REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE January 1 2005	3. REPORT TYPE AND DATES COVERED Final: May 15 2003 to February 14 2004

4. TITLE AND SUBTITLE
Fuel From Self-Degrading Bioengineered Packaging

5. FUNDING NUMBERS
DAAD19-03-1-0091

6. AUTHOR(S)
Richard A. Gross

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Richard A. Gross, Polytechnic University, CBSE Dept
Six Metrotech Center
Brooklyn, NY 11201

8. PERFORMING ORGANIZATION REPORT NUMBER
RGFuel6

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
45474.1-CH-DRP

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Research focused on developing an enzyme that could be embedded within a single-use plastic and convert the plastic after use to products of value such as fuel to troops in the field. Polylactic acid (PLA) was selected for this work since it is commercially available, degradable by an enzyme (Proteinase K), and is known to have properties compatible with those required by the military for many single-use materials. Engineering of Proteinase K for PLA degradation required a low throughput assay that accurately selects variants that offer improvements in a wide range of enzyme characteristics including the structure of products formed, activity at low pH, and thermal stability. This required a new technological approach to protein engineering. Using an integrated substitution choice method and sequence-activity models we obtained dramatic improvements in enzyme performance while screening only 100 variants. The small number of variants required allowed us to accurately assess the distribution of products formed and enzyme behavior while varying different physical parameters of importance for final process design such as temperature and pH. A new LC-MS method was developed that provided information on degradation product chain length, concentration, and stereochemical composition. Improvements in enzyme performance relative to the wild-type achieved by this new protein engineering method include identification of variants with: i) 12-fold improvement at low pH (5.5), ii) 5-fold improvement in dimer yield, and iii) 2-fold improvement in heat stability at 65°C. This contrasts to other methods such as directed evolution that requires screening of thousands of variants per cycle to obtain improvement factors normally 1.5 to 2-fold.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Waste-to-energy, Fuel, Biodegradable Plastics, Protein Engineering, Proteinase K, Polylactic acid, Self-Degrading Plastics, Single-use plastics, embedded enzymes	13
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Technical Report – Research Agreement No. DAAD19-03-1-0091

Fuel from Self-Degrading Bioengineered Packaging

Submitted by Richard A. Gross
Principle Investigator

Research focused on developing an enzyme that efficiently converts plastics to products of value as fuels or other substances that would otherwise need to be provided to troops in the field. Polylactic acid (PLA) was selected for this work since it is a commercially available biodegradable plastic with useful physical attributes both in fiber and film form. For troops to successfully convert PLA to useful products, an enzyme is needed that can meet the requirements below:

- Completely convert PLA to degradation products within a practical time period (i.e. 24 hours).
- Directly convert PLA to lactic acid for use in fuel cells.
- Increase the enzymes tolerance to low pH environments that will be encountered in non-buffered bioreactors during PLA degradation.
- Increase the activity of the enzyme to function at lower temperatures thus enabling the development of conversion processes that require low energy input.

The above goals call for a protein engineering strategy that can rapidly acquire a broad range of information to address all of the above enzyme requirements. Furthermore, testing of engineered enzymes (variants) for all these activities requires a new approach. This is illustrated by the need to identify variants that produce products with different structures. Such information cannot practically be attained using high-throughput methods. Thus, for the first time a powerful protein engineering technology was used that generates small sets (<50) of designed variants instead of the 1 000 to 10 000 sets required by conventional directed evolution approaches. The protein engineering technology integrates structural, phylogenetic and statistical sequence information to select an initial set of function-modifying substitutions. No one piece of information (such as a crystal structure) is essential, making substitution choice more robust than for any other method. A set of 100 variants was designed, each incorporating 3-6 amino acid substitutions into the wild-type sequence. This greatly increased the information content over library-based methods and eliminated the need for high-throughput screening. We were thus able to evaluate the clones under many experimental conditions using multiple replicates and methods such as size-exclusion-chromatography that are slow, tedious and impossible to convert to a high-throughput format. Multiple new algorithms for determining sequence-activity relationships were designed and used to construct a second set of 15 variants that are currently under study.

The results from the seedling project have been a resounding validation of this approach. The accomplishments from just one cycle of variant sequence and testing are summarized below. If not otherwise specified, activities are expressed relative to the “wild type” enzyme from the fungus *Tritirachium album* (heterologously expressed in *E. coli*). The wild type enzyme is sample "NS9".

Enzyme engineering goals

The goals of the enzyme engineering part of this project are to produce a protease capable of degrading a useful packaging plastic, to produce an enzyme that yields a specific distribution of products when it degrades the plastic, to improve the stability of the enzyme to higher temperatures required for plastic manufacture and to increase the activity of the enzyme at the lower pH values that it will encounter due to the acid generated during plastic degradation.

We set out to accomplish the following goals:

1. Construct a proteinase K that is well expressed and easily extractable in a test production system.
2. Engineer proteinase K to improve its ability to degrade polylactic acid.
3. Engineer proteinase K to improve its tolerance to acidity.
3. Engineer proteinase K to improve its thermostability.
4. Engineer proteinase K to control the PLA oligomers that it produces, to increase the production of those oligomers that have enhanced value as liquid fuels.

1. Proteinase K production system

Escherichia coli is a convenient host for protein engineering: it grows rapidly and its molecular genetics are well understood. The cDNA and the chromosomal gene encoding proteinase K from *Tritirachium album* Limber have previously been cloned in *E coli*. The protein was properly secreted into the periplasm with subsequent processing to the mature proteinase (Gunkel and Gassen, 1989, *Eur. J. Biochem.* **179**:185). As is often the case with genetic constructs described in the scientific literature, it proved impossible for us to obtain the plasmid constructed by Gunkel and Gassen. We also wished to make some improvements to the published system, which we did by re-designing and then synthesizing the entire gene.

