on partially correlated landscapes. Proc. Natl. Acad. Sci. USA. 92:9657-9661.

Trakulnaleamsai, S., Yomo, T., Yoshikawa, M., Aihara, S., and Urabe, I. 1995. Experimental sketch of landscapes in protein-sequence space. J. Ferm. Bioeng. 79:107-118.