Title
Investigation of the substrate specificity of trypsin

Author
Evnin, Luke Basil

Publication Date
1990

Peer reviewed|Thesis/dissertation

Investigation of the Substrate Specificity of Trypsin

by

Luke Basil Evnin

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

Date SEP 9 19[illegible] University Librarian

Degree Conferred: . . . . . .

To Deann K. Wright who is, after all of this, still willing to be my wife.

Special thanks must be extended to several people who have made this work possible. Charly Craik has helped me develop better experimental systems and achieve a more sophisticated understanding of proteases. In June of 1985 when I began, the research was conducted in a single small room and Charly's office was a converted closet. Although Charly's office is still a closet, the lab is substantially larger, and our research has attracted outstanding post-docs and graduate students to continue these efforts. Robert Fletterick has been a source of inspiration and encouragement throughout the peaks and deep valleys that characterize experimental science. The style and thoughtfulness of his work has served as a model. The sage advice of Bob Stroud has helped me keep my thinking on track on several occasions.

My most intimate and fruitful collaboration has been with John Vásquez. Our numerous hours of discussion that date from the beginning of our experience are the substance of this dissertation. In this vein, I owe thanks to Jeff Higaki who has also been a member of the lab since the beginning. My experiments have strived to emulate the care and consistency of Jeff's work. Lily Babé and Dianne Decamp have brightened the lab and have helped me maintain my composure on many difficult days. Lily performed the silver staining of *S. griseus* trypsin. Christopher Cilley has been a friend and a collaborator (when I could steal his time) and is responsible for the purification of the trypsin D189C. I am indebted to Sergio Pichuantes for all the work involving yeast expression. Jason Rosé deserves credit for working with me. I would like to wish the best of luck to

John Perona who intends to carry this work to the next peak. Good luck to Scott Willet and David Cory who will also continue the trypsin work after the departure of the three initial lab members. Good luck also to Ann Eakin in her quest to stay at the cutting edge of molecular biology technology. Guo Hong who's name I still do not pronounce correctly and Dan Wanders are former lab members that I would also like to thank.

Investigation of the Substrate Specificity of Trypsin

Luke Basil Evnin

Trypsin hydrolyzes peptide and ester bonds carboxy terminal to arginine and lysine amino acids. The exclusive specificity exhibited toward these basic amino acids is derived from the strength of the enzyme-substrate interactions in both the ground state and the transition state. From X-ray crystallographic analysis and enzymological studies, aspartic acid 189 was identified as a major determinant of this specificity. In this dissertation, I experimentally address this conclusion and analyze the contribution of other components in the primary substrate binding pocket. Furthermore, I examine how trypsin minimizes any preference for the structurally disparate amino acids, arginine and lysine. Efficient bacterial expression of rat anionic trypsin has been attained using the *tac* promoter and the *trpA* terminator to control expression and the *hisJ* signal peptide to direct secretion to the periplasm. Using oligonucleotide-directed mutagenesis, two libraries of mutants were generated. To facilitate sorting through the libraries, screens based on both amidolysis and esterolysis were developed and a genetic selection based on proteolysis was devised. The activity of trypsin was shown to be critically dependent on the presence of a negative charge at position 189 by assaying the activity of 18 trypsins differing only at position 189 (the first library). This role is best satisfied by aspartate although cysteine and glutamate mutants are active, retaining 1.4% and 0.01% the activity of trypsin, respectively, at pH 8.0. From the second set of mutant trypsins, those substituted

at position 189 and position 190, 15 active mutants were identified. Either aspartate or glutamate in either position 189 or 190 could satisfy the requirement for a negative charge at the base of the pocket although aspartate at position 189 was optimal. With the favored aspartate at position 189, several amino acids, C, T, G, V, and P, could substitute for serine at position 190. However, the differential specificity toward arginine and lysine substrates was shown to depend on the amino acid at position 190, and this regulatory effect on the differential substrate specificity was identified in the substrate binding pockets of naturally occurring trypsin homologs.

Chair and Advisor- Dr. Charles S. Craik, Assistant Professor in the Departments of Pharmaceutical Chemistry and Biochemistry and Biophysics, University of California-San Francisco

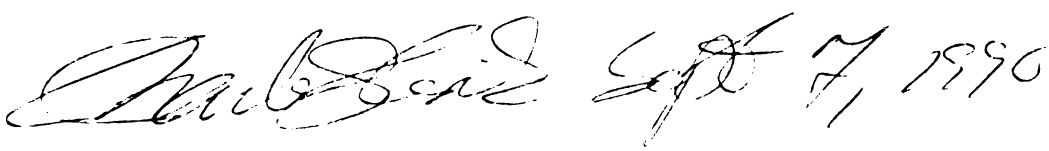


Table of Contents **Page No.**

List of Tables

Introduction

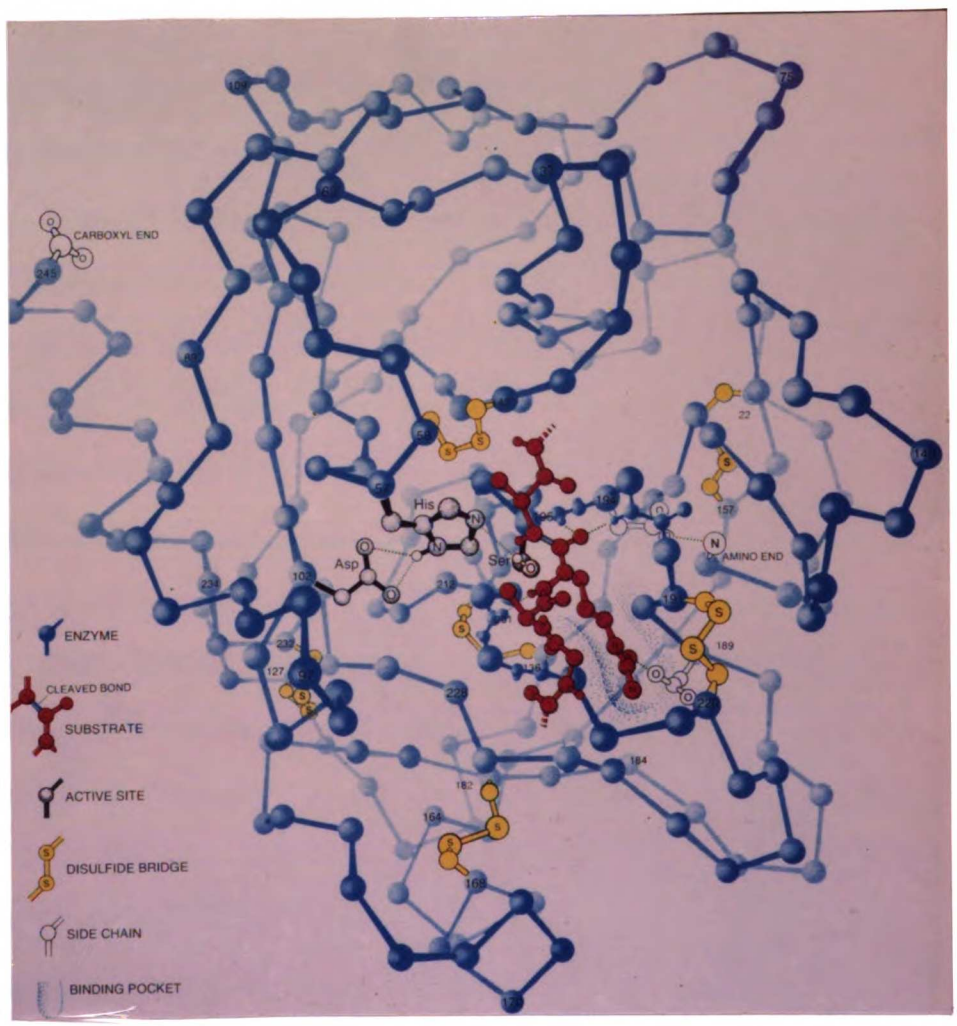
Towards the Study of Substrate Specificity in Trypsin

Catalysis is achieved by lowering the free energy barrier between the reactants and products of a reaction. Enzymes achieve this goal by binding specifically to the appropriate substrate to form a complex and by stabilizing the transition state between the reactants and products (Fersht, 1985). In general terms, the enzyme acts as a polar solvent, superior to water, for the cognate substrate and not other substrates (Warshel, Åqvist et al. 1989). A complex mix of forces is involved in this solvation including hydrogen bonds, salt bridges, and the hydrophobic effect. The description of the amino acids involved in these interactions and the magnitude of their contribution is the subject of this dissertation. These questions have been addressed using a single enzyme, trypsin, that is a superb catalyst and has a selective substrate specificity.

The serine proteases are a well studied family of enzymes which are defined by the arrangement of three side chains, aspartic acid, histidine, and serine (the catalytic triad), that activate the serine hydroxyl oxygen. This triad has evolved in the trypsin and the subtilisin class of enzymes which are otherwise unrelated to each other. Enzymes from both families have been investigated intensively; I have focused exclusively on an enzyme in the former class. Structural characterization of trypsin (Figure i-1) (Kreiger, Kay et al. 1974; Ruhlmann, Kukla et al. 1973; Marquart, Walter et al. 1983), chymotrypsin (Matthews et al. 1967), elastase (Sawyer, Shotton et al. 1978), thrombin (Bode, Mayr et al. 1989), kallikrein (Chen and Bode 1983), the bacterial homologs

Figure i-1. The Three Dimensional Structure of Trypsin.

A model of the solution of the three dimensional structure of bovine cationic trypsin complexed to a peptide substrate. The model is based on the solution of the three dimensional structure of trypsin with benzamidine (Kreiger, Kay et al. 1974) and the peptide substrate was accurately modeled from consideration of the requisite geometry and data including the structure of the BPTI-trypsin complex (Ruhlmann, Kukla et al. 1973). The numbering of the mature form begins at 16 and includes some insertions and deletions relative to chymotrypsin which is the numbering standard. Consequently, the terminal amino acid is position 245. Note the six disulfide bonds in yellow and the important aspartic acid residues at positions 189 and 194 in white. The side chains that form the catalytic triad are shaded. The key is shown at the lower left of the Figure. This illustration was created by Irving Geis for the use of Dr. Robert Stroud (Stroud 1974). Permission to reproduce this illustration was obtained from both the artist and Dr. Stroud.


CARBOXYL END
245
His
Asp
Ser
AMINO END
ENZYME
CLEAVED BOND
SUBSTRATE
ACTIVE SITE
DISULFIDE BRIDGE
SIDE CHAIN
BINDING POCKET

Streptomyces griseus protease A (Brayer, Delbaere et al. 1978; James, Sielecki et al. 1980) and *Streptomyces griseus* protease B (Delbaere, Hutchinson et al. 1975), the α-lytic protease (Brayer, Delbaere et al. 1978; Bone, Silen et al. 1989), tonin (Fujinaga and James 1987) and *Streptomyces griseus* trypsin (Read and James 1988) have been completed and at least some functional characterization has been completed for trypsin (for review see Fersht 1985), chymotrypsin (for review, see Fersht 1985), elastase (see Creighton 1983, pg 433), kallikrein (Ascenzi, Menegatti et al. 1982; Fiedler 1987), *S. griseus* proteases A (Bauer, Thompson et al. 1976a; Bauer, Thompson et al. 1976b) and protease B (Bauer 1978)[1], α-lytic protease (Bauer, Brayer et al. 1981; Bone, Silen et al. 1989), and thrombin (Baird, Curragh et al. 1965; Weinstein and Doolittle 1972; Ascenzi, Menegatti et al. 1982). In contrast, some functional characterization of urokinase (Ascenzi, Menegatti et al. 1982), and plasmin (Weinstein and Doolittle 1972) has been completed but no X-ray crystallographic solution of these structures is available. Various other proteases in this family have been identified by sequence homology but have yet to be studied in detail (Doolittle and Feng 1987).

Trypsin is one of three principle digestive proteases (pepsin and chymotrypsin) that functions in the small intestine of all mammals (Neurath, Walsh et al. 1967). Trypsin is also found in invertebrates such as the crayfish (Titani, Sasagawa et al. 1983) and even in bacteria (Delbaere, Hutchinson et al. 1975). It is

---

[1]*Streptomyces griseus* A (SPGA) is the same enzyme as *Streptomyces griseus* 3 (SPG3). SGPB is identical to SPG1 (Bauer et al, 1976a).

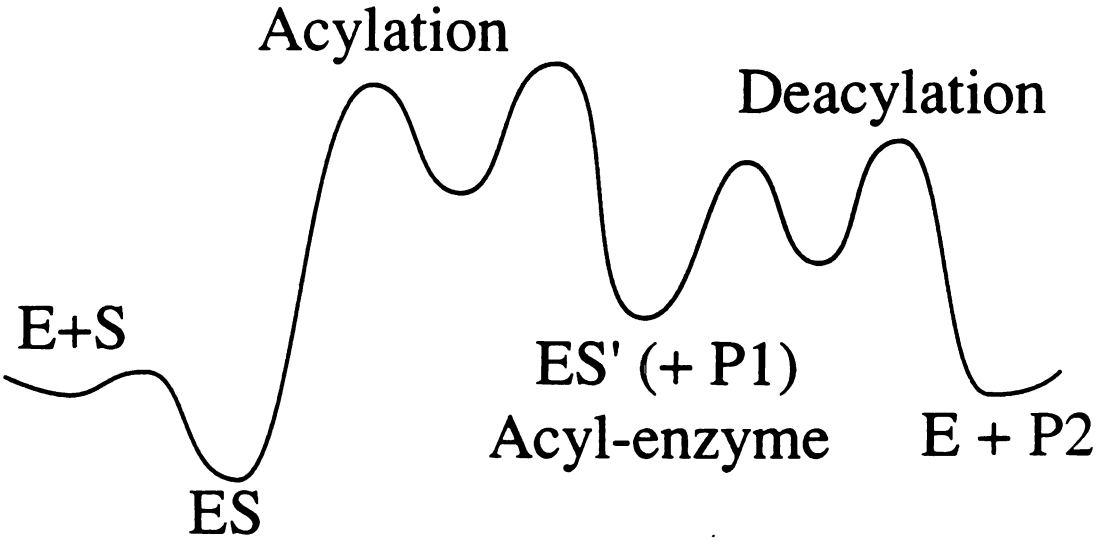
synthesized as a zymogen and, in mammals, is excreted to the small intestine from the pancreas as the pro enzyme. Activation of trypsin occurs by cleavage of a short amino terminal peptide which exposes an isoleucine (position 16) at the new amino terminus. The amino terminal peptide is of varying length but concludes with $(Asp)_4$-Lys (Neurath, Walsh et al. 1967) which is recognized by the physiological activator enterokinase (and by trypsin itself). In most organisms there are a number of trypsin isozymes including one cationic and one anionic form (Neurath, Walsh et al. 1967; Fletcher, Alhadeff et al. 1987). These forms differ substantially in amino acid composition but function equivalently. The bovine cationic form is the enzyme that has been studied most extensively. The mature, active protease molecule includes 223 amino acids with an apparent molecular weight of 23,800 (Stroud 1974). Serine protease hydrolysis of the peptide (or ester) bond proceeds through an acyl-enzyme intermediate (Figure i-2). The oxygen atom of the activated serine in the catalytic triad (Asp102-His57-Ser195) attacks the carbon atom of the peptide bond carboxy terminal to a lysine or arginine amino acid in the substrate to form a covalent complex. Breakdown of this intermediate by reaction with water renews the enzyme (Figure i-2). Only the amino acids arginine and lysine are effective natural substrates for trypsin (Keil-Dlouhá, Zylber et al. 1971a; Keil-Dlouhá, Zylber et al. 1971b)[2]. Among the other natural amino acids, the specificity towards tyrosine is superior (Gráf, Craik et al. 1987). However, relative to

[2]Incubation of glucagon for 50 hours at pH 8.0, 25°C yielded no detectable cleavage of bonds other than those carboxy terminal to Lys or Arg.

Figure i-2. The Trypsin Reaction Coordinate.

According to convention, the vertical axis is the free energy of the system. The horizontal axis is the reaction coordinate. The trypsin reaction has two transition states in which bonds are broken and made and has a single intermediate, the acyl-enzyme. As the products have a lower internal energy than reactants under physiological conditions, the reaction proceeds with net energy release, and, so, is essentially irreversible.