MKKLLFAIPLVVPFYSHST*MAPAVEQRSEAAPLIEARGEMVANKYIVKFKEGSALSALDAAMEKISGKPDH*
*VYKNVFSGFAATLDENMVRVLRAHPDVEYIEQDAVVTIN*AAQTNAPWGLARISSTSPGTSTYYYDESAGQGS
CVYVIDTGIEASHPEFEGRAQMVKTYYYSSRDGNGHGTHCAGTVGSRTYGVAKKTQLFGVKVLDDNGSGQYS
TIIAGMDFVASDKNNRNCPKGVVASLSLGGGYSSSVNSAAARLQSSGVMVAVAAGNNNADARNYSPASEPSV
CTVGASDRYDRRSSFSNYGSVLDIFGPGTSILSTWIGGSTRSISGTSMATPHVAGLAAYLMTLGKTTAASAC
RYIADTANKGDLSNIPFGTVNLLAYNNYQAHHHHHH

Figure 1. Amino acid sequence of proteinase K.

The redesigned amino acid sequence of proteinase K is shown in Figure 1. The first 18 amino acids (underlined) are an *E coli* signal peptide derived from the pBAD/gIII vector, followed by two linker amino acids. An *E coli* signal sequence was used to improve secretion of the heterologous protein. The *Tritirachium album* proteinase K propeptide is italicized, followed by the sequence of the mature protein. Finally we added a histidine tag (underlined) to the C-terminus of the protein. This allows convenient and rapid purification of the proteinase K variants, allowing us to assess their

activities in the absence of contaminating proteases endogenous to *E coli* and of other proteins present in the *E coli* periplasm which could act as competitive inhibitors of the proteinase K.

To further improve expression of proteinase K, we redesigned the DNA encoding the protein. Different organisms preferentially use different codons to specify the same amino acid. These differences are particularly marked in E coli, and the presence of codons that are not generally used by E coli often causes significant reduction in expression levels of heterologous proteins. We therefore used codons at the frequencies that they are found in highly expressed E coli genes, and used no codon that was used less than 10% of the time in these genes. The redesigned DNA sequence for proteinase K from the alanine at position 20 to the alanine preceding the histidine tag is shown in Figure 2.

```
GCACCGGCCGTTGAACAGCGTTCTGAAGCAGCTCCTCTGATTGAGGCACGTGGTGAAATGGTAGCAAACAA
GTACATCGTGAAGTTCAAGGAGGGTTCTGCTCTGTCTGCTCTGGATGCTGCTATGGAAAAGATCTCTGGCAA
GCCTGATCACGTCTATAAGAACGTGTTCAGCGGTTTCGCAGCAACTCTGGACGAGAACATGGTCCGTGTACT
GCGTGCTCATCCAGACGTTGAATACATCGAACAGGACGCTGTGGTTACTATCAACGCGGCACAGACTAACGC
ACCTTGGGGTCTGGCACGTATTTCTTCTACTTCCCCGGGTACGTCTACTTACTACTACGACGAGTCTGCCGG
TCAAGGTTCTTGCGTTTACGTGATCGATACGGGCATCGAGGCTTCTCATCCTGAGTTTGAAGGCCGTGCACA
AATGGTGAAGACCTACTACTACTCTTCCCGTGACGGTAATGGTCACGGTACTCATTGCGCAGGTACTGTTGG
TAGCCGTACCTACGGTGTTGCTAAGAAAACGCAACTGTTCGGCGTTAAAGTGCTGGACGACAACGGTTCTGG
TCAGTACTCCACCATTATCGCGGGTATGGATTTCGTAGCGAGCGATAAAAACAACCGCAACTGCCCGAAAGG
TGTTGTGGCTTCTCTGTCTCTGGGTGGTGGTTACTCCTCTTCTGTTAACAGCGCAGCTGCACGTCTGCAATC
TTCCGGTGTCATGGTCGCAGTAGCAGCTGGTAACAATAACGCTGATGCACGCAACTACTCTCCTGCTAGCGA
GCCTTCTGTTTGCACCGTGGGTGCATCTGATCGTTATGATCGTCGTAGCTCCTTCAGCAACTATGGTTCCGT
CCTGGATATCTTCGGCCCTGGTACTTCTATCCTGTCTACCTGGATTGGCGGTAGCACTCGTTCCATTTCCGG
TACGAGCATGGCTACTCCACATGTTGCTGGTCTGGCAGCATACCTGATGACCCTGGGTAAGACCACTGCTGC
ATCCGCTTGTCGTTACATCGCGGATACTGCGAACAAAGGCGATCTGTCTAACATCCCGTTCGGCACCGTTAA
TCTGCTGGCATACAACAACTATCAGGCT
```

Figure 2. DNA sequence encoding proteinase K optimized for expression in E coli.

After constructing the gene as shown above, we placed it under control of the arabinose promoter. *E coli* transformed with the resulting construct produced mature active proteinase K secreted into the periplasm. This proteinase K can then be purified using commercially available nickel-NTA affinity matrices. Thus we have successfully constructed a production system suitable for the analysis of proteinase K variants.

2. Creation of functional proteinase K variants

To engineer into proteinase K the properties that will be required for it to usefully degrade polylactic acid, we are adopting an approach in which small numbers of variants are synthesized and tested; the relationship between sequence and activity is then modeled mathematically and used to design variants that are further improved. This approach was chosen because some of the parameters that are important in this project, particularly the range of oligomers produced by the proteinase, are not readily amenable to high-throughput screening. To obtain meaningful polylactate degradation data requires growing about 1 liter of cells; feasible for tens or even a hundred variants, but certainly not for the thousands that are generally needed in each round for directed evolution methods.