Trypsin Reaction Coordinate



lysine, tyrosine is a worse substrate by 1 x 10^5 fold with a kcat/Km value of 10.5 versus 7.3 x 10^5 min $^{-1}\mu M^{-1}$ on 3-carboxypropanoyl-Ala-Ala-Pro-X-AMC at 37°C in 50 mM Tris-HCl pH 8.0, 10 mM $CaCl_2$ (Gráf, Jancsó et al. 1988). Investigation of this strong preference for arginine and lysine using synthetic substrates has established that the length and charge of the side chain are critical features of an effective substrate (Table i-1).

Interaction between the protein surface and the amino acids of the substrate both amino terminal and carboxy terminal to the arginine or lysine contribute to the specificity of the reaction[3]. Generally, longer substrates exhibit decreased Km and elevated kcat values although the differences are generally small in comparison with the overall 10^{10} fold rate enhancement achieved by trypsin relative to the rate of amidolysis in neutral water (Kahne and Still 1988). However, exceptional combinations of amino acids do have particularly strong effects on the hydrolysis of the P1-P1' peptide bond. The kcat/Km values of a set of substrates differing only at the P2 or P3 position have been determined, and show that the worst and best substrates differ by approximately seven fold at each position (Tanaka, McRae et al. 1983; Fiedler 1987). For chymotrypsin, amino acids as distant from the active site as P3 and P2' can apparently exert an effect on catalysis

---

[3]According to convention, the substrate is numbered such that the P1-P1' peptide bond is cleaved (Schechter and Berger, 1967). The amino acids are numbered sequentially in both the amino terminal (P2, P3...) and carboxy terminal (P2', P3', ...) directions from P1 and P1'. The subsites on the enzyme correspond to the amino acid bound, so the the P1 amino acid binds in the S1 pocket, the P2' amino acid in the S2' pocket. For synthetic substrates, the reporter group is the P1' residue.

Table i-1. Characterizing the S1 Pocket Using Synthetic Substrates

Reading from left to right, the first column details the substrate that was synthesized. Only experiments reporting results with single residue ester substrates are listed so that comparisons are possible. No effort was made to judge the quality of the synthesis or of the trypsin utilized. In most cases, the concentration of trypsin was obtained by calculation from the A_{280} reading rather than active site titration. In the second column, the conditions of the assay are summarized. Most of the experiments were done in a pH-stat. The observed Km and kcat are recorded in the next two columns. Where data was unobtainable, a dash is used. The two Michaelis-Menten parameters were obtained by a regression analysis of the kinetic data. The Km is expressed in M and the kcat is per sec. The second order rate constant kcat/Km is calculated from the Km and kcat. The source of the data is listed in the final column. Relative to the lysine substrate, heptyline has a carbon in place of the η nitrogen. Citrulline is the uncharged analog of arginine as it has an amide group at the terminus of the side chain. Homoarginine has an additional methylene group relative to the three present in the arginine side chain while amidobutyric acid has one fewer. The ornithine side chain is shorter than that of lysine by one methylene group.

Substrate	Conditions	observed Km (M)	observed kcat (/sec)	kcat/Km (/ sec / M)	references
N-acetyl-glycine-OMe	pH 6.6, 25°C, 0.1 M KCl	1.1	0.013	0.01	Inagami, Mitsuda, 1964
N-acetyl-glycine-OEt	pH 7.0, 25°C, 0.1 M KCl	1.4	0.041	0.03	Sanborn, Hein , 1968
N-acetyl-alanine-OEt	pH 6.6, 25°C, 0.1 M KCl	0.88	0.85	1.0	Vajda, Szabó,1976
N-acetyl-norvaline-OEt	pH 6.6, 25°C, 0.1 M KCl	0.14	0.88	6.3	Vajda, Szabó, 1976
N-acetyl-norleucine-OEt	pH 6.6, 25°C, 0.1 M KCl	-	-	9.2	Vajda, Szabó, 1976
N-acetyl-heptyline-OMe	pH 7.0, 25°C, 0.1 M KCl	4.90E-2	4.40E-1	9.0	Sanborn, Hein , 1968
N-acetyl-lysine-OMe	pH 7.0, 25°C, 0.1 M KCl	2.50E-2	5.20E+1	2080	Sanborn, Hein , 1968
N-benzoyl-heptyline-OMe	pH 7.0, 25°C, 0.1 M KCl	1.00E-4	9.40E-3	94	Sanborn, Hein , 1968
N-benzoyl-lysine-OMe	pH 7.0, 25°C, 0.1 M KCl	5.50E-5	1.80E+0	3.27E+4	Sanborn, Hein , 1968
N-tosyl-lysine-OMe	pH 7.0, 25°C, 0.1 M KCl	1.90E-2	1.40E+0	74	Sanborn, Hein , 1968
N-tosyl-lysine-OMe	pH 6.8, 21°C, O.1 M phosphate	3.20E-4	1.00E+1	3.13E+4	Ascenzi et al 1982
N-acetyl-lysine-OMe	pH 6.8, 21°C, O.1 M phosphate	3.20E-4	9.00E+0	2.81E+4	Ascenzi et al 1982
N-tosyl-ornithine-OMe	pH 7.0, 25°C, 0.1 M KCl	1.90E-2	1.40E+0	74	Sanborn, Hein , 1968
	pH 7.0, 25°C, 0.1 M NaCl, 2 mM CaCl2	1.50E-1	3.36E+0	22	Baird et al, 1965
TAME: N-tosyl-arginine-OMe	pH 8.0, 25°C, 0.1 M KCl,	1.30E-05	6.00E+01	4.62E+6	Trowbridge et al, 1963
	pH 6.8, 21°C, O.1 M phosphate	3.00E-4	1.00E+01	3.33E+4	Ascenzi et al 1982
	pH 8.4, 25°C, 0.1 M NaCl, 2 mM CaCl2	6.40E-6	75	1.17E+7	Baines et al, 1964; Hummel 1959
N-benzoyl-arginine-OMe	pH 6.8, 21°C, O.1 M phosphate	2.80E-4	14	5.00E+4	Ascenzi et al 1982
N-benzoyl-arginine-OMe	pH 7.0, 25°C, 0.1 M KCl,	9.60E-6	9.3	9.69E+5	Sanborn, Hein, 1968
N-benzoyl-arginine-OEt	pH 6.8, 21°C, O.1 M phosphate	3.00E-4	11	3.67E+4	Ascenzi et al 1982
γ-guanidino-N-tosyl-amidobutyric methyl ester					
	pH 7.0, 25°C, 0.1 M NaCl, 2 mM CaCl2	4.54E-4	18.5	4.07E+4	Baird et al, 1965
	pH 8.0, 25°C, 0.1 M NaCl, 2 mM CaCl2	6.90E-4	38.1	5.52E+4	Baird et al, 1965
methyl ß-guanidino-tosyl-amidobutyric methyl ester					
		-	-	20 <	Baird et al,1965; Izumiya et al, 1959
N-tosyl-homoarginine-OMe	pH 8.4, 25°C, 0.1 M NaCl, 2 mM CaCl2	3.32E-4	4.04	1.22E+4	Baines et al, 1964 (Kitagawa and Izumiya, 1959)
N-tosyl-citrulline-OMe	pH 7.0, 25°C, 0.1 M KCl	9.10E-2	0.39	4	Sanborn, Hein , 1968
N-benzyl-citrulline-OMe	pH 7.0, 25°C, 0.1 M KCl	4.10E-2	0.014	0.34	Sanborn, Hein , 1968

(J. Powers, personal communication) and this seems to be true for trypsin as well (Fiedler 1987).

The S1 pocket is a deep cylindrical crater in the spherical molecule measuring roughly 12 Å deep with a diameter of 5 Å (Kreiger, Kay et al. 1974; Ruhlmann, Kukla et al. 1973). It is punctuated at its base by a negative charge, the carboxylate of aspartic acid 189 (Figure i-3, i-4). The walls of the pocket are formed largely by main chain atoms although tyrosine 228 contributes its phenol ring to wall-off part of the pocket. Each oxygen of the D189 carboxylate apparently forms a direct a hydrogen bond to a nitrogen of the Arg substrate guanidinium group. In contrast, the D189 carboxylate must interact with the Lys substrate amino group via a structural water molecule. Other atoms that are hydrogen bonding partners from the base of the pocket to the substrate include the serine 190 O and Oγ, the glycine 219 O, the tyrosine 228 hydroxyl O, and two structural waters.

Although informative, the three dimensional structure alone is not sufficient to quantify the forces that characterize the enzyme-substrate interactions. The use of synthetic substrates is limited by chemistry. Application of site-directed mutagenesis to the cloned gene, expression, and characterization of the mutant enzyme, which together constitute protein engineering, is the complementary tool. Several protein engineering studies have focused on the nature and the magnitude of the interactions between enzymes and their substrates in an effort to understand specificity. Pioneering work was performed on tyrosyl tRNA synthetase which catalyzes the formation of tyrosyl adenylate.

Figure i-3. The Trypsin Substrate Binding Pocket

Ektacolor 100 picture taken using the Silicon Graphics picture system and the graphic display package Insight. The structure is of the rat anionic trypsin S195C which was solved to an average R factor of 16% with good stereochemistry (McGrath, Wilke et al. 1989). The orientation is the same one that would be observed from the tip of the P1 side chain approaching the S1 binding pocket. The backbone is traced in white except for the glycine amino acids 216 and 219 in pink. Double bonds, as in the carbonyl, are indicated with with a second white line; double bond character, as in the peptide backbone and the tyrosine ring, are indicated with a second dotted line. The catalytic triad is shown at the top of the picture for orientation and atoms are shown with their appropriate van der Waal's surfaces. The main chain atoms of these amino acids are in grey. The aspartic acid 102 carboxylate is in orange and the van der Waal's surface is displayed in pink; the histidine 57 imidazole is partially visible in white and the van der Waal's surface is displayed in blue; the serine 195 methoxy is in pink and the van der Waal's surface is displayed in orange. The primary specificity pocket is formed by amino acids in the carboxy terminal ß barrel of the two that comprise trypsin, namely, amino acids 214-220, 225-228, and 189-195. Structural waters also contribute and these are shown as white stars. At glycine 216, the ß strand is kinked and runs away from the pocket to return at position 219. This excursion is larger in chymotrypsin and elastase. There is a disulfide bridge connecting cysteines 191 and 220 which is shown in yellow. A kink, in the protein backbone at proline 225, positions

glycine 226 to form part of the pocket. This ß strand continues away for the pocket although the entire tyrosine 228 side chain (in grey with the hydroxyl in pink) forms part of the pocket. The side chains of aspartic acid 189 (in red with the van der Waal's surface in red) and serine 190 (in pink and with the van der Waal's surface displayed in orange) define the base of the pocket. The glutamine 192 side chain (not shown) helps define the mouth of the pocket. The main chain atoms of amino acids 189 and 190 are in grey. A ß turn including amino acids 193-195 creates the small pocket that stabilizes the oxyanion in the transition state.

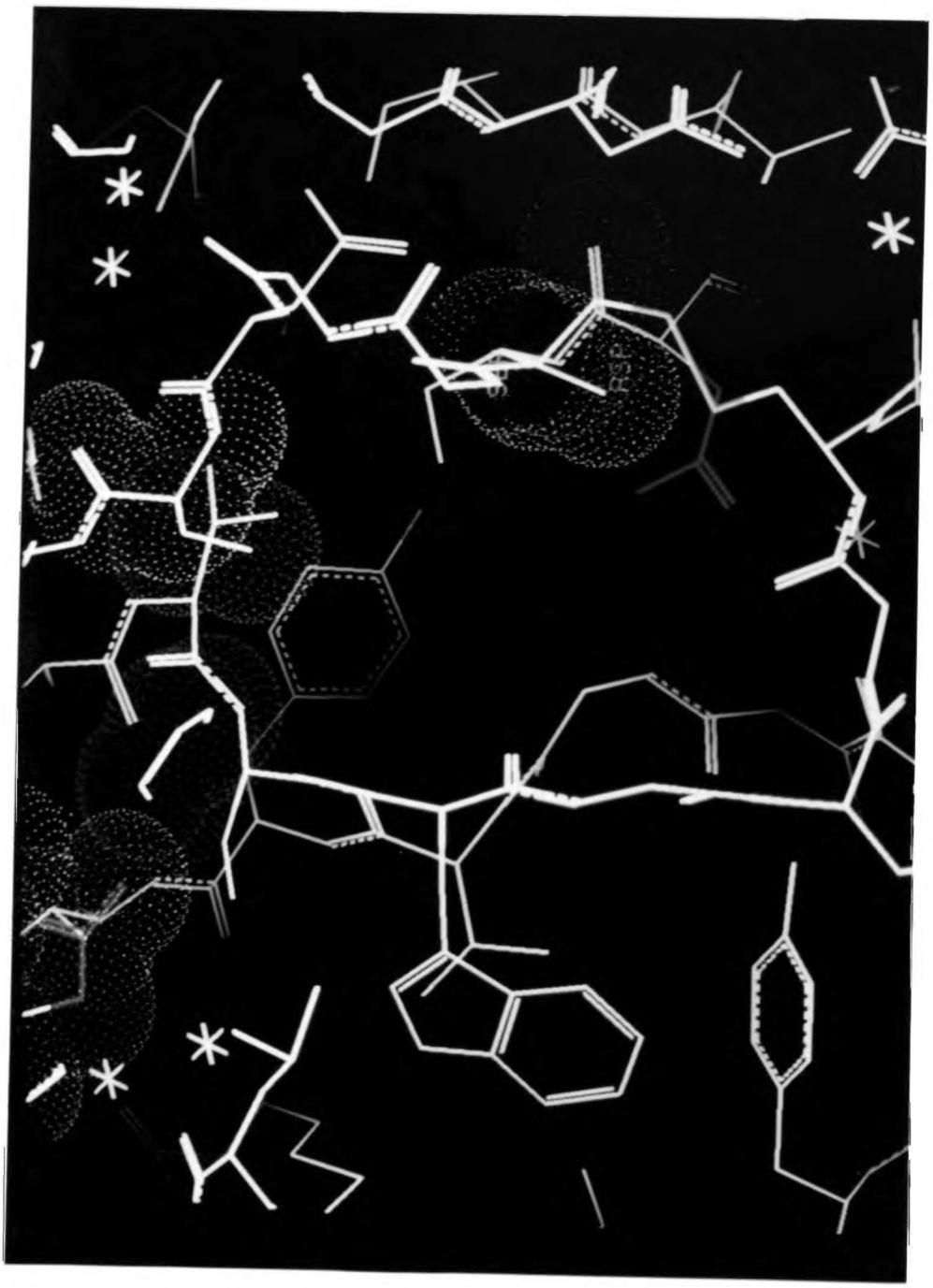


Figure i-4. Schematic of the Trypsin-Substrate Interactions in the Binding Pocket

The main chain and some of the side chain atoms that comprise the substrate binding pocket are shown. The position numbers are placed at the Cα carbons. Atoms that contribute directly to the hydrogen bonding between enzyme and substrate are depicted explicitly. The eponymous Ser195 is included for perspective. Waters are denoted as W. The disulfide bond pictured in the foreground connects amino acid positions 191 and 220. The substrate arginine and lysine amino acids are also shown; the peptide backbone goes into the page 5' to 3'. The two structures were solved in the same laboratory allowing ready comparison of the numbered solvent molecules.

A.) A representation of the pocket with an Arg side chain bound. The atom positions were taken from the structure of the Arg15 analogof BPTI complexed to trypsinogen in the presence of Val-Val (File 4TPI in the Brookhaven Protein Data Bank; Bode, Walter et al. 1984). The position of the Arg Nη nitrogen atoms are superimposable with the nitrogens of the amidino group of benzamidine in the trypsin-benzamidine complex (File 3PTB in the Brookhaven Protein Data Bank).

B.) A representation of the pocket with a Lys substrate bound. The atom positions are taken from the structure of BPTI complexed to trypsin (File 2PTC in the Brookhaven Protein Data Bank; Ruhlmann, Kukla et al. 1973). The depiction of this complex relative to Fig. i-4a is rotated 30° in the Y plane and -30° in the X plane to depict additional features of the binding pocket.