To obtain a sequence activity relationship, it is first necessary to create a set of variants in which each mutation occurs more than once and in combination with different mutations. These combinations of sequences and their associated activities are then amenable to analysis by mathematical tools developed in other fields of engineering.

We used several criteria to select positions within the proteinase K sequence for modification. These included work such as structural studies published in the scientific literature, changes that are likely to be acceptable based on empirically-derived amino acid substitution matrices, phylogenetic analysis of homologs and principal component analysis of the sequences of homologs with known desirable properties, particularly thermostability. The first 24 selected changes are shown in Table 1.

As an initial set of variants we designed and synthesized 24 genes, each of which incorporated 6 of these changes. We also ensured that each change was represented in 6 of the genes. We used casein as a sensitive substrate simply to assess whether or not the variant proteinase K proteins were at all active. From our initial set we obtained 3 variants with some activity, two of which were more active degraders of casein than the original protein.

Variation	Primary reason for inclusion
N95C	Structural stability at higher temperature: from published literature
P97S	P to S for flexibility and structural perturbabtion
S107D	Seen in active homologs
S123A	Found in the thermostable consensus sequence
I132V	Substitution matrix-derived conservative change (controlled perturbation)
E138A	Seen in experiments in the literature
M145F	Seen in experiments to improve thermostability
Y151A	Seen in experiments to improve thermostability
V167I	Substitution matrix-derived conservative change (controlled perturbation)
L180I	Substitution matrix-derived conservative change (controlled perturbation)
Y194S	Varaiation observed in highly active clone
A199S	Substitution matrix-derived conservative change (controlled perturbation)
K208H	PCA modelling of thermostable homologs collected from GenBank.
A236V	PCA modelling of thermostable homologs collected from GenBank.
R237N	From experiments to improve thermostability (in literature)
P265S	P to S for flexibility and structural perturbabtion
V267I	Substitution matrix-derived conservative change (controlled perturbation)
S273T	Multiple sources identify this change. (thermostability and other)
G293A	Thermostability considerations (observed in thermitases)
L299C	Could form a disulphide bridge with N95C (from literature)
I310K	From structural studies
K332R	Thermostability considerations (observed in thermitases)
S337N	Thermostability considerations (observed in thermitases)
P355S	P to S for flexibility and structural perturbabtion

Table 1. The first 24 positions in proteinase K chosen for variation.

NS#	25	34	50	55	95	97	107	123	132	138	145	151	167	180	194	199	208	236	237	265	267	273	293	299	310	332	337	355	362
	H	S	N	S	C	S	D	A	V	A	F	A	I	I	S	S	H	V	N	S	I	T	A	C	K	R	N	S	M
NS1																													
NS9	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS10	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	S	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS12	Y	A	D	N	N	P	S	A	I	E	M	A	V	L	Y	A	K	A	R	P	V	S	A	L	K	R	S	S	L
NS13	Y	A	D	N	N	P	S	S	V	E	F	A	I	L	Y	A	K	A	R	P	V	T	G	L	I	K	N	P	L
NS14	Y	A	D	N	N	P	D	S	I	E	M	A	V	I	S	S	K	A	R	P	I	S	G	L	I	K	S	P	L
NS15	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	N	P	V	T	A	L	I	R	S	S	L
NS16	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS18	Y	A	D	N	N	P	S	S	I	A	M	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS19	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	H	A	R	P	V	S	G	L	I	K	S	P	L
NS20	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	N	P	V	S	G	L	I	K	S	P	M
NS21	Y	A	N	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	N	P	V	S	G	L	I	K	S	P	L
NS22	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	S	V	S	G	L	I	K	S	P	L
NS23	Y	A	D	N	N	P	S	S	I	E	M	A	I	I	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS24	H	A	D	N	N	P	S	S	I	E	M	A	I	I	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS25	Y	A	D	S	N	P	S	S	I	E	M	A	I	I	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS26	Y	S	D	N	N	P	S	S	I	E	M	Y	V	L	S	S	K	A	R	P	I	S	G	L	I	K	S	P	L
NS27	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	T	A	L	K	K	S	P	L
NS28	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	R	N	S	L
NS29	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	R	N	S	L
NS30	Y	A	D	N	N	P	D	S	I	E	M	A	V	L	S	A	K	A	R	P	V	T	G	L	I	R	S	P	L
NS31	Y	A	D	N	N	P	D	S	I	E	M	A	V	L	S	A	K	A	R	P	V	T	G	L	I	R	S	P	L
NS32	Y	A	D	N	N	P	S	A	I	E	M	Y	I	L	Y	S	K	A	R	P	V	S	A	L	I	K	N	P	L
NS33	Y	A	D	N	N	P	S	A	I	E	M	Y	I	L	Y	S	K	A	R	P	V	S	A	L	I	K	N	P	L
NS34	Y	A	D	N	N	P	D	S	I	E	M	Y	I	L	Y	A	K	A	R	P	I	T	G	L	I	K	N	P	L
NS35	Y	A	D	N	N	P	D	S	I	E	M	Y	I	L	Y	A	K	A	R	P	I	T	G	L	I	K	N	P	L
NS36	Y	A	D	N	N	P	S	A	I	E	M	Y	V	I	S	A	K	A	R	P	V	S	A	L	I	K	S	S	L
NS37	Y	A	D	N	N	P	S	A	I	E	M	Y	V	I	S	A	K	A	R	P	V	S	A	L	I	K	S	S	L
NS38	Y	A	D	N	N	P	S	S	I	E	F	Y	I	L	S	A	K	A	R	P	V	S	G	L	I	K	S	P	L
NS39	Y	A	D	N	N	P	S	S	I	E	M	Y	V	I	Y	S	K	A	R	P	V	T	G	L	I	K	S	P	L
NS40	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	I	S	A	L	I	R	S	P	L
NS41	Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	I	S	A	L	I	R	S	P	L
NS42	Y	A	D	N	N	P	D	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	S	G	L	K	K	N	P	L
NS43	Y	A	D	N	N	P	S	A	I	E	M	A	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	S	S	L
NS44	Y	A	D	N	N	P	S	A	I	E	M	A	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	S	S	L

Table 2. Sequence variations in active proteinase K variants

Using this information we designed and synthesized an additional 24 genes, 8 of which were designed to test the amino acids that we had not seen represented in any of the active variants, the remaining 16 were designed to test different combinations of amino acid changes that appeared to allow some proteinase activity. The combinations of amino acid changes that resulted in enzymes with some activity in the degradation of casein are shown in Table 2. Combinations of amino acids that yielded proteinase K variants with no apparent protease activity are shown in Table 3.