228
195
Oγ
416W
190
189
Oγ
Oδ2
191
226
Nζ
Oδ1
I15
414W
215
220
216
219

228
226
416W
Oγ
190
Oδ2
189
Nη2
Oδ1
215
195
Oγ
Nη1
216
220
I15
ARG
219

(tyr-AMP) via nucleophilic attack of the tyrosine carboxylate on ATP (Fersht, Shi et al. 1985). Numerous site-directed mutants were generated which were altered at amino acids that appeared, in the three dimensional structure, to be in direct contact with one of the two substrates. For example, the phenolic hydroxyl of tyrosine 169 appeared to form a hydrogen bond to the α-amino group of the amino acid substrate in the ground state. Deletion of this hydroxyl by replacing tyrosine with phenylalanine coordinately reduced the binding interaction between the enzyme and substrate, transition state, and product. In contrast, replacement of histidine 45 with alanine revealed a hydrogen bonding interaction confined to the transition state of the reaction. From these and other mutants, it was concluded that the specificity of an enzyme is mediated by hydrogen bonding to or from charged or polar residues. An unpaired charge reduces the specificity by a factor of 1000 while a loss of 2.5 - 15 fold is incurred by leaving uncharged hydrogen bonding groups without partners. It was also noted that steric repulsion figured importantly in specificity as van der Waal's repulsion energies are extremely sensitive to interatomic distances.

Yet, other experiments have indicated that hydrogen bonding is not the dominant force in determining specificity. The serine protease subtilisin[4] preferentially hydrolyzes the peptide bond carboxy terminal to large, hydrophobic amino acids such as

---

[4] The subtilisin serine proteases are not structurally related to the trypsin-like enzymes.

tyrosine or phenylalanine[5]. Twelve non-charged mutations were made at position 166 at the base of the S1 cleft (Estell, Graycar et al. 1986). It was observed that as the volume of the side chain at position 166 increases, the substrate preference shifts from large to small P1 side chains. For example, leucine in place of glycine 166 was a much more active enzyme on Met substrates than Phe or Tyr. In addition, the substrate specificity (kcat/Km) of some enzyme substrate pairs exceeded the optimal wild type pair (glycine 166 on tyrosine substrates). However, any enzyme-substrate pairs that exceeded the capacity (volume) of the pocket exhibited dramatically lower reaction rates. The data indicated that the drive to sequester hydrophobic moieties from water, an entropic effect, was the driving force in the specificity of the broad S1 cleft of subtilisin.

The specificity of trypsin toward substrates with an Arg and Lys at the P1 position is apparently strongly influenced by charge-charge interactions between the positively charged terminus of the P1 side chain and the aspartic acid 189. The nature of charge-charge interactions, as they relate to specificity, has been addressed by protein engineering in trypsin and other systems. Trypsin D189K was recombinantly generated to examine the effects of a reversal of the electrostatic interaction between the lysine and the aspartic acid 189. It was inactive on Lys or Arg substrates, and showed no specificity for a Glu substrate (Gráf, Craik et al. 1987). More detailed analysis revealed that the D189K

---

[5] Subtilisin exhibits a broad specificity that is strongly influenced by the amino acid at the P4 site.

mutant retained similar Phe-specific activity to trypsin indicating that the catalytic triad of the enzyme was still functioning. In addition, trypsin D189K showed Leu-specific activity 100 fold greater than that observed for trypsin (Gráf, Craik et al. 1987) although the kcat of this reaction was reduced by 10^5 fold relative to the reaction of trypsin on its preferred substrates. Thus, polarity reversal in trypsin could not be achieved with this single amino acid substitution and probably requires additional remodeling of the binding pocket and may even require changes of the amino acids outside of the substrate binding pocket.

E. coli aspartate aminotransferase (transaminase) was mutated in an effort to reverse the preference of acidic amino acid substrates to basic ones (Cronin, Malcolm et al. 1987). This pyridoxal phosphate-dependent enzyme catalyzes the reversible transamination of α-amino to α-keto acids. Aspartate aminotransferase R292D was engineered and evaluated on lysine and arginine substrates, and on substrates, such as aspartate or α-ketoglutarate, preferred by the wild type enzyme. Against Arg, the mutant is 15.5 fold more specific (kcat/Km) than wild type. However, this new specificity was acquired at the expense of a drop in kcat of 10^5. Also, the R292D mutant preferred Arg over Glu substrates only 6 fold where the aspartate aminotransferase was able to discriminate 10^5 fold against Arg in favor of its preferred substrates. The activity of the mutant enzyme toward citrulline, an uncharged analog of arginine, suggested that the preference of the mutant for basic amino acids resulted not from

the introduction of a favorable interaction, but from the removal of an unfavorable one (Cronin and Kirsch 1988).

In apparent contrast to the character of the above results to alter charge specificity, the charge specificity of the broad S1 pocket of subtilisin was malleable. Two amino acids in the S1 cleft, glutamic acid 156 and glycine 166, were changed alone and in concert to glutamine and serine for position 156 and to D, N, M, G, K, and R for position 166. Differences were observed as large as 1000 fold among the amino acid combinations toward a given substrate. For example, subtilisin G166D was 5 fold more specific for Lys than for Met while the mutant E156Q exhibited a 50 fold preference for Met over Lys substrates (Wells, Powers et al. 1987). Also, the kcat value of these reactions was within 100 fold of the kcat for the optimal substrate and the wild type enzyme. However, it is important to note that subtilisin has a less discriminating binding pocket than that found in trypsin and catalyzes reactions with a range of amino acid substrates (Estell, Graycar et al. 1986). Furthermore, high level activity in subtilisin is dependent on subsites unlike the activity of trypsin (Estell, Graycar et al. 1986).

The energetic cost of reversing the polarity of the ion pair in an enzyme active site that has evolved to accomodate a specific ion pair has been calculated semi-quantitatively (Hwang and Warshel 1988). These results indicated that the polarity reversal can not be stabilized in the microenvironment that evolved for the original

polarity[6]. The explanation of the apparent manipulation of the charge specificity in subtilisin (Wells, Powers et al. 1987) relates to the broad specificity of the S1 cleft which has not evolved to stabilize a charged side chain for efficient catalysis.

Other components in the roughly equivalent specificity of trypsin for Arg and Lys substrates have been examined in a separate set of experiments. Two glycines at the base, positions 216 and 226, of the trypsin S1 pocket were mutated to alanines (Craik, Largman et al. 1985). Kinetic analysis of trypsin G216A revealed that addition of a methyl group at this position caused a substantial but equivalent decrease in the ability to bind Arg and Lys side chains (elevated Km). However, the ability to stabilize the transition state of the reaction with Lys (kcat) was perturbed to a greater extent. The Arg specificity was reduced to a greater extent than the Lys in trypsin G226A. The Lys amino group doesn't hydrogen bond to aspartate 189, but is connected to that carboxylate only through water 414. This water is probably displaced by the G216A mutation. Arg directly interacts with the aspartic acid 189. Insertion of the methyl group at position 226 creates a critically short contact with aspartic acid 189 and may cause deformation of that portion of the pocket. The differences in how the two side chains interact with aspartic acid 189 may be the

[6]In fact, a prize of $100,000 from Alliant Inc. has been set aside for anyone who successfully achieves polarity reversal (Hwang and Warshel, 1988).

basis of the differential catalytic effects observed for the G226A mutant[7].

The techniques of protein engineering have ensured that an almost limitless array of mutants can be generated. However, characterization of trypsin mutants which generally involves kinetic analysis is time consuming. Consequently, broadly targeted engineering experiments, such as examination of the P1 binding pocket, would benefit by a technique that could define a class of informative mutants. Use of genetic selection or screening to identify informative mutants for structure::function study is established (Helinski and Yanofsky 1963). In principal, a genetic screen or selection allows a search for critical amino acids or amino acid combinations by casting a far wider net than is possible using screens or clonal analysis. Nonetheless, the number of possible mutations that can be examined simultaneously, even with a genetic selection, is still limited. For example, to examine n positions, a library of mutants must contain 20^n members assuming the nucleotides of the n codons are equally represented[8]. Thus, the library of double mutations at two positions is 400 but the library for 6 positions is 64 x 10^6. Assuming that one plate in the selection scans 10^5 possible solutions, 640 plates are required to cover the library a single time. Consequently, it is particularly useful to refine the variables in the problem so as to maximize the

---

[7] A three dimensional structure of trypsin G226A complexed to benzamidine indicates that a novel network of hydrogen bonds is present (Wilke et al, 1990).

[8] This represents something of an oversimplification of the problem of obtaining a representative collection of mutants as the genetic code is biased toward some amino acids.

power of the selection. For example, one region of the protein can be mutated fairly heavily, or the entire protein can be mutated less severely. Alternatively, the entire set of a defined collection of amino acid positions can be mutated to saturation.

Use of experimental systems in which a three dimensional structure allows protein engineering and in which a genetic selection allows broad searching of mutant libraries is still rare. This combination has never been used for the study of substrate specificity[9] although other issues have been considered. Two systems which have benefited from this approach are ones involving triose phosphate isomerase and one involving λ repressor. Triose phosphate isomerase (TIM) catalyzes the interconversion of dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate. An aspartic acid mutant was uncovered at the catalytic glutamic acid 165 that retained 0.1% the potency of the wild type enzyme. Contiguous regions of TIM of 23-34 amino acids were mutated with a likelihood of recovering approximately 30% single and double DNA substitutions, and 20% triple DNA substitutions for a maximum of three amino acid changes in the region. The search across 7 different pools of mutants allowed recovery of six second site repressors of the E165D mutant. When mapped to the three dimensional structure, all six mutants were found within 5 Å of the active site. The conclusion was advanced that different arrangements of catalytic groups from a stable scaffolding may achieve a functional

---

[9] A study of the substrate specificity of ß-lactamase prior to the solution of the three dimensional structure was accomplished (Schultz and Richards, 1986).

arrangement. Furthermore, only amino acids nearby the catalytic regions of the molecule are sensitive to mutation while the other amino acids are important only as part of the evolutionarily optimized scaffold.

The λ DNA repressor binds to the cognate DNA operator and functions as the linchpin in the bacteriophage decision to lytically or lysogenically infect a bacterium (Delbruck and Weigle 1951; Lederberg 1951). A set of 7 core amino acids and a stretch of 17 amino acids that lie at or near the dimer interface and away from the DNA binding face of the molecule were examined (Reidharr-Olson and Sauer 1988; Bowie, Reidharr-Olson et al. 1990). Using cassette mutagenesis, every amino acid was changed singly to all others and various combinations were also tested. The pool of mutant repressors was searched by selection and the amino acid replacements were catalogued. Most solvent accessible positions tolerated multiple substitutions even between amino acids that reversed the polarity or the hydrophobicity of the position. In contrast, substitution of the core, while still readily achieved, maintained the hydrophobic character of the position. Apparently, the structural information at most positions is degenerate.

This dissertation represents a quest for a detailed description of the forces involved in the stabilization of the substrates in the primary substrate binding pocket of trypsin. I have attempted to use tools to investigate this problem that have never been used in the analysis of serine proteases. In particular, the development of a genetic selection has been a technical objective. The selections devised for peptidase N (Miller, Green et al. 1982) and

carboxypeptidase Y (Rothman and Stevens 1986) delineated relevant starting points; however, it was not clear how to get *E. coli* to depend on trypsin for survival. Potential problems included the necessity for high level expression in bacteria of an active protease. Intellectually, several elements of the comprehensive question have been pursued, including an assessment of what are the critical determinants of specificity. Also, how do these amino acids interact to achieve the precise positioning of the substrate? In particular, how can roughly equivalent catalytic activity toward structurally disparate amino acids, arginine and lysine, be maintained?

In the first of four chapters, the tools and techniques for expression and mutagenesis are presented; in the second chapter, the application to the study of position 189 in substrate specificity is presented. The third and fourth chapters follow the same pattern although the techniques described in chapter three involve genetic selection. The fourth chapter describes an investigation of how position 189 and 190 define the base of the pocket. I have attempted to write each chapter so that it could be considered independently; as it is said, "write for the grant, write for the paper, write for the thesis, write for the review, and, always, write for posterity."

Chapter One

A System for Generating, Expressing, and Screening Trypsin and Trypsin Mutants

ABSTRACT

To facilitate the dissection of the arginine and lysine substrate specificity exhibited by the serine protease trypsin, a system has been developed that allows over-expression of both trypsinogen and mature trypsin in *Escherichia coli*. Expression and secretion are controlled by the alkaline phosphatase promoter and signal peptide allowing accumulation of trypsin to levels of 0.2 mg of trypsin per A_{600} of culture (~3.0 OD of cells /L). This system has been adapted to generate and express trypsin mutants in the same vector by incorporation of the bacteriophage M13 origin in the vector. A rapid, single-substrate gel overlay activity assay has been developed which is sensitive to picogram quantities of trypsin. As the trypsin is separated via SDS-PAGE prior to activity analysis, specific detection of trypsin is enabled. This screen concept can be adapted for analysis of different substrate specificities. In preparation for the analysis of the specificities of trypsins mutated at position 189, a library of trypsin mutants differing only at position 189 was created using oligonucleotide-directed mutagenesis. The aspartic acid at this position is a major determinant of substrate specificity. Efforts were made to ensure equal representation of each mutant in the library but were only partially successful. Of the 19 possible mutants, 15 were identified.

INTRODUCTION

Biochemical and biophysical characterization of sets of point mutants have been particularly useful in establishing how an amino acid in a particular environment can function, and this work has begun to build a data base from which general principles may be deduced (Alber, Dao-ping et al. 1987; Reidharr-Olson and Sauer 1988). Hydrogen bonding groups participating in the ground and transition state of the reaction were identified and the strength of these hydrogen bonds was analyzed (Fersht 1985). Hydrophobic interactions have also been shown to be important and even dominant in certain enzyme substrate interactions (Estell, Graycar et al. 1986). This approach toward structure::function investigation depends on knowledge of the three dimensional structure and on the capability to readily perform site-directed mutagenesis of the cloned gene and obtain the purified mutant protein.

Characterization of a protein by mutant analysis is expedited by limiting examination to the study of functional proteins. A genetic screen or high throughput analysis of extracts can be used in this procedure. Examples of genetic screens include release of 5-bromo-4-chloro-3-indoyl from X-gal for ß-galactosidase detection (Miller 1972), precipitation of a diazonium salt such as fast garnet red with free ß-naphthylamide for peptidase N detection (Miller, Green et al. 1982), and the reaction between free nucleotides and toluidene blue for staphylococcal nuclease detection (Shortle and Lin 1985). In yeast cells, carboxypeptidase Y activity in the periplasm was measured by following a color change using L-amino acid oxidase, ß-diansidine, and horseradish peroxidase (Rothman and Stevens

1986). Classical examples of high throughput screens include the use of casein overlay gels to detect protease activity (Sarath, Motte et al. 1989) and the array of spectrophotometric assays (Young, Liu et al. 1987).

The serine protease trypsin hydrolyzes peptide and ester bonds following arginine and lysine amino acids. The catalytic mechanism and the substrate specificity have been investigated extensively (for review see Kraut 1977). The three-dimensional structures of bovine cationic trypsin (Kreiger, Kay et al. 1974; Ruhlmann, Kukla et al. 1973; Marquart, Walter et al. 1983) and rat anionic trypsin (Sprang, Standing et al. 1987) have been solved to high resolution, and the rat gene has been cloned (Craik, Choo et al. 1984) and expressed in mammalian cells (Craik, Largman et al. 1985) and in bacteria (Gráf, Craik et al. 1987)[10]. There is 40% identity between the amino acid sequences of trypsin, chymotrypsin, and elastase although the backbone structure and the amino acids involved in the catalytic and binding sites are highly conserved. The primary substrate binding pocket of these enzymes is defined by main chain atoms in three ß strands in the carboxy terminal ß barrel of the molecule. In contrast to chymotrypsin and elastase, the base of the trypsin pocket is also defined by the aspartic acid 189 carboxylate and serine 190 methoxy groups. The positioning of an aspartic acid at the base of the deep substrate pocket was determined to be the essential determinant of the specificity trypsin exhibits toward

---

[10]The bovine cationic trypsin gene has also been cloned and expressed in *E. coli*, but as the expression levels are substantially lower than those for the rat enzyme, the focus has been on the rat anionic trypsin (J. Higaki, personal communication).

basic amino acids (Kreiger, Kay et al. 1974; Ruhlmann, Kukla et al. 1973).