25	34	50	55	95	97	107	123	132	138	145	151	167	180	194	199	208	236	237	265	267	273	293	299	310	332	337	355	362
Y	A	D	N	C	P	S	S	I	E	F	Y	I	L	Y	S	K	A	N	P	V	T	G	L	I	K	S	P	L
Y	A	D	N	N	S	S	S	I	A	M	Y	V	I	S	A	K	V	R	P	I	S	G	L	I	K	S	P	L
Y	A	D	N	N	P	D	S	V	E	M	Y	V	L	Y	A	H	A	R	S	V	S	G	C	I	K	N	P	L
Y	A	D	N	N	S	D	S	I	E	M	Y	V	I	Y	A	K	V	N	P	V	S	G	L	K	K	S	P	L
Y	A	D	N	N	P	S	A	I	A	M	Y	V	L	Y	S	H	A	R	S	V	S	G	L	I	K	S	S	L
Y	A	D	N	C	P	S	S	I	E	M	Y	V	L	S	A	K	A	R	P	I	S	A	C	I	R	S	P	L
Y	A	D	N	C	P	S	S	V	A	F	Y	I	L	Y	A	H	A	R	P	V	S	G	L	I	K	S	P	L
Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	V	N	P	V	T	A	L	I	R	S	S	L
Y	A	D	N	N	S	S	A	I	E	M	Y	V	L	Y	A	K	A	R	S	V	S	G	C	K	K	N	P	L
Y	A	D	N	C	P	D	A	I	E	M	Y	V	I	S	A	K	A	R	P	V	S	G	L	I	K	N	P	L
Y	A	D	N	N	P	S	S	I	A	M	A	I	L	Y	S	H	A	R	P	V	S	G	C	I	K	S	P	L
Y	A	D	N	N	S	S	S	I	E	F	Y	V	L	Y	A	K	A	N	P	V	T	A	L	K	K	S	P	L
Y	A	D	N	N	P	S	S	V	E	M	Y	V	L	Y	A	K	V	R	S	I	S	G	L	I	R	S	S	L
Y	A	D	N	N	S	S	S	I	E	M	A	V	L	Y	S	K	V	R	P	V	S	G	C	I	K	S	S	L
Y	A	D	N	C	P	D	S	I	E	M	Y	I	I	Y	A	K	A	R	P	V	S	A	L	K	K	S	P	L
Y	A	D	N	N	P	S	S	I	E	F	Y	V	L	Y	A	K	A	N	S	I	S	G	L	I	R	N	P	L
Y	A	D	N	N	P	S	A	V	A	M	Y	V	L	S	A	H	A	R	P	V	T	G	L	I	K	S	P	L
Y	A	D	N	N	P	S	A	I	E	M	Y	V	L	Y	A	H	V	R	P	I	S	A	C	I	K	S	P	L
Y	A	D	N	N	P	D	S	V	A	F	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	N	S	L
Y	A	D	N	N	S	S	S	I	E	M	Y	V	I	S	S	K	A	R	S	V	S	G	L	K	K	S	P	L
Y	A	D	N	C	P	S	S	I	E	M	A	I	L	Y	A	K	A	N	P	V	T	G	L	I	R	S	P	L
Y	A	D	N	C	P	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
Y	A	D	N	N	S	S	S	I	E	M	Y	V	L	Y	A	K	A	R	P	V	S	G	C	I	K	S	P	L
Y	A	D	N	N	S	S	S	I	E	M	Y	V	L	S	A	K	A	R	P	V	S	G	C	I	K	S	P	L
Y	A	D	N	N	P	S	S	I	E	M	Y	V	L	Y	A	K	V	R	P	V	S	G	L	I	K	S	P	L
Y	A	D	N	N	P	D	A	I	E	F	Y	V	L	Y	A	K	A	R	P	V	S	G	L	I	K	S	P	L
Y	A	D	N	N	P	S	S	I	E	F	Y	V	I	Y	A	K	A	R	P	I	S	G	L	K	K	S	S	L
Y	A	D	N	N	P	S	S	I	E	F	A	V	L	Y	S	K	A	R	P	V	S	G	L	K	R	S	P	L

Table 3. Sequence variations in inactive proteinase K variants

3. Use of preliminary screens

The aim of this project is to optimize proteinase K for the degradation of polylactic acid. All of the variants shown in Table 2 have now been transferred to the Gross lab for analysis of their polylactic acid degrading properties. Because some of the properties of interest may potentially be measured independently we have used preliminary screens to test the sequence-function modeling approach on which the project is based.

We have chosen two preliminary screens. The first of these is the ability of E coli strains carrying plasmids encoding proteinase K variants to hydrolyze casein. Casein is a protein in skimmed milk whose incorporation into the media on which E coli grows causes a cloudy appearance to the plate. Hydrolysis of casein results in a clear zone surrounding the colony. We use this screen as a crude test for proteinase activity.