However, initial attempts to convert the specificity of trypsin to the specificity exhibited by the homologous protease chymotrypsin has shown that the specificity of trypsin is dependent on other elements in addition to the amino acid at position 189 (Gráf, Craik et al. 1987; Gráf, Jancsó et al. 1988)[11]. Consequently, trypsin must function in a different fashion than subtilisin (Estell, Graycar et al. 1986) or α-lytic protease (Bone, Silen et al. 1989). In these enzymes, changing the amino acid at the base of the substrate specificity pocket has been shown to affect substrate specificity in a predictable fashion.

A novel expression system allows mature trypsin to be expressed to high levels in the periplasmic space without apparent effect on the bacteria. We have also created a system to facilitate the generation, expression, and characterization of mutants in trypsin. This system involves the use of a portion of the M13 sequences (Dotto and Zinder 1981) that confers on the plasmid the ability to form single stranded copies in the presence of the appropriate set of replication proteins (Dente, Cesareni et al. 1983). This same plasmid is used for expression. Characterization of the trypsin (and other serine proteases) can be accomplished directly from the crude periplasmic extract with the use of a novel gel

---

[11]Additional single and multiple mutations including D189S/Q193M have been made so that the trypsin binding pocket would more closely resemble that of chymotrypsin. These experiments have yet to substantially enhance the intrinsic chymotryptic activity of trypsin (László Szilágyi, personal communication 1989) which was characterized in the study of trypsin D189K (Gráf et al. 1987).

MATERIALS AND METHODS

Materials

Restriction enzymes, T4 DNA ligase, and T4 polynucleotide kinase were purchased from New England Biolabs or Boehringer Mannheim Biochemicals. T4 DNA polymerase was obtained from New England Biolabs Inc. (Beverly, MA). T7 DNA polymerase (form II) was from United States Biochemicals Corp. (Cleveland, OH). Nα-*p*-tosyl-arginyl methyl ester (TAME), acetyl-phenylalanine methyl ester (Ac-Phe OMe), acetyl-tyrosine methyl ester (Ac-Tyr OMe), benzyl-alanine methyl ester (Bz-Ala OMe), and acetyl-Ala-Ala-Ala methyl ester (Ac-$(Ala)_3$ OMe) were from Sigma Chemical Co. (St. Louis, MO). Phenol red (phenolsulfonephthalein) was from Fisher Scientific Corp. The *trpA* transcription terminator (Christie, Farnham et al. 1981) was purchased as a 28 bp DNA duplex from Pharmacia Inc. (Piscataway, NJ); the sense strand was 5'-AGCCCGCCTAATGAGCGGGCTTTTTTTT-3', and the anti-sense was the complement. Bovine cationic ß-trypsin was purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN), while bovine α-chymotrypsin and porcine pancreatic elastase were from Sigma Chemical Co.. Plasmid pBluescript (KS, plus), VCS M13 and R408 helper phage were products of Stratagene Inc. (La Jolla, CA).

Bacterial Strains

E. coli SM138 (F^-/ araD139, Δ(lac)169, relA1, rpsL, phoR) and SM133 (F^-/ phoR, proC::Tn5, trp(ochre), rpsL) were gifts of Dr. S. Michaelis (University of California- San Francisco). MH1 (F^-/ araD139, ΔlacX74, galU, galK, hsr^-, hsm^+, rpsL) was from Dr. M. Hall

(University of California- San Francisco). LE101 (F^-/ araD139, ΔlacX74, galU, galK, hsr^-, hsm^+, rpsL, phoR, proC::Tn5) was obtained by P1 transduction (Silhavy, Berman et al. 1984) of the phoR and Tn5 (kanamycin resistance) markers from SM133 into MH1 to combine the high DNA transformation efficiency property of MH1 with the ability to constitutively produce proteins regulated by the *phoA* promoter (phoR repressor). CJ236 (dut, ung, relA, pCJ105 (pilis formation, Cm^r)) was provided by Biorad Laboratories (Richmond, CA).

Plasmid Constructions

The enzymes were used according to suppliers recommendations except that ligations involving blunt-ended fragments were done at room temperature in a buffer including 5% PEG 6000 (Pheiffer and Zimmerman 1983). Oligonucleotides were synthesized on a 380B DNA synthesizer (Applied Biosystems Inc., Foster City). Purification of the oligonucleotides was accomplished by electrophoresis through a 20% polyacrylamide gel in Tris-borate buffer, excision of the appropriate gel slice, and elution of the DNA, by diffusion, into 1x TE (10mM Tris pH 8.0, 0.1mM EDTA). Oligonucleotides were further purified on a Sep-Pak C18 cartridge (Waters Associates, Milford).

The *trpA* terminator oligonucleotides were boiled and placed at 60°C for one hr in 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM ß-mercaptoethanol. The pBsTn vector had been digested with Cla I and blunt-ended with the Klenow fragment of DNA polymerase I (Maniatis, Fritsch et al. 1982). Successful ligation of the terminator

occurred at a 50 fold molar excess over the vector fragment. Identification of positive clones was by colony screening using radioactively labelled terminator oligonucleotides as the probe.

The alkaline phosphatase promoter and signal peptide were from pFOG402, a gift of H. Inouye (Inouye, Barnes et al. 1982). pGAP was a gift of Dr. S. Rosenberg (Chiron Corporation, Emeryville, CA). pBS (KS orientation of the polylinker, plus orientation of the M13 origin) was a product of Stratagene Inc. (La Jolla, CA).

Single-stranded DNA Isolation

Site-directed oligonucleotide mutagenesis without phenotypic selection (Kunkel 1985; Kunkel, Roberts et al. 1987) was adapted for use with vectors that also contain an M13 origin of replication (Dente, Cesareni et al. 1983). Single-stranded plasmid DNA was obtained as described (Dente, Cesareni et al. 1983) with substantial modifications. The plasmid to be mutagenized was transformed into strain CJ236 using a rapid transformation protocol (Hanahan 1985). VCS M13 helper virus (Messing 1983) was added to a logarithmically growing culture at a multiplicity of infection (MOI) of 1:1 to 10:1 virus to cells; the lower MOI provided equivalent yields of plasmid with less helper virus. Although the culture can be initiated from a single colony, a 1:10 dilution of a saturated overnight culture to a A_{600} of 0.2-0.3 is more convenient. Virus were harvested 6-8 hours later; cells were eliminated by two rounds of centrifugation or by centrifugation followed by filtration of the supernatant through 0.2 or 0.45μM filter. Virus were precipitated by addition of a 1/4 volume of a solution of 20% PEG,

3.5M NH_4OAc. Approximately, 0.25 mg of uracil-laden single-stranded DNA per ml of culture was typically recovered. Although slightly higher yields of single stranded plasmid DNA were obtained with the R408 helper phage (Russell, Kidd et al. 1986), the ratio of helper phage DNA to plasmid was greater and so VCS M13 was routinely used. Optimization of these conditions was a collaborative effort with Biorad Laboratories (Geisselsoder, Witney et al. 1987; McClary, Witney et al. 1989). The presence of uracil was confirmed by comparing the efficiency of transformation of the isolated single-stranded DNA against control DNA obtained from a *dut+ ung+* host. Addition of uridine to the growth media is not necessary. Approximately 0.1% of the colonies obtained with the control DNA were obtained with an equal amount of the uracil impregnated DNA.

Mutagenesis

Oligonucleotides were phosphorylated in a reaction cocktail containing 50 mM Tris pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM spermidine, 0.4 mM rATP, 1 U T4 polynucleotide kinase, and 100-200 pM of DNA by incubation at 37°C for 45 min (Maniatis, Fritsch et al. 1982). Phosphorylated oligonucleotides and the uracil-laden template were added together at a molar ratio of 25:1 (6.0 pM in 1x phosphorylation buffer and 0.4 pM in 1x TE respectively), boiled, and then allowed to anneal for 15-20 minutes in 10mM $MgCl_2$, 20mM Tris pH 7.4 at a temperature empirically determined by a prior analysis using dideoxy sequencing (Sanger, Nicklen et al. 1977) (for oligonucleotides 17-25 nucleotides in length, 37°C was

often used). The annealed primer and template were then placed on ice while the reagents necessary for chain extension by T4 or T7 DNA polymerase and ligation by T4 DNA ligase were added. For T4 DNA polymerase, the reaction cocktail was 0.4 mM in each dNTP, 0.75 mM in rATP, 20 mM Tris pH 7.4, 5 mM $MgCl_2$, 1.5 mM DTT, 3-10 U of T4 DNA polymerase and 5U T4 DNA ligase in 20 µl (McClary, Witney et al. 1989). For T7 DNA polymerase (form II[12]), the reaction cocktail was 0.15 mM each dNTP, 0.75 mM rATP, 50 mM Tris pH 7.4, 10 mM $MgCl_2$, 5 mM DTT, 2.5 U of T7 DNA polymerase (Engler, Lechner et al. 1983) and 5 U T4 DNA ligase. Chain extension and ligation were accomplished in a 60-90 min incubation at 37°C. Of the 20 µl reaction, 2 µl were transformed into $CaCl_2$ competent (Maniatis, Fritsch et al. 1982) *dut+ ung+* cells; usually LE101.

Single Substrate Gel Overlay Activity Assay

The cell pellet from 1.5 mls of a saturated LE101 culture was resuspended in 20 µl of 25% sucrose, 20mM EDTA and lysozyme was added to 2 mg/ml. The suspension was incubated at room temperature for 30 min and spun at 14,000 g for 5 min. The pellet was discarded leaving the periplasmic extract. An equal volume of 2x Laemmeli buffer lacking 2-mercaptoethanol was added to the extract and the mixture was heated for 3 min at 65°C to break up protein aggregates.

---

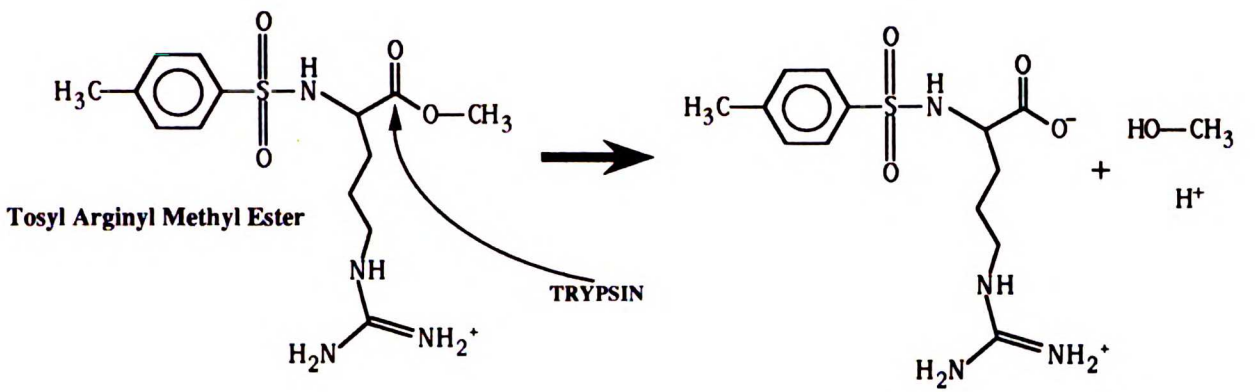
[12]Form II is purified such that the strand displacement activity of the T7 polymerase is inactivated (Engler et al, 1983).

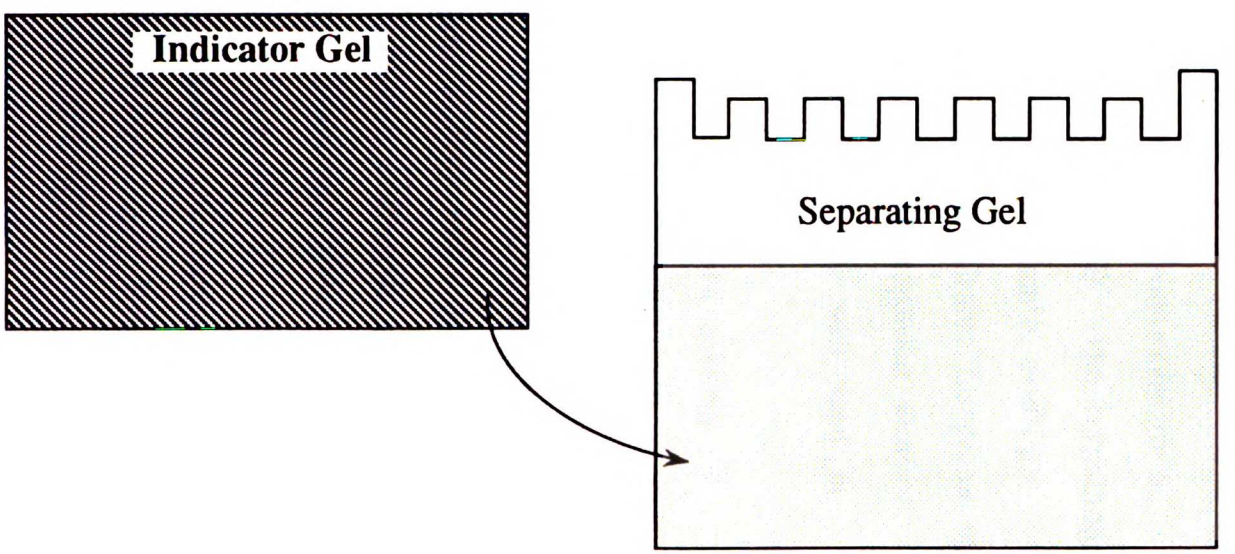
The proteins were separated by SDS-PAGE. The separating gel was soaked in 2.5% Triton X-100 for 35 min, 5 mM $CaCl_2$ for 7 minutes, 5 mM $CaCl_2$ for 1 min, and then overlaid with an 8% polyacrylamide gel containing 65 mM TAME, 1 mM phenol red (4.3 ml 29.2% acrylamide:0.8% bis, 5 mls 200 mM TAME, 6 ml 2 mM phenol red, 0.08 ml of 1 M NaOH, 0.125 ml 10% ammonium persulfate, .0135 ml TEMED). The final pH of the gel solution is 9.3 where phenol red is purple. The sandwich was sealed with Saran Wrap and incubated at 37°C for 0.5-3 hours or overnight at room temperature. Acetyl-phenylalanyl methyl ester, benzyl-alanyl methyl ester, and acetyl-$(Ala)_3$ methyl ester were added in 5 ml of 10% DMF, 50% DMF, or water respectively; 1 M NaOH was added to pH 9.3 (Figure 1-1).

Commercially available bovine trypsin, bovine α-chymotrypsin A and porcine pancreatic elastase were made as 1 mg/ml solutions in 1 mM HCl and serially diluted for examination of the sensitivity of the overlay gel assay.

Figure 1-1. Single Substrate Overlay Gel Activity Analysis

The reaction depicted is the cleavage of the ester substrate TAME by trypsin to its acid and alcohol constituents. This substrate is present in the indicator gel which is overlaid on the separating gel after SDS-PAGE. Phenol red, present in the indicator gel, records the drop in pH as the acid is generated. The pKa of methanol is 15.3 and so is present as the alcohol while the acid is deprotonated (pKa approximately 3).


Tosyl Arginyl Methyl Ester
TRYPSIN
HO—CH3
H+


Indicator Gel
Separating Gel

RESULTS

Expression Plasmids

The first step in establishing a system for the genetic manipulation of trypsin was the isolation and expression of the coding region for the enzyme. A complete sequence for rat anionic trypsin II was obtained by fusing a genomic clone (Craik, Choo, et al. 1984) comprising the signal peptide and the first exon to a partial cDNA clone (MacDonald, Stary et al. 1982). The trypsinogen minigene which comprised the entire pretrypsinogen coding sequences, 55 bp of the 3' untranslated region (Craik, Largman et al. 1985), and a 300 bp Pst I- Hinc II pBR322 fragment was assembled in the vector mp8-Tg (Figure 1-2) that was used in subsequent constructions.

A transition vector from which the trypsinogen coding sequences could be readily transferred was generated via a three component ligation reaction. The resultant vector, pGAP-Tg, was a product of the Eco RI-Sal I trypsin fragment from mp8-Tg, the Sal I-Nco I vector fragment and a synthetic Nco I-Eco RI 67 bp adaptor (Figure 1-2). This adaptor encoded Nco I, Bam H1, and Hind III restriction sites and the trypsinogen sequences from the signal peptide to the Eco R1 site.

A bacterial expression vector, pTRAP, was created by splicing the Bam HI-Sal I trypsinogen DNA fragment into Bam HI-Sal I digested pFOG402 (Inouye, Barnes et al. 1982). In this configuration, trypsinogen expression is regulated by the *phoA* promoter; the *phoA* signal peptide directs the secretion of the enzyme (Figure 1-2). We presumed that a secretion system would

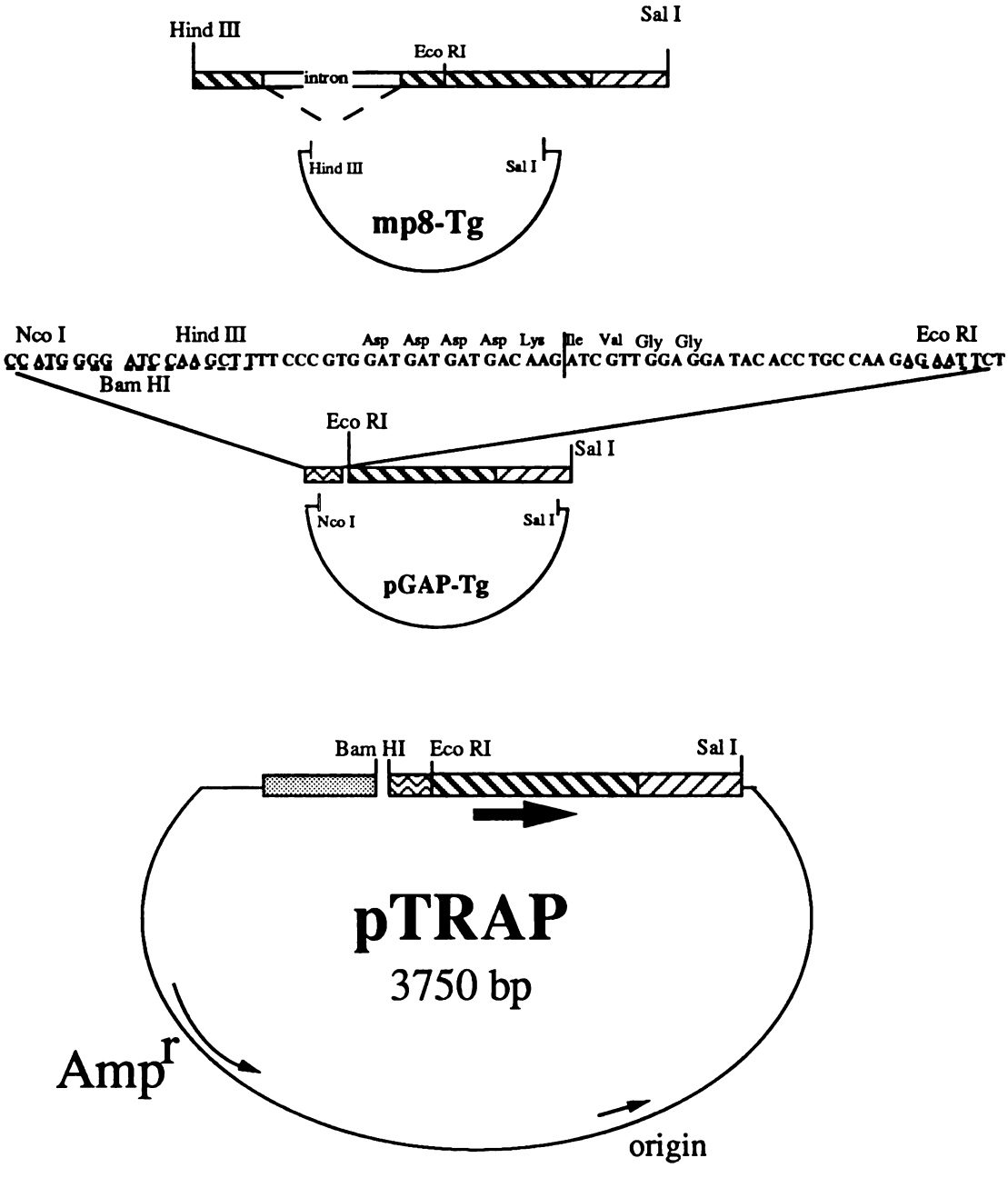
Figure 1-2. A Trypsinogen Expression Vector

The trypsinogen coding sequences are denoted by the heavy slanted lines. The Eco RI site within trypsin is marked in all instances as a landmark. The thin slanted lines denote sequences from pBR322 which were carried into the M13mp8 vector in order to take advantage of compatible restriction sites. The intervening sequence or intron is denoted by the empty box and is labeled "intron". Wavy lines denote sequences introduced using synthetic oligonucleotides. The stipled box denotes sequences encoding the alkaline phosphatase promoter and signal peptide. Direction of transcription of the ß-lactamase gene which provides ampicillin resistance (Amp^r) is indicated; the orientation of the origin of replication is indicated. The direction of transcription is the same in all vectors and is indicated in pTRAP by a heavy arrow. The intron initially introduced into this construction with the genomic sequences was subsequently deleted by oligonucleotide directed mutagenesis to finish mp8-Tg.

pGAP-Tg is a pBR322 based vector. The DNA sequence including the 67 bases of the Nco I-Eco RI adaptor is displayed above the vector. The bases encoding the Nco I restriction site are singly underlined; the bases of the Bam HI site are doubly underlined; the bases encoding the Hind III site are singly underlined. The trypsinogen protein sequence adjacent to the enterokinase cleavage site $(Asp)_4$-Lys is identified and the processing site is marked by the vertical line. Ile-Val-Gly-Gly is the start of mature trypsin.

pTRAP is a derivative of pFOG-402 (Shortle 1983) which uses the alkaline phosphatase promoter and signal peptide (Inouye,

Barnes et al. 1982) to regulate staphyloccocal nuclease expression. It is a pBR322 derivative with the alkaline phosphatase (*phoA*) promoter and signal peptide extending from the remnants of the Eco RV restriction site.


Hind III
Sal I
Eco RI
intron
Hind III
Sal I
mp8-Tg
Nco I
Hind III
Asp Asp Asp Asp Lys Ile Val Gly Gly
Eco RI
CC ATG GGG ATC CAA GCT TTT CCC GTG GAT GAT GAT GAC AAG ATC GTT GGA GGA TAC ACC TGC CAA GAG AAT TCT
Bam HI
Eco RI
Sal I
Nco I
Sal I
pGAP-Tg
Bam HI
Eco RI
Sal I
pTRAP
3750 bp
Amp r
origin

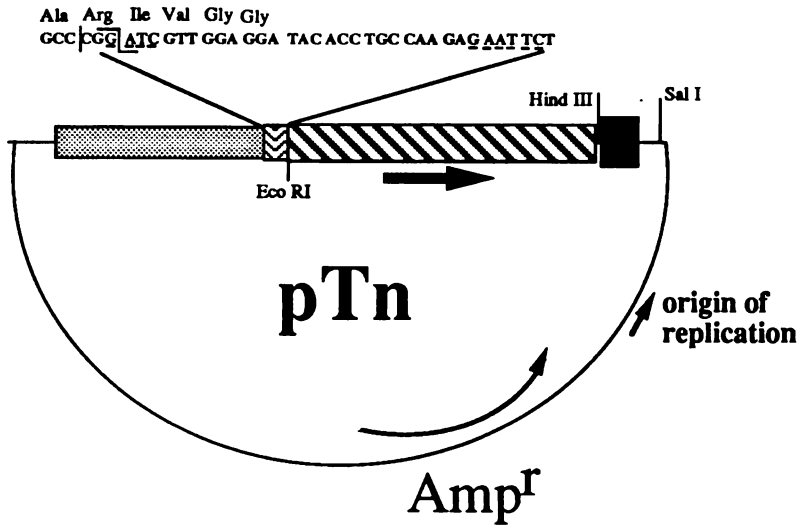
be the most favorable one for the expression of a mammalian protease containing disulfide bonds as the reducing environment of the cytoplasm would prevent correct disulfide formation while degradative intracellular proteases would have ample access to heterologous trypsin localized to the cytoplasm. Also, active trypsin formed inside the cell might be deleterious to the host. As the periplasmic contents comprise approximately 1% of the total cell protein, harvesting the periplasm provides a 100 fold purification over a total cell extract. The *phoA* promoter is induced in low phosphate and in strains lacking a functional repressor (i.e. SM138) which is a product of the *phoR* gene (Inouye, Barnes et al. 1982). The zymogen can be activated using enterokinase. Pure trypsin was isolated from this system for kinetic characterization of trypsin D189K (Gráf, Craik et al. 1987) and trypsin S195C (Higaki, Gibson et al., 1987).

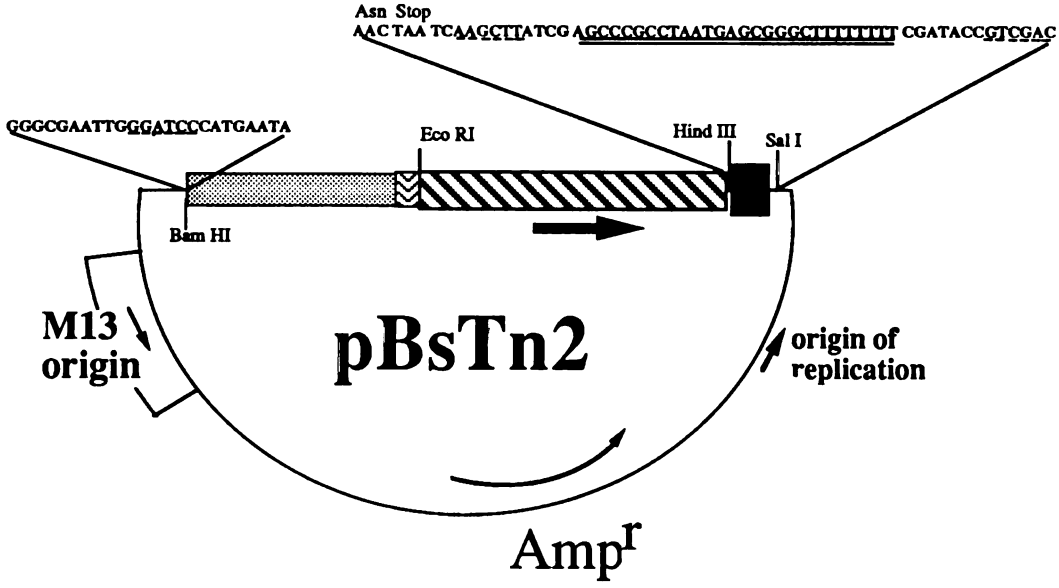
To eliminate the need for zymogen activation, the pTn vector was created to direct secretion of mature trypsin. In this system, direct screening of cultures for trypsin activity would be possible. The first codon of mature trypsin was abutted to the prepeptide sequence of alkaline phosphatase by replacing the Bam HI-Eco RI trypsinogen fragment with a synthetic 28 bp Sau3A I-Eco RI adaptor (Figure 1-3). Neither constitutive nor regulated secretion of the mature enzyme seemed to affect the viability of the *E.coli* at the ~0.2 mg/L level of expression obtained in this system. Protection against the action of a protease in the periplasm may be afforded

Figure 1-3. Trypsin Expression Vectors

The boxes with stipled and slanted lines denote alkaline phosphatase, and trypsin sequences as in Figure 1-2. The *trpA* transcription terminator is denoted by the black box. Arrows indicate orientation of the ß-lactamase gene and the origins of replication. pTn is a pBR322 based vector; pBsTn2 is a pUC derivative and as such is present at higher copy number than a pBR322 based vector would be.

The expanded sequence in the pTn depiction shows the junction between the *phoA* signal peptide and the mature trypsin sequence. The vertical line indicates the site of signal peptidase cleavage while the "Z" line indicates the processing necessary to create mature trypsin. The Sau3A I and Eco RI sites are underlined. The expanded sequences in pBsTn2 illustration show the 5' junction between pBS and the alkaline phosphatase-trypsin expression cassette and the 3' junction between pBS and the expression cassette. At the 5' end the Bam HI site is underlined. At the 3' end, the Hind III and Sal I restriction sites are singly underlined while the terminator sequence is doubly underlined.


Ala Arg Ile Val Gly Gly
GCC CGG ATC GTT GGA GGA TAC ACC TGC CAA GAG AAT TCT
Hind III
Sal I
Eco RI
pTn
origin of replication
Amp^r


Asn Stop
AAC TAA TCAAGCTTATCG AGCCCGCCTAATGAGCGGGCTTTTTTTT CGATACCGTCGAC
GGGCGAATTGGGATCCCATGAATA
Eco RI
Hind III
Sal I
Bam HI
M13 origin
pBsTn2
origin of replication
Amp^r

by the separation of the membranes into domains (Clark, Mel et al. 1987; MacGregor and Hunt 1986)[13].

As with pTRAP, pTn directs the expression of trypsin plus a leader. Signal peptidase cleaves the signal peptide C-terminal to an alanine at position (-1) leaving an arginine at the amino terminus (Inouye, Barnes et al. 1982) (Figure 1-3). Apparently, the amino terminal arginine residue is removed. This secondary processing is analogous to the physiological activation of trypsin; the loss of the arginine residue mimics the loss of the propeptide. The aminopeptidase IAP is responsible for the amino terminal differences among the isozymes of alkaline phosphatase (Nakata, Shinagawa et al. 1982) and may be the enzyme involved in this processing. Evidence for the creation of correctly folded trypsin is provided by the activity of the trypsin which has been demonstrated on an arginine ester substrate (Figure 1-4) and by the ability of this trypsin to bind pancreatic trypsin inhibitor (Higaki, Gibson et al. 1987). Subsequent studies in which trypsin has been purified to homogeneity have detected a single active form (Vásquez and Craik 1990).

The trypsin is periplasmic. The processing associated with membrane translocation occurs, and generation of spheroplasts from intact trypsin-producing cells releases the enzyme into the supernatant. Some of the trypsin may remain in the cytoplasm or attached to the membrane. However, experiments with pTRAP

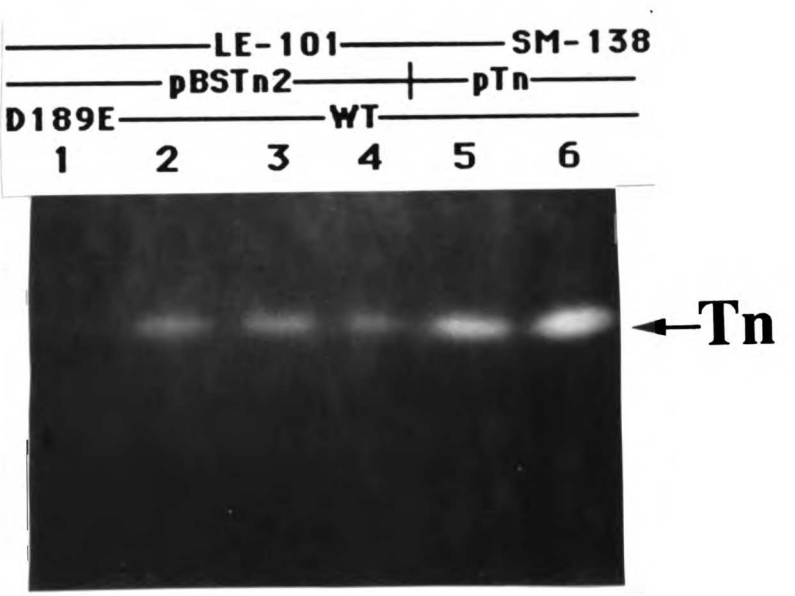
---

[13] The segmentation of lipid bilayer is a fairly novel concept that has not been generally accepted. It is directly opposed to the conventional view of the lipid bilayer as providing for unlimited and rapid exchange in two dimensions (Singer and Nicolson 1972).

Figure 1-4. Levels of Trypsin Expression

Black and white high contrast photograph of an activity gel. White on the black background corresponds to yellow on purple background. The only yellow band on the gel is the one representing the trypsin activity on this arginine ester substrate.

Samples were obtained from 1.5 ml of saturated culture. Both LE101 and SM138 are *phoR* strains and so produce trypsin constitutively from the *phoA* promoter. The band in lane 1 represents the activity from a culture of strain LE101 carrying the vector pBsTn2 with the mutant trypsin D189E. Lanes 2, 3 and 4 contain extract from LE101 pBsTn2 with wild type (WT) trypsin. Lane 5 represents the activity from strain LE101 carrying pTn. Lane 6 represents the activity present in strain SM138 carrying pTn. Vector pTn2 was created by replacing the pUC backbone sequences of pBsTn2 with those form pBR322. Expression of trypsin form this vector in SM138 cells appeared to be stable in contrast to pBsTn2. Also, pTn2 produced roughly two fold the active trypsin produced by pTn in SM138 and so was the superior vector in this series (Vásquez, Evnin et al. 1989).