The second preliminary screen that we have used is the cleavage of *p*-nitro anilide peptides, which results in an increase in absorbance at 405 nm. These substrates are not intended as a surrogate for activity of proteinase K towards polylactic acid. Indeed we see clear differences in the activity of some of our variants towards casein (a large and heterogeneous substrate) and *p*-nitro anilide tripeptides. Some variants prefer casein and others prefer the small peptides. This simply emphasizes the importance of screening variants using the substrate of interest, in this case polylactic acid. However, we are not only interested in the substrate preferences of the proteinase K, we are also interested in modifying its physical properties. Specifically we would like to increase the tolerance of proteinase K to heat treatment, in order to allow its incorporation into the polylactic acid during manufacture. We would also like to increase the activity of proteinase K at low pH, so that it can function in the presence of the lactic acid that will be produced during its degradation of polylactic acid.

4. Increase of acid tolerance of proteinase K

We tested all of the variants shown in Table 2 for their activity towards a p-nitro anilide substrate at three different pHs (4.5, 5.5 and 7.5). Our aim was to see whether the

variations that we had introduced were suitable for modifying the physical properties of the enzyme.

Figure 3 shows that four of the variants synthesized have 3-4 times the activity of the wild type enzyme at pH4.5, compared with less than 2-fold improvement over wild type at pH7.5. This shows that we have identified and combined some residues that shift the pH profile of proteinase K. We may therefore anticipate that we will be able to increase the activity of proteinase K for the degradation of polylactic acid, specifically for improving its tolerance to the low pH resulting from lactic acid formation.

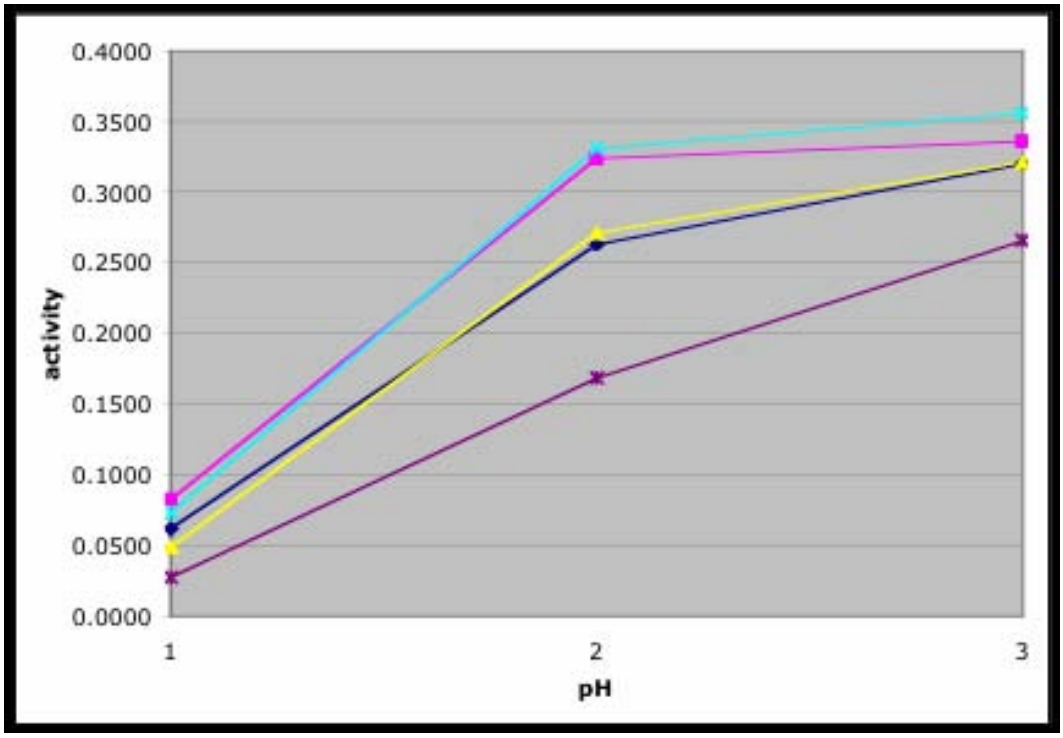

Figure 3. Activity of four variants and wild type proteinase K towards AAPL-p-nitro anilide at three different pHs. The same amount of purified proteainse K enzyme (approximately 20 ng) was incubated with AAPL-*p*-nitro anilide at pH 4.5 (1), pH 5.5 (2) and pH 7.5 (3). Activity was measured as an increase in the absorbance at 405 nm. Four of the variants that possessed increased activity at pH 7.5 (in descending order of activity: NS40 in turquoise, NS23 in pink, NS15 in blue and NS35 in yellow compared with wild type in purple) also had broadened pH activity profiles.

5. Increase of thermostability of proteinase K

We tested all of the variants shown in Table 2 for their activity towards a p-nitro anilide substrate following a 5 minute treatment of the purified enzyme at 65°C. Our aim was to see whether the variations that we had introduced were suitable for modifying the thermal tolerance of the enzyme, a necessary property for the proteinase to withstand the manufacturing process.