LE-101
SM-138
pBSTn2
pTn
D189E
WT
1
2
3
4
5
6
Tn

indicate that greater than 90% of the total trypsinogen can be recovered by harvesting the contents of the periplasm (Vásquez, Evnin et al. 1989)[14].

A vector for mutagenizing, sequencing, and expressing trypsin was created in two steps by subcloning a blunt-ended Cla I-Sal I fragment encoding the alkaline phosphatase-trypsin fusion into Sma I-Sac I digested pUC18. The Cla I end was blunt ended with T4 DNA polymerase. A vector containing both an M13 origin and a high copy ColEl origin (pUC) was created by subcloning a blunt ended Bam HI-Sac I DNA fragment containing the alkaline phosphatase-trypsin fusion into Sac I- Eco RI digested pBS. The resultant vector, pBsTn has eliminated all the 3' untranslated DNA and the 300 bp of pBR322 sequence that had been carried into pTn (Figure 1-2, 1-3). Sequence analysis of the vector revealed that both the 5' and 3' ends of the alkaline phosphatase-trypsin fusion had been altered; 17 bases had been deleted from the 5' terminus and one from the 3' terminus. Fortunately, the 5' sequences lost were beyond the *phoA* promoter sequences and the stop codon was

---

[14]The hallmark of the activation of trypsin is the freeing of isoleucine at position 16 as the amino terminus. The activation has been analyzed using the X-ray crystallographic solutions of the static three dimensional structures of trypsin and trypsinogen. It was suggested that the salt bridge that forms between the new amino terminus and the carboxylate of Asp194 drives the conformational changes that are observed in the substrate binding pocket. This latter reconfiguration allows productive binding of the substrates. Recent experiments in which asparagine was substituted for the aspartic acid at position 194 indicate that the exact fit of the Ile16 (and Val17) into a hydrophobic cleft in the enzyme (which is punctuated by the Asp194 carboxylate) provides the energy to drive the conformational changes. Salt bridge formation may play only a supporting role (J. Vásquez, personal communication, 1990).

preserved at the 3' end although the last base was changed (TAG to TAA).

We were concerned that a relatively strong promoter on a high copy number plasmid would be deleterious to the host. To minimize this effect, the strong, rho independent *trpA* transcription terminator (Christie, Farnham et al. 1981) was subcloned into a blunt-ended Cla I site of pBsTn 12 bp downstream from the end of the trypsin coding sequences. This construct was designated pBsTn2 (Figure 1-3). Dideoxy sequence analysis confirmed the insertion and identified a clone with the desired orientation.

Mutagenesis

Our approach towards understanding the dynamics of trypsin-substrate interactions relies on generating many mutants in the substrate binding pocket. Anticipating this need, a system to mutate and express without subcloning was developed. The procedure employs a strain of *E. coli* deficient in the enzymes dUTP nucleotidohydrolase (dUTPase) encoded by *dut* and uracil N-glycosidase (or glycosylase) encoded by *ung* (Kunkel 1985; Kunkel, Roberts et al. 1987). A *dut ung E. coli* host is transformed with the desired substrate for mutagenesis, in our case, pBsTn2. The resultant ampicillin resistant cells are grown in liquid media until the culture attains logarithmic growth and then infected with defective F1 phage. These helper virus code for the production of replication and packaging proteins but are themselves inefficiently replicated. However, pBsTn2 with its wild type M13 origin is efficiently replicated and efficiently packaged. Consequently, at the

harvesting of the virus pBsTn2 outnumber VCS M13 (Messing 1983) or R408 (Russell, Kidd et al., 1986) virus by roughly 25:1 and 5:1 respectively. The DNA, readily obtained by phenol extraction, contains uracil in place of approximately 14% of the thymine bases (Warner, Duncan et al. 1981).

The uracil-laden DNA is annealed to a mutagenic primer which serves as a starting point for the chain elongation by T4 or T7 DNA polymerase (form II). These two polymerases are well suited for mutagenesis as they are highly processive and will not strand displace (Nossal 1974; Lechner, Engler et al. 1983). The presence of uracil in the parental strand of the heteroduplex allows strong selection against this strand in the mismatch repair that occurs prior to replication of the transformed DNA (Kunkel 1985). Although some mismatches are better recognized and repaired than others (Kramer, Kramer et al. 1984; Dohet, Wagner et al. 1985), most colonies (typically 70%) harbor only site-specifically mutated plasmids. Some colonies do contain both mutant and wild type plasmids which results when replication occurs prior to repair. Purely wild type colonies (background) presumably arise from "endogenous" priming by random contaminating RNA and DNA and not because the newly synthesized strand is repaired instead of the parental strand. Support for this claim comes from experiments in which the amount of contaminating DNA and RNA is dramatically reduced by incubation of the intact virions with RNase and DNase. After this incubation, the number of wild type colonies recovered is markedly reduced.

With the efficiencies obtained, screening for mutants is not required. From the colonies obtained upon transformation of the mutagenesis reaction into competent *E. coli* cells, DNA is prepared and subjected to DNA sequence analysis. As the procedure is oligonucleotide-directed, very few spurious or unwanted alterations in the flanking sequences are observed[15]. This method is superior to cassette mutagenesis (Lo, Jones et al. 1984) for generating libraries of random mutations since unique and proximal restriction sites are necessary for the cassette. Subcloning of the mutated sequences is not required unlike M13 based mutagenesis stratagies. Also, only one primer is required as the DNA polymerases exhibit no strand displacement activity and the presence of T4 DNA ligase allows the formation of closed covalent circles protecting the ends of the synthesized strand from cellular exonucleases.

Construction of Position 189 Library

To create a bank of mutants from which the variants at position189 could be isolated, we chose to make two separate libraries each containing half the total pool. In principal, this route avoids preferential hybridization which can occur when a subset of the mutagenic oligonucleotides in the pool has a better match to the template sequence than other oligonucleotides in the pool. As preferential hybridization can increase the representation of a small number of the mutants in the library, the number of

---

[15] The presence of uracil in the parental strand increases the rate of spontaneous mutation approximately two fold (Kunkel, 1985).

sequencing reactions required for the isolation of the set of mutants is increased.

Specifically, two mutants D189S (GAT to AGT) and D189E (GAT to GAA), generated in other experiments, served as templates. The total pool of oligonucleotides necessary to code for all the amino acids and a stop codon at a particular position is x-N-N-G/C-x (where x denotes flanking sequences and N denotes all four nucleotides). The total set was synthesized in two parts: Pool 1 contained x-N $^{A}/_{T}$ $^{C}/_{G}$ -x and pool 2 contained x-N $^{C}/_{G}$ $^{C}/_{G}$-x. Consequently, the best possible match for pool 1 to the D189S template and for pool 2 to the D189E template was only one of three bases in the codon for position 189 (Table 1-1).

Mutagenesis with the pool 1 oligonucleotides recovered 8 of the 13 possible enzymes; mutagenesis with the pool 2 oligonucleotides yielded 8 of 8 possible mutants. From the library generated by priming of the D189S template with pool 1, 57 members were picked and the sequence of the trypsin gene around position 189 was determined. As Table 1-1 enumerates, 40 of these were mutants. The remainder were -TCT- (coding strand) with a single exception which contained a change in the codon 5' to position 189. From the pool 2 library, 16 members were sequenced and all were mutants. Thus, the frequency of recovering mutants from the pool 1 library was 70% and from the pool 2 library was 100%. The recovery of mutants from pool 2 was anomalously high and may reflect the relatively small sample size.

Table 1-1. Identification of Mutants in the Position 189 Library

This table depicts the results obtained from sequencing isolates from the two halves of the position 189 library. In the first, leftmost, column of Table 1, the sequences at position 189 (coding strand) of the template are shown (5' to 3') and labeled according to the amino acid that is encoded by that particular codon. The strand isolated was the non-coding strand. In the second column, the pool of mutagenic oligonucleotides is shown 5' to 3'. The total pool of oligonucleotides at position 189 is x-NN$^{C}/_{G}$-x where "x" indicates identity with the coding strand. Pool 1 has $^{A}/_{T}$ at the second position and pool 2 has $^{C}/_{G}$ at that position. The third column lists by codon the amino acids that could be generated from the pool of oligonucleotides. The last column indicates the number of independent isolations of each amino acid; in parentheses are tabulated the number of times a particular codon was identified if multiple codons are possible for an amino acid.

In the first column, the sequences at position 189 of the template are shown (5' to 3') and labeled according to the amino acid that is encoded by that particular codon. In the second column, the pool of oligonucleotides that was annealed to the template (column 1) is shown 5' to 3' (N denotes all four bases and x denotes the flanking, complementary nucleotides present in the oligonucleotides). The third column lists the amino acids that could be generated from the pool of oligonucleotides (column 2) and the respective codon(s). The last column shows the number of independent isolations of each amino acid and the number of isolations of each codon is tabulated where applicable.

Template at Position 189	Mutagenic Primer	Amino Acids and Codons	Number of Isolations
Asp189Ser x-TCT-x	Pool 1 x-N(A/T)(C/G)-x	Phe: TTC	16
		Leu: TTG, CTC, CTG	10 (5,2,3)
		Ile: ATC	2
		Met: ATG	6
		Val: GTC, GTG	2 (0,2)
		Tyr: TAC	0
		stop: TAG	2
		His: CAC	0
		Gln: CAG	1
		Asn: AAC	1
		Lys: AAG	0
		Asp: GAC	0
		Glu: GAG	0
Asp189Glu x-GAA-x	Pool 2 x-N(C/G)(C/G)-x	Ser: TCC, TCG, AGC	2 (0,0,2)
		Pro: CCC, CCG	2 (0,2)
		Thr: ACC, ACG	3 (1,2)
		Ala: GCC, GCG	1 (1,0)
		Cys: TGC	3
		Trp: TGG	1
		Arg: CGC, CGG, AGG	4 (1,3,0)
		Gly: GGC, GGG	1 (1,0)

Screens for Trypsin Activity

To expedite the search for especially informative mutants, an assay was developed allowing detection of a specific esterase activity of an enzyme in a mixture of proteins. A protein extract, the periplamic fraction, is fractionated by SDS-PAGE. The separating gel is treated to remove Tris buffer and SDS detergent and it is overlaid by a polyacrylamide gel containing the substrate TAME and the pH indicator phenol red (Figure 1-1, 1-4, 1-5). Trypsin activity can be identified in the presence of other arginine specific esterases because of the fractionation. However, the *E. coli* esterases are not stable in 0.1% SDS unlike trypsin and no additional bands are observed in the overlay gel (Figure 1- 4)[16]. The TAME substrate is used at high concentration because it potentiates the activity of enzyme as much as six fold, presumably by interaction distant from the active site (Trowbridge, Krehbiel et al. 1963). Hydrolysis of TAME generates acid (Nα-*p*-tosyl-arginine) and methanol moieties (Figure 1-1) and, so, in the region of the overlay gel adjacent to the trypsin, the pH falls. This drop is recorded by the phenol red as it changes from purple to yellow (Bryan, Rollence et al. 1986).

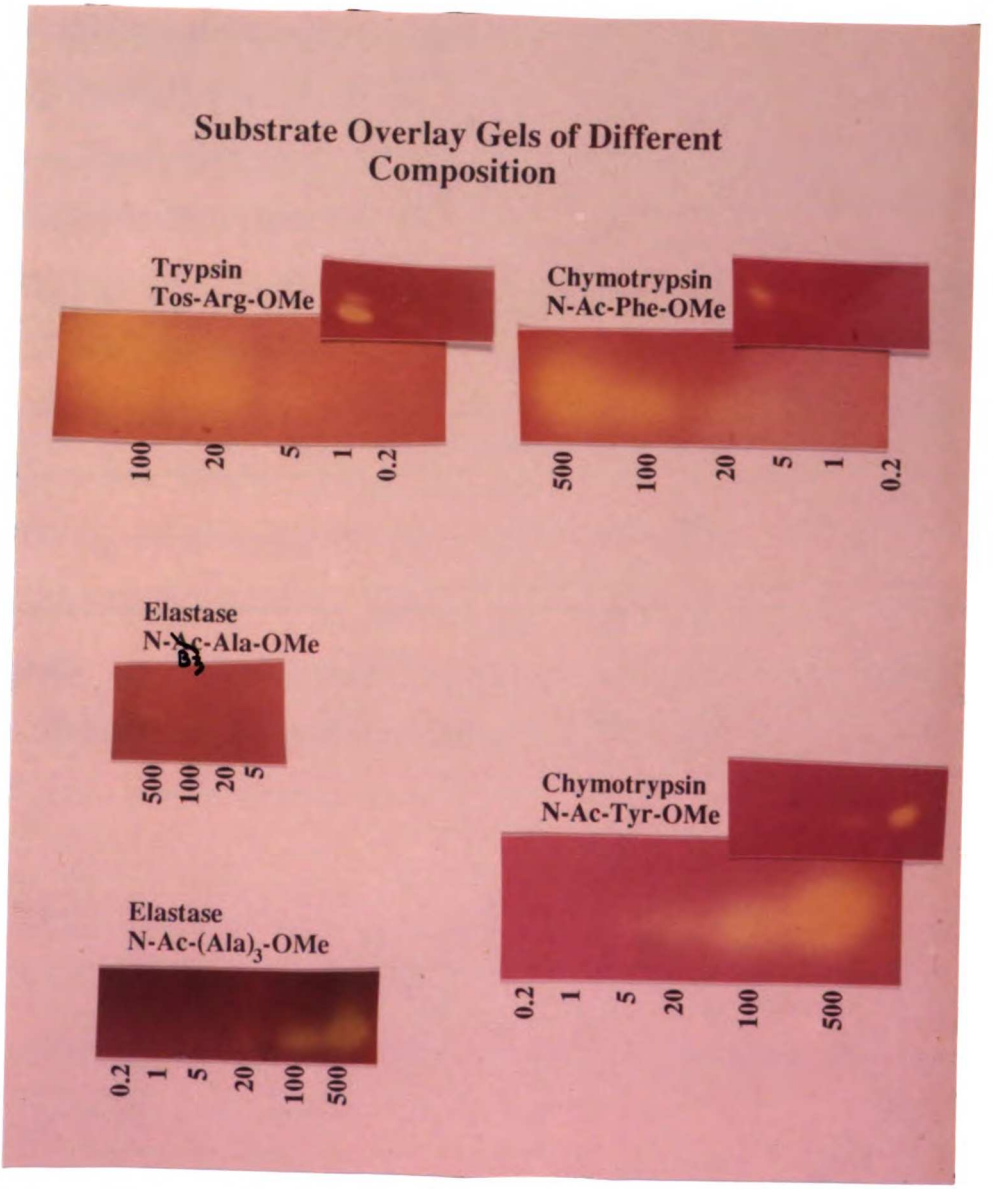
Figure 1-5 shows the limit of detection of trypsin using a dilution series commercially available bovine trypsin. The

---

[16]S. Willet in this laboratory has performed experiments in which the activity of the total cell extract as well as the periplasmc extract was examined for trypsin acitivity. The activity gel assay could specifically detect the trypsin activity in total cell extracts. However, all the trysin activity produced was present in the periplasm (personal communication, 1990).

Figure 1-5. Single Substrate Overlay Gels of Different Composition

The enzyme tested is listed along with the ester substrate above each overlay activity gel. The yellow band on the purple background is due to the action of the protease. In the smaller inset color picture of the arginine and tyrosine gels, a picture of the gel after 45 min (37°C) and the larger pictures and the pictures of the alanine ester gels are shown after incubation for 2.5 hr. TAME was present at 65 mM, Ac-Tyr methyl ester, Bz-Ala methyl ester, and Ac-$(Ala)_3$ methyl ester were present at 50 mM, 100 mM, and 50 mM respectively. The overlay gels contained DMF which was a necessary solvent at 0%, 2%, 2.6%, and 0% respectively. The total amount of enzyme in nanograms is listed below each lane. The large band in the inset is the highest concentration used and the inset is in the orientation as the larger picture of the gel.