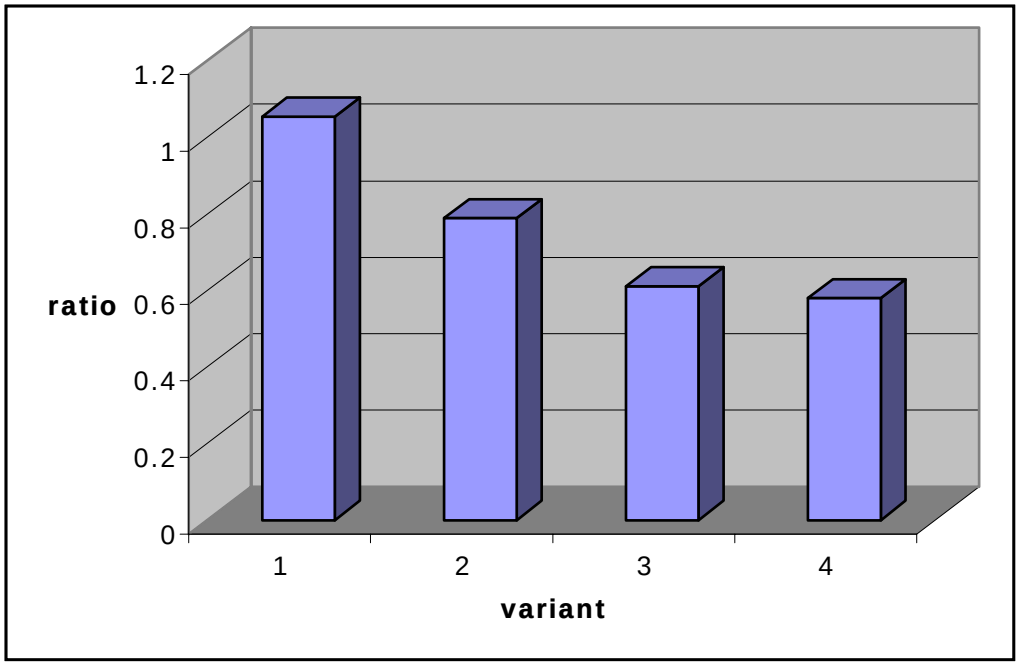

Figure 4. Activity of four variants and wild type proteinase K towards AAPL-p-nitro anilide following heat treatment. Duplicate samples of purified proteinase K variants were either incubated on ice for 5 minutes or incubated at 65°C for five minutes. The enzymes were then incubated with with AAPL-*p*-nitro anilide at pH 7.5. Activity was measured by an increase in absorbance at 405 nm. The activity of the variant following heat treatment was divided by its activity without heat treatment to calculate the stability ratio, shown on the y-axis. Variants were NS40 (1), NS32 (2), NS19 (3) and wild type (4).

Figure 4 shows that three of the variants synthesized are more stable to heat treatment than the wild type enzyme. This shows that we have identified and combined some residues that improve the thermal tolerance of proteinase K. We may therefore anticipate that we will be able to increase the activity of proteinase K following its incorporation into polylactate during manufacture.

Performance of variants at the optimal pH of Proteinase K.

Proteinase K has an optimal pH of 8.6 under native conditions. To identify key amino acids that improve the performance of this enzyme under optimal conditions, Poly lactic acid was hydrolysed with wild type and mutant enzymes at 37 C for 48 hours. At the end

of the time period the mixture was filtered and the pH was recorded. The filtrate was then analysed by LC-MS using a water acetonitrile mobile phase. The product profile was evaluated using lithium lactate as a monomer standard, lactic acid lactide as a dimer standard and commercial lactic acid (mixture of monomer, dimer and trimer) as a trimer standard. The yield of lactic acid monomer, dimer (LL and LD) and trimer was determined in mmoles and the background activity was subtracted using a buffer control. In order to observe the improvement of various mutations, the yields were expressed relative to that of NS9 (the wild type enzyme).

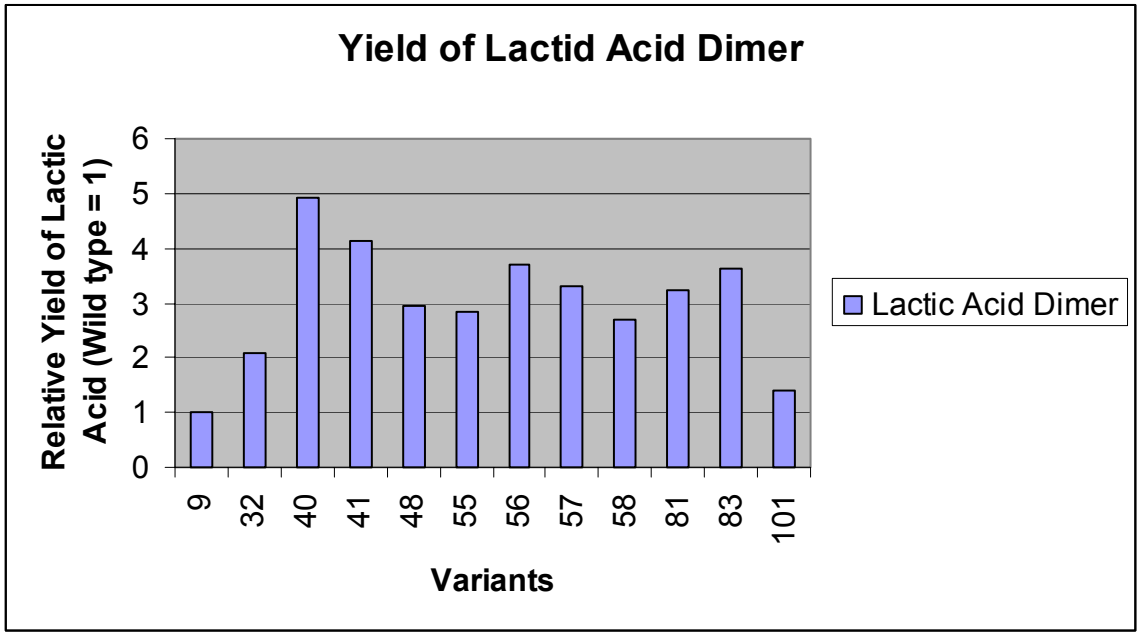

Figure 5. Relative yield of Lactic acid dimer (with reference to the yield of the wild type enzyme, NS9).

The analysis of the hydrolysis products at pH 8.6 indicated that a significant improvement had been achieved with variant NS 40 having a near 5-fold improvement in activity (See Figure 5). While a number of variants had lower yields than the wild type enzyme, a large group had better yields than the wild type enzyme. This clearly indicated that the enzyme improvement strategy was useful in developing better candidates.