Substrate Overlay Gels of Different Composition
Trypsin
Tos-Arg-OMe
100
20
5
1
0.2
Chymotrypsin
N-Ac-Phe-OMe
500
100
20
5
1
0.2
Elastase
N-Ac-Ala-OMe
500
100
20
5
Chymotrypsin
N-Ac-Tyr-OMe
0.2
1
5
20
100
500
Elastase
N-Ac-(Ala)3-OMe
0.2
1
5
20
100
500

detection limit extends with time, and is approximately 0.5 ng. In these experiments, trypsin runs as a doublet (the upper band may be an intermolecular dimer, perhaps formed between cysteines at position 191).

The ability to detect the specific hydrolysis of other proteases was also investigated. A dilution series of chymotrypsin and elastase were subjected to SDS-PAGE. The separating gels were treated with in the same manner as the TAME gel and the chymotrypsin gels were overlaid with an acrylamide gel containing either 50 mM Ac-Tyr-OEt or Ac-Phe-OMe. The limit of detection in both enzyme assays is approximately 0.5 ng (Figure 1-5). Elastase activity on the single residue Bz-Ala-OMe substrate was detected with 500 ng of protein and was detected at 0.5 ng with the tripeptide Ac-$(Ala)_3$-OMe illustrating the dependence of the activity of elastase on additional subsites (for a summary see Creighton 1983). Both elastase and chymotrypsin are visualized as a single band.

The trypsin expression levels of the pTn and pBsTn2 plasmids were compared by activity gel analysis (Figure 1-4) (Vásquez, Evnin et al. 1989). In LE101, the lower copy number pTn vector generates approximately twice as much active trypsin as pBsTn2. The presumed increased in trypsin RNA levels from the pUC-based vector does not compensate for other limitations of the pBsTn2 system. In particular, highly variable trypsin expression levels from SM138 pBsTn2 were observed, and occasionally, no trypsin was obtained. Other work has indicated that in strain SM138 among other *phoR* strains (do not make repressor), that additional regulation of *phoA* (alkaline phosphatase) expression can result in variable rather than constitutive expression (Wanner, Wilmes et al. 1988). However, these conclusions may not be relevant to the problems encountered in the trypsin expression since the DNA site that binds the additional regulatory protein (*phoM*) may not be present in the DNA fragment of the phoA promoter that is present in the pTn constructions[17].

Other Screens

A way to screen for trypsin directly on plates was pursued. The basis for the single substrate overlay gels proved ineffective as endogenous esterase activity masked the trypsin activity. The periplasmic proteases in *E. coli* are reported to be zinc requiring. Thus, a metal chelator such 1,10 (or *o*)-phenanthroline would

---

[17]Alkaline phosphatase (phoA gene product) production is followed by the hydrolysis of X-P (5-bromo-4-chloro-3-indoyl-phosphate-*p*-toluidine). When repressor is not produced, *phoA* is induced, and hydolysis of the X-P follows which renders the cells blue.

presumably inhibit their activity. However, addition of this reagent to plates containing TAME and phenol red did not substantially enhance the discrimination between active and inactive trypsins. The known soluble proteases in the periplasm include Do, Re, Mi and Pi. Do, Re and Mi are casein- and globin-degrading enzymes and Do and Re are present in both the periplasm and cytoplasm. Pi is also present in both subcellular fractions but it degrades insulin and not globin or casein. Pi is identical to protease III (Cheng and Zipser 1979).

Other efforts focused on the precipitation and associated color change caused by the reaction between fast garnet red and free ß-naphthylamide. Although this reaction works well for the detection of the peptidase N (Miller and MacKinnon 1974), it was ineffective for trypsin screening. The failure may be due to differences in localization of trypsin which is periplasmic, and peptidase N which is cytoplasmic.

Efforts were made to detect the release of *p*-nitroaniline or 4-methyl-7-amino coumarin from trypsin substrates (Slots 1987). In these instances, the expense of the substrates dictated that the substrate be sprayed on after cell growth rather than being incorporated into the solid growth media. Alternatively, the substrate was incorporated into enzyme-directed overlay membranes which are created by dipping cellulose diacetate paper in the appropriate substrate present at 20 mM in DMF. However, in all instances, the hydrolysis observed in untransformed *E. coli* rendered these assay unreliable.

Various efforts were made to utilize the detection system of the activity gel directly on the colonies. The colonies were grown on nirocellulose which was then transferred both face up and face down, both with and without lysis of the cells, onto a gel containing TAME and phenol red at pH 9.0. Alternatively, the nitrocellulose was dipped in a cocktail containing the TAME and the phenol red at pH 9.0. In all cases, the sensitivity was too low and the background activity due to endogenous esterases was too high to get reliable results. An attempt was also made to utilize a clearing assay on rich plates with 0.5% skim milk protein embedded (Silen, Frank, et al. 1989)[18]. Again, the sensitivity of the assay was too low and background was too high. The development of a plate assay remains a goal of future research.

[18]J. Silen reported that this assay works well for α-lytic protease, but the assay must be conducted at room temperature over 5 days. These assays were done at 37°C over 24-36 hours.

CONCLUSIONS

The full length coding sequence for rat trypsin has been assembled and placed in a bacterial expression vector that codes for the secretion of active trypsin. A streamlined site-directed oligonucleotide-mediated approach has been established for generating both single mutations and libraries of trypsin mutations. As a prelude to a detailed investigation of the substrate binding pocket of trypsin, a library of mutants at position 189 was generated. We attempted to prevent any bias in the representation of the library but were only partially successful. Nonetheless, 15 of 19 possible mutants were isolated from the libraries. Apparently, purine-purine mismatches are far more deleterious than pyrimidine-purine mismatches. Only 7.5% of the 40 mutants identified in pool 1 have A in the second position which must pair with G in the template. Among the codons that do have a T in the second position, it was again favorable to avoid a purine-purine mismatch in the first position as 72% have a pyrimidine (58% have T which forms a Watson-Crick pair). The noxious effect of the purine-purine pair is exemplified in the absence of Tyr mutants (TAC). The first position provides the canonical T::A match but the second position introduces a purine-purine (A::G) pair. More specifically, the purine-pyrimidine pair is G::T whch is an energetically favorable pairing although far less so than the standard Watson-Crick base pairs and is observed in RNA hairpins and in the wobble base pair interactions of codons and anticodons (Saenger 1984). In pool 2, where all three bases in the template (non-coding strand) were pyrimidines, these strong preferences

were not observed. However, it is notable that a canonical base pair in the first position did not enhance the recovery of those mutants. For example, Gly (GGN) and Ala (GCN) were not recovered more efficiently than the other possibilities.

The skewing problem might be avoided by additional subdivision of the pool of oligonucleotides necessary for complete representation. Thus, there might be a separate pool for the x-$NT^{C}/_{G}$-x oligoncleotides. Alternatively, the entire region to be mutagenized could be deleted in a single round of mutagenesis prior to the library mutagenesis reaction. No mismatches between the template and the primer would remain.

A rapid screening assay for extracts containing trypsin (or other proteases) was developed that relied on the acidity of the products of the hydrolysis of an ester substrate. In this assay, an overlay gel containing a pH indicator is used to monitor the pH change[19]. The limit of detection for trypsin on the arginine substrate (and for chymotrypsin with Phe or Tyr and for elastase with Ala-Ala-Ala) was approximately 0.5 ng. All of the enzymes contain six disulfides which helps retain the enzyme structure in the presence of detergent. The precise alignment of the substrate is achieved by providing sufficient chemical contacts with the enzyme. For the deep pockets of chymotrypsin and trypsin, the single side chain is sufficient; for (porcine pancreatic) elastase which prefers small hydrophobic substrates, additional contacts

[19]In an analogous fashion, a pH-stat records the addition of base necessary to maintain a given pH in a reaction involving tryptic hydrolysis of an ester substrate.

provided by an extended substrate are necessary to achieve efficient catalysis (Thompson and Blount 1970). These results are consistent with previous spectrophotometric assays involving these enzymes and these ester substrates. Trypsin hydrolysis of TAME at pH 8.0 and 25°C was characterized by a kcat=60 sec^{-1} and a Km=12.5 μM[20] (Trowbridge, Krehbiel et al. 1963). Chymotrypsin exhibited a kcat=192 sec^{-1} on Ac-Tyr-OMe with Km=0.66 mM at pH 8.0 and 25°C and exhibited a kcat=97.1 sec^{-1} with a Km=0.93 mM on Ac-Phe-OMe under identical conditions (Fersht 1985b). Elastase hydrolyzes Ac-Ala-OMe with kcat= 8.2 sec^{-1} and Km= 170 mM (Thompson and Blount 1970) but hydrolyzes the Ac-$(Ala)_3$-OMe substrate with a kcat=120 sec^{-1} and Km= 0.4 mM (Thompson and Blount 1970) at pH 9.0, 37°C such that it has a similar kcat/Km to the other enzymes on their single residue substrates. The sensitivity is sufficient to allow detection of mutants with 1 x 10^{-4} the activity of trypsin assuming that the mutant is expressed to similar levels as trypsin. Additional sensitivity may be possible using different substrates such as dipeptide or tripeptide esters or with different indicators such metacresol purple (*m*-cresolsulfonephthalein) or 4,4'-bis (4-amino-1-naphtholazo) 2,2' stilbene disulfonic acid. These chemicals exhibit complete color changes in a narrower pH range than phenol red (Handbook of Chemistry and Physics 1978). There is one fluorescent indicator,

---

[20] However, there is a secondary binding site for the TAME which has an apparent Km= 53 mM. The rat of hydrolysis increases to 340 sec^{-1} with occupancy of this second site. It is unclear which rate relates to the conditions in the overlay gel, but the turnover number is likely to be intermediate between the 60 and 340 sec^{-1} as the initial substrate concentration is 65 mM (Trowbridge et al 1963).

coumaric acid, that might be useful as it changes from green at pH 9.0 to colorless at pH 7.2 (Handbook of Chemistry and Physics 1978).

Future efforts will include synthesis of the remaining position 189 mutants (W, G, H, and Y). Other efforts will be focused on the development of a selection which will allow direct assessment of the trypsin activity synthesized by a colony. Also, the 20 trypsins (including wild type) will be tested with the gel overlay activity gel screen against several substrates. The information obtained using these mutants will provide insights into the atomic details of the functioning of the substrate binding pocket.

Chapter Two

Examination of the Substrate Specificity of Trypsins Mutated at Position 189

ABSTRACT

A quantitative description of the forces that determine the narrow specificity of trypsin has proved elusive although many of the key interactions between the protein and substrate have been enumerated as a result of solution of three dimensional structure and kinetic analysis. An improved bacterial expression system which generates approximately 1 mg trypsin/ L of culture (~3.0 OD of cells/ L) has promoted the analysis of a set of 18 trypsins differing at aspartic acid 189. Using a gel overlay screen sufficiently sensitive to detect enzymes with 10^{-5} the activity of trypsin, five mutants were identified as active towards arginine, trypsin D189C, E, S, N, and A. Partial purification of these mutants allowed determination of the kinetic parameters kcat and Km which indicated that loss of the negative charge as in trypsin D189N causes a drop of activity of 10^5. The two most active mutants trypsin D189C and D189E retained 1.4% and 0.01% the activity of trypsin against Tos-GPR-AMC. Apparently, the γ sulfhydryl mimics the γ carboxylate. A pH-activity profile of trypsin D189C was obtained with Z-GPR-AMC to examine the titration of the cysteine which has a nominal pKa= 8.0. However, no additional inflection points in the kcat, Km or kcat/Km profiles were observed.

INTRODUCTION

Trypsin, a paradigm for the family of serine proteases, hydrolyzes the peptide bond C-terminal to L-arginine and lysine amino acids at a rate that is approximately 10^{10} fold greater than the amidolysis observed in neutral water (Kahne and Still 1988)[21]. Among the remaining natural amino acids, the preferred substrate is tyrosine (Gráf, Craik et al. 1987) which is discriminated against by 10^5 relative to arginine and lysine (Gráf, Jancsó et al. 1988). The specificity of trypsin is determined by the S1 pocket[22] which binds the arginine and lysine side chains although binding sites amino and carboxy termial to the S1 site do contribute somewhat to specificity. The substrate length directly correlates to activity (Fiedler 1987), and effects as large as 7 fold at each of the P2 and P3 positions have been characterized (Fiedler 1987; Tanaka, McRae et al. 1983)[23].

Use of synthetic inhibitors has defined two components of the S1 pocket, an anionic site and an adjacent hydrophobic slit (Mares-Guia and Shaw 1965; Mares-Guia and Shaw 1967). Studies of neutral derivatives of the substrates lysine and arginine side chains such as heptyline and citrulline (Sanborn and Hein 1968) indicated that the positive charge on the substrate was essential. Studies

---

[21] D-amino acids are very poor substrates and are generally considered competitive inhibitors with K_i constants in the 1-10 mM range (Fersht, 1985).

[22] The nomenclature for the enzyme-substrate interaction is according to convention with the substrate numbered P1 at the catalytic site and P1', P2' and forward on the carboxy terminal side and, upstream of the P1 amino, from P2. The cognate sites in the protein have the S prefix instead of P (i.e. the S2' subsite binds the P2' amino acid) (Schechter and Berger 1967).

[23] Analysis of chymotryptic interactions at secondary binding sites indicates that amino acids from P5 to P3' have effects on catalysis (J. Powers, personal communication).

involving γ-guanidino-amidobutyrate (Baird, Curragh et al. 1965), homoarginine (Baines, Baird et al. 1964), and ornithine (Baird, Curragh et al. 1965; Sanborn and Hein 1968) continued the investigation of the hydrophobic opening of the S1 pocket and confirmed that the anionic site and the scissile carbon were rigidly separated by a precise distance (Table i-1).

Solution of the three dimensional structure of α-chymotrypsin (Matthews et al. 1967) and subsequently trypsin (Kreiger, Kay et al. 1974; Ruhlmann, Kukla et al. 1973) pinpointed the aspartic acid at position189[24] as the determinant of the anionic character of the P1 binding pocket. In conjunction with the serine 190 methoxy group, the carboxylate defines the base of the pocket. The structures of trypsin complexed to the inhibitors benzamidine and BPTI (basic pancreatic trypsin inhibitor) have been solved and reveal a number of hydrogen bonds. Examination of the steric packing of the BPTI lysine 15 show that extensive hydrophobic contacts to the walls of the pocket are formed with the side chain.

Replacing the glycine 216 and glycine 226 separately and in combination with alanine (Craik, Largman et al. 1985) confirmed that the base of the pocket is very constrained in volume. Additional methyl groups are accommodated only with the loss of activity. Also, these experiments underlined the importance of the structural water molecules in the ability of a single binding pocket to retain rough parity in the specificity toward both Arg and Lys

---

[24]The amino acids are numbered according to the accepted standard which is based on the numbering of the amino acids in chymotrypsinogen (Hartley et al., 1965).

substrates. Replacing the anion with a cation, the D189K mutation (Gráf, Craik et al. 1987), did not reverse the specificity of the enzyme. At least in this case, the binding pocket did not function as a symmetric tube whose character is solely determined by the amino acid at the base of the pocket.

To more fully assess the function of the aspartate at position 189 in the enzyme-substrate interaction, 17 point mutants were made by oligonucleotide-directed mutagenesis. A screen for activity against arginine amide and ester substrates allowed rapid assessment of the activity of the variants. Trypsin mutants that retained Arg-specific catalytic activity were partially purified and kinetically characterized. The trypsins were also screened against a panel of amide and ester substrates to determine if the specificity of the enzyme had been re-directed by the substitution at position 189.

MATERIALS AND METHODS

Materials

Restriction enzymes were obtained from New England Biolabs (Beverly, MA), Boehringer Mannheim Biochemicals (Indianopolis, IN), or Promega (Madison, WI). T4 DNA ligase was obtained from New England Biolabs or United States Biochemical Corp. (Cleveland, OH). Isopropyl ß-D-thiogalactopyranoside (IPTG), Nα-carboxybenzyl-seryl methyl ester, acetyl-glutamyl methyl ester, acetyl-leucyl methyl ester, benzyl-alanyl methyl ester, and Nα-*p*-tosyl arginyl methyl ester (TAME) were obtained from Sigma Chemical Co. (St. Louis, MO). The substrates Nα-carboxybenzyl (Z)-Gly-Pro-Arg-7-amido-4-methyl coumarin, succinyl-Ala-Ala-Pro-Phe-7-amido-4-methyl coumarin, tertiary butyloxycarbonyl (tBOC)-Leu-Gly-Arg-*p*-nitroanilide, succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide, succinyl-Ala-Ala-Pro-Ala-*p*-nitroanilide, and acetyl-$(Ala)_3$ methyl ester were obtained from Bachem Biosciences Inc. (Philadelphia, PA). SeaPlaque low gelling temperature agarose was from FMC Bioproducts Corp. (Rockland, ME). *p*-chloromecuro-benzoate (*p*CMB) was purchased from ICN Biomedicals, Inc. (Costa Mesa, CA).