Altered activity at lower pH conditions.
As the optimal pH of Proteinase K is 8.6, this enzyme is quickly inhibited at lower pH values. We set out to screen the variants at pH 6.5 and 5.5 to identify those that would not be inhibited by the accumulation of an acidic product such as lactic acid or its oligomers.

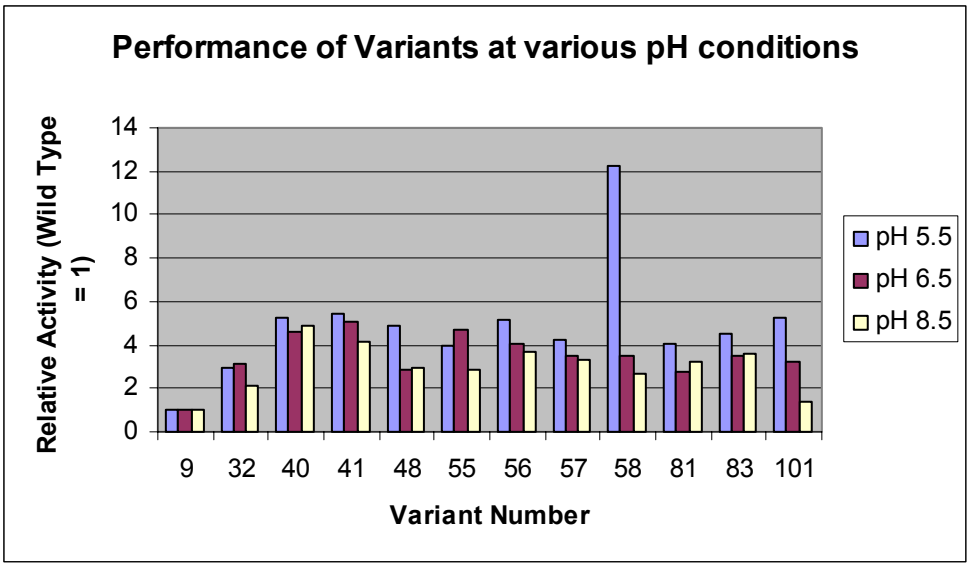

Figure 6. Relative yield of Lactic acid dimer (with reference to the yield of the wild type enzyme, NS9) at pH 5.5, 6.5 and 8.6.

The yield of lactic acid dimer (LL and LD) was determined in mmoles and the background activity was subtracted using a buffer control. In order to observe the improvement of various mutations, the yields were expressed relative to that of NS9 (the wild type enzyme). As there was no monomer detected at pH 5.5, the performance of the variants was evaluated by the relative yields of the lactic acid dimer (LL and LD). A number of variants were found to perform very well at pH 5.5. The highlight of this screen is variant NS 58 which has a 12 fold improvement over the wild type and out performs even commercial preparations of Proteinase K purchased from Sigma chemical Co St Louis.

Thermal stability

Industrial and field applications of biocatalysists call for a very stable enzyme. To screen our variants for heat-stability, we heat-denatured the enzyme preparations at 60 ^{o}C for 15 minutes prior to hydrolysis of PLA at 37 ^{o}C and pH 8.6. All the variants were screened under these conditions to identify those that would be thermostable.

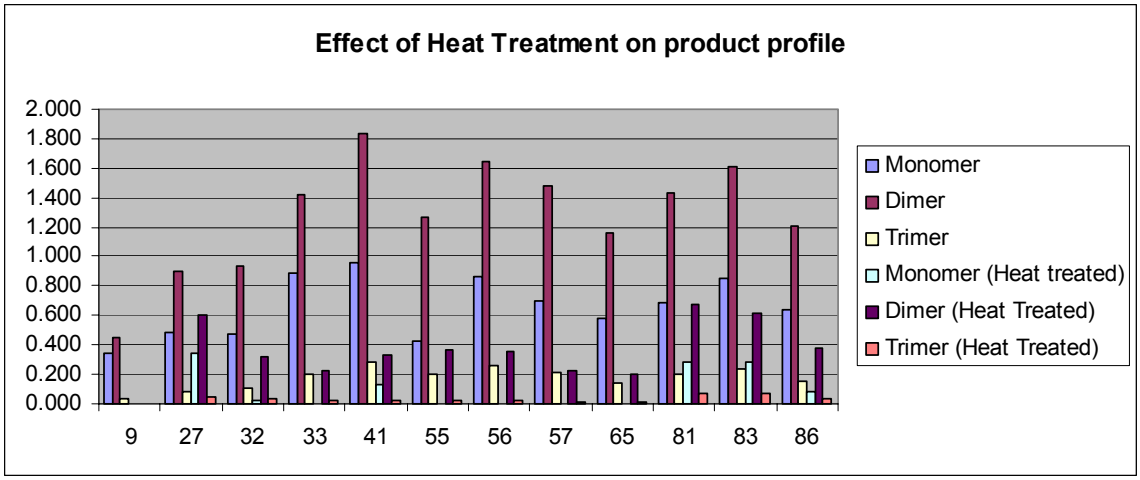

Figure 7. Yield of Lactic acid Oligomers (monomer, dimer and trimer) in mmoles at pH 8.6 after a 15 minute heat denaturation at 60 ^{o}C.

While most enzymes succumbed to heat denaturation and lost all their activity as in the case of the wild type enzyme (see Figure 7), certain variants remained active after heat-denaturation. The most striking variant in this study was variant NS 27 which retained 67 % of its activity after heat-denaturation (see Figure 8).

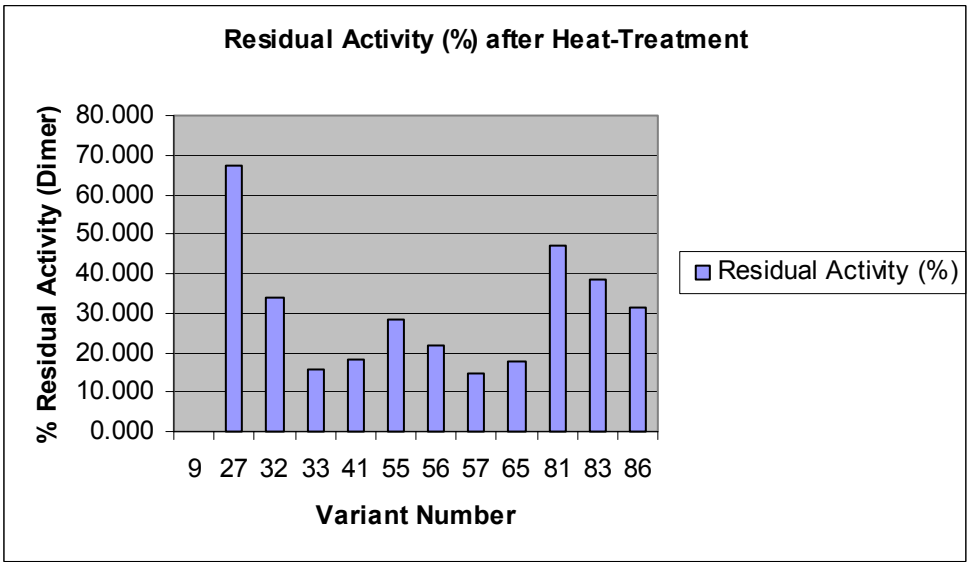

Figure 8. Residual activity after heat denaturation: Ratio of the yield of dimer (LL and LD) after heat denaturation to that of the untreated enzyme.

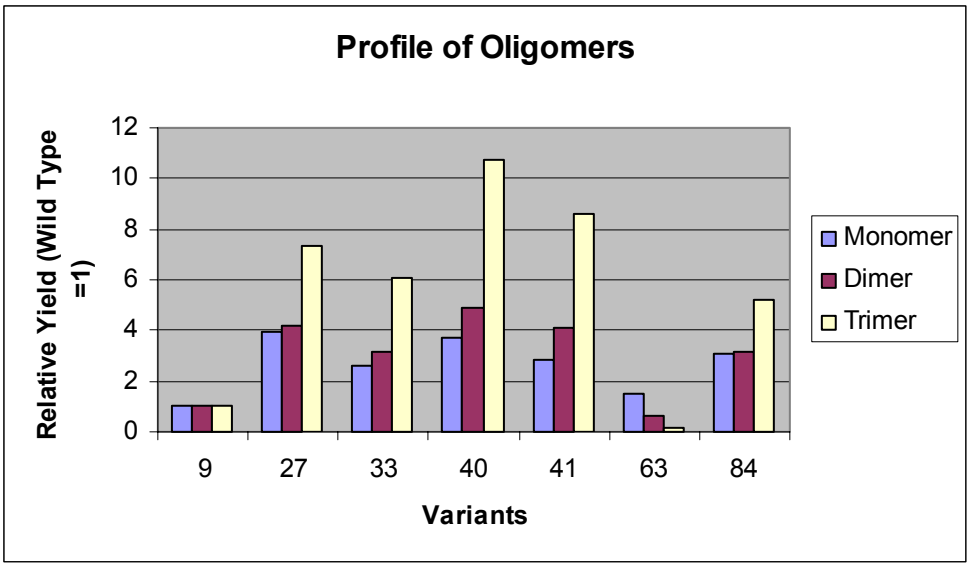

Figure 9. Profile of lactic acid oligomeric products (monomer, dimer and trimer) relative to the wild type enzyme (NS 9).

A key feat in protein engineering is the ability to alter the product profile of an enzyme. Such an approach can yield novel products and can be used to tailor the reaction towards specific products. We have analyzed the products of all our variants at pH 8.6 and looked for variants which differed significantly from the wild type enzyme. While a majority of the variants had increased yields of the lactic acid trimer (see Figure 9) such as NS 27, 33 40, 41 and 84, NS 63 though lower in total activity had and increased yield of lactic acid monomer. This result indicated a significant change in product profile from the general trend.

Summary

Analysis of variant structures resulting in the above improvements showed that amino acid substitutions important for one function are different from those that are important to improve another. This emphasizes the importance of using low-throughput, high information-content screens. It is well known that high-throughput surrogate screens increase the noise and possibility of false-positives. This leads to lost time as information is wrongly interpreted and used in the design of next round variant pools. Equally dangerous is the omission of variants that are not recognizing for the beneficial changes they provide. These difficulties are largely overcome by this new protein engineering strategy for which small variant sets become the focus of careful and comprehensive studies to characterize variant activities. Furthermore, 8 different algorithms for sequence activity modeling have already been tested for sequence activity modeling and are generally proving to be robust and capable of predicting optimal sequences as well as designing information-rich variants.

Conclusion

Directed evolution approaches must screen thousands of variants per cycle to obtain improvement factors that normally are 1.5- to 2-fold. Such large numbers of variants makes it impractical to screen multiple enzyme performance characteristics. Furthermore,

the results from high through-put screening are often are tainted by false positives and negatives. The project undertaken to engineer PLA degrading enzymes needed low throughput assays that could accurately select variants that offer improvements in a wide range of enzyme characteristics including the structure of products formed. This required a new technological approach to protein engineering. Using an integrated substitution choice method, and sequence-activity models, we have obtained a dramatic shift in enzyme activity, while screening only a total of 100 variants. The small number of variants required allowed us to accurately assess the distribution of products formed and enzyme behavior while varying different physical parameters that are important for final process design such as temperature and pH.