Bacterial Strains

Escherichia coli strain MH1 (araD139, ΔlacX74, galU, galK, hsr^-, hsm^+, rpsL) was a gift of Dr. M. Hall (University of California-San Francisco). JA221I (F' $lacI^q$, lacZ, pro/ lpp, trpE5, leuB6, lacY, recA1) (Nakamura, Masumai et al. 1982) was a gift of Dr. P. Babbitt (UCSF). Dr. C. Desplan (UCSF) provided strains DG98 (F' lacZ▼M15,

proAB/ thi, hsdR17, endA44, supE44Am, lacI^q, lacZ▼M15, proC::Tn10) and DG101 (F$^-$/ thi, hsdR17, endA44, supE44Am, lacI^q, lacZ▼M15). *E.coli* X90 (F' lac I^q, lacZY, pro AB/ Δ(lac-pro), ara, nalA, argEam, thi, rif^r) (Amann, Brosius et al. 1983) and *E. coli* SY903 (F' lac I^q, lacZ::Tn5, proAB/ recA1, srl::Tn10, Δlac-pro, rif^r, nalA) (Sauer, Smith et al. 1988) were gifts from Dr. A. Vershon (University of California- San Francisco). *E. coli* CJ236 (F' CJ105/dut1, ung1, relA1) was a gift of BioRad Laboratories Inc. (Richmond, CA). *E. coli* XLI-blue (F' proAB, lac I^q, lacZ▼M15, Tn10/ recA1, endA1, gyrA96, thi, hsdR17 (rk$^-$, mk$^+$), supE44, relA1, λ$^-$, lacZ) was purchased from (Stratagene, La Jolla, CA). *E. coli* SM138 (F$^-$/ araD139, Δ(lac)169, relA1, rpsL, phoR) was a gift of Dr. S. Michaelis (UCSF).

Plasmid Constructions

Four vectors were used in these experiments, pBsTn2 (Evnin and Craik 1988), pTn2 (Vásquez, Evnin et al. 1989), pFA54 (gift of K. Nikaido and unpublished although analogs of this vector are discused in (Ames, Nikaido et al. 1989), and pT3. Vector pBsTn2 contains the DNA sequence of mature trypsin fused at its 5' end to the promoter and signal peptide of alkaline phosphatase and at its 3' end to the *trpA* transcription terminator. Plasmid pTn2 contains the alkaline phosphatase-trypsin-terminator cassette present in pBsTn2 spliced into the Bam HI and Sal I sites of pBR322. Plasmid FA54 was constructed by inserting the coding sequence for *Salmonella typhimurium hisJ* into the polylinker that separates the *tac* promoter and the ribosomal RNA transcriptional terminator in

pKK177-3 (Brosius and Holy 1986). The Shine Dalgarno sequences of both *lac* (*tac* promoter) and *hisJ* are present.

pT3 was constructed in three steps. First, a 460 bp Bam HI-Eco RI DNA fragment from pFA54 encoding the *tac* promoter (deBoer, Comstock et al. 1983), *his* J signal peptide, and the first 30 amino terminal residues of the mature *his* J coding sequence was blunt-end ligated into pBsTn2 at the unique Bam HI site located 5' to the alkaline phosphatase promoter (Evnin & Craik, 1988). This construction was isolated from transformation of strain DG-101 but could not be isolated in MH-1. The presence of additional lac repressor protein which was produced by the *lacI^q* allele may be the important difference.

Second, precise fusion of the *his* J signal peptide sequence to the mature trypsin sequence (Ile 16) was achieved by oligonucleotide-directed deletion of 431 intervening base pairs. However, the established mutagenesis scheme (Evnin and Craik 1988) based on mutagenesis without phenotypic selection could not be used. CJ236 cells could not be coaxed to generate uracil-laden single stranded DNA of this composite plasmid. Consequently, single stranded DNA from *E. coli* DG98 was employed. Enrichment for the mutant duplexes present in the mutagenesis reaction was possible since the mutagenesis eliminated a Kpn I restriction site.

Finally, the Ssp I-Sal I 917 bp DNA fragment encoding the *tac* promoter, the *his*J and trypsin sequences, and the *trp*A terminator from the pUC-based pBsTn2 vector were spliced into a 2010 bp Pvu II- Sal I digested pBR322 DNA fragment.

For high-level expression of mutant trypsins, the 327 bp Xho I-Sty I DNA fragment of each trypsin gene was subcloned from pBsTn2 into pT3. Two to three μg of the pBsTn2 DNA was digested with Xho I and Sty I and the DNA was electrophoresed in 0.8% low gelling agarose, 2.4 μg/ml ethidium bromide, 1x TAE, at 4°C . Approximately 1/3 of the fragment (15 μl) was added to a ligation cocktail (5% PEG 8000, 50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT) containing the purified vector DNA fragment from pT3 (100 ng or 0.04 picomoles). T4 DNA ligase was added and the reaction was incubated for 2-16 hr at ambient temperature (King and Blakesley 1986). Isolates were identified by restriction digestion and the presence and identity of the position 189 mutation was confirmed by dideoxy sequence analysis (Sanger, Nicklen et al. 1977).

Each position 189 mutant is stored as the pT3 plasmid; most of the mutants are also stored as Xho I-Sty I fragments embedded in low temperature gelling agarose. Additionally, *E. coli* strain X90 harboring each plasmid is stored as a glycerol stock at -70°C. The desired plasmid can be readily re-introduced into X90 or other cells using the Hanahan quick transformation procedure (Hanahan 1985). Briefly, approximately 10^8 cells from a fresh streak of the desired strain are resuspended in TFB buffer (100 mM KCl, 45 mM MnCl•4 H_2O, 10 mM $CaCl_2$•2 H_2O, 3 mM $HACoCl_3$ (hexamine cobalt), 10 mM potassium buffered MES pH 6.2). The suspension is incubated at 0°C for 10 min, then, with DNA, for 10 min. Heat shock and recovery of cells follows as with $CaCl_2$ competent cells (Mandel and Higa 1970).

For these experiments, trypsin was harvested from X90 cells carrying the pT3 plasmid grown in LB, 50 mg/ml ampicillin, 1 mM IPTG. From 15 mls of culture, 400 µl of periplasmic extract were obtained as described (Evnin and Craik 1988). The trypsins were stored frozen in the solution used for harvesting: 25% sucrose, 10 mM Tris-HCl pH 8.0, 5 mM EDTA, 5 mg/ml (hen egg white) lysozyme.

Screening

Activity gel screen (see Chapter One). Trypsin activity from putative positive isolates was assayed by activity gel analysis as described (Evnin and Craik 1988). In brief, the activity of the trypsin which is embedded in a SDS polyacrylamide gel is monitored by observing a color change in a polyacrylamide gel overlay impregnated with the ester substrate *N*α-*p*-tosyl-L-arginine methyl ester (TAME) and the pH indicator phenol red. Incubation of the gel sandwich continued at 37°C for 1-3 hr. The stock solution of TAME was 200 mM (final concentration 65 mM); acetyl-phenylalanyl methyl ester, acetyl-leucyl methyl ester, benzyl-alanyl methyl ester, carboxybenzyl-seryl methyl ester and acetyl-glutamyl methyl ester were dissolved to 50 mM (final concentration, 16.5 mM) in 25% DMF, 50% DMF, 50% DMF, 40% DMF, and water, respectively.

Microplate screen. For each assay, 3 or 15 µl of the extract were added to 180 µl of 100mM Tris-HCl pH 8.0, 1.5 mM *t*-butyloxycarbonyl-Leu-Gly-Arg-*p*-nitroanilide, 3% DMF in one well of a 96-well microplate. Succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide

and succinyl-Ala-Ala-Pro-Ala-*p*-nitroanilide were also tested. Trypsin activity was measured in duplicate by following the release of *p*-nitroaniline at 405 nm using the UVmax Kinetic Microplate Reader (Molecular Devices Corp., Redwood City). *p*CMB assays were conducted by the addition of a 25 mM solution to a final concentration of 2.2 mM to the assay mixture prior to enzyme addition. The rate was obtained immediately over a minimum of 30 min except in the case of trypsin where product inhibition would be significant.

Enzyme Purification

Partial purification. *E. coli* X90 cells containing the desired mutant plasmid were inoculated into 500 ml of LB, 50 mg/L ampicillin, 1 mM IPTG and cultured for 12-16 hr. The periplasmic fraction from these cultures was collected into 25% sucrose, 10mM Tris-HCl pH 8.0, 5 mM EDTA as described (Vásquez, Evnin et al. 1989), and was dialyzed exhaustively at 4°C in 10 mM sodium citrate, pH 2.8. The acid-precipitate was separated by centrifugation for 20 min at 12,000 x g and the supernatant containing the trypsin was concentrated by ultrafiltration in an Amicon (Danvers, MA) Centriprep 10 (10,000 MW cutoff). From approximately 20 ml of dialysate, 0.5 ml of extract was obtained. This concentrated trypsin solution was spun at 14,000 x g for 5 min and the precipitate discarded. The trypsin was stored in 10 mM sodium citrate pH 2.8 at 4°C.

The concentration of active sites was determined by following the release of 4-methylumbelliferone from 4-methylumbelliferyl

p-guanidinobenzoate (MUGB) (Jameson, Adams et al. 1973). No burst was observed with an extract from cells carrying pLIB (Figure 2-1).

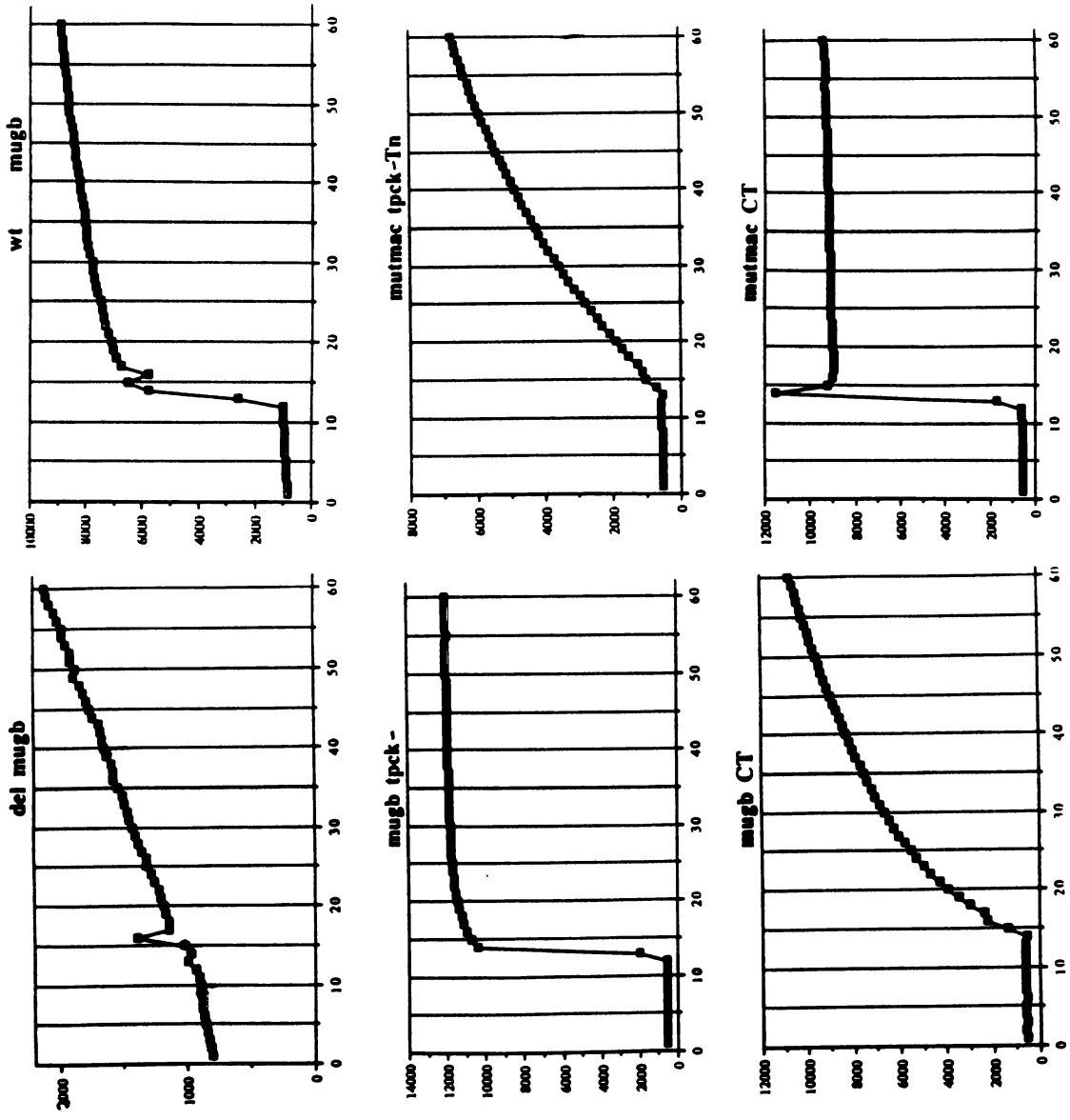
A comparison of the kinetic parameters obtained from analysis of acid-precipitated extract and of affinity purified material was completed on Z-GPR-AMC. The apparent Km constants obtained with acid-precipitated extracts were within 10% of each other and with the affinity purified trypsin (18 μM); the observed kcat for the purified extract (6240 min^{-1}) was approximately twice the observed kcat for affinity-pure trypsin (kcat=2370 min^{-1}) (Higaki, Evnin et al. 1989). The rate amplification was also observed for the Tos-GPR-AMC substrate. However, the Arg/Lys substrate preference derived from the kinetic constants for the purified extract and the affinity-pure trypsin were indistinguishable. Hydrolysis of Z-Gly-Pro-Arg-AMC by extracts from cells containing pLIB was negligible.

Affinity Purification. Trypsin D189C was affinity purified from a ten liter fermentor growth in three steps as described (Higaki, Evnin et al. 1989). X-90 cells with the pT3-D189C plasmid were grown in a modified MOPS media at 37°C with glucose fed at 1 ml/min in a 10 L fermenter. The presence of 2 mM IPTG caused constitutive expression of the mutant protein. The cells were harvested at an A_{600} of 3.2.

The periplasmic proteins were extracted by osmotic shock to yield a final solution of proteins in water, which was buffered with

Figure 2-1. Active Site Titrations Using MUGB and MUTMAC

The vertical axis of the graphs is arbitrary units of fluorescence; note some of these scales are different to emphasize different aspects of the profiles. The horizontal axis is time. The 50 mM Tris-HCl pH 8.0, 100 mM NaCl buffer was maintained at 26°C and the data collection was begun at the time of addition of 10 µl of the titrant stock. After approximately 30 sec (10 time points), enzyme was added and the data acquisition was continued for the remainder of the 3 min assay period. The fluorescence absorbance was calibrated at the beginning of each day using an 4-methylumbelliferone enabling calculation of the concentration of the active sites. In the top two graphs are depicted the reaction of MUGB with extracts from cells carrying the pLIB construction which has a deletion at position 188-190 and cells carrying the wild type trypsin plasmid. Note the classical burst kinetics that characterizes the single turnover of MUGB by trypsin on the left. In contrast, no burst is observed in the control extract on the right. In the bottom four panels are depicted controls for this experiment. In the top left and right are the MUGB and MUTMAC profiles, respectively, obtained using bovine cationic trypsin treated with the chymotrypsin modification reagent tosyl-phenylalanine chloromethyl ketone (TPCK). Trypsin can be titrated using MUGB but not with MUTMAC. In the bottom panels, are the profiles obtained using chymotrypsin with the same two substrates. As reported, chymotrypsin hydrolyzes MUGB but can be titrated with MUTMAC.


wt mugb
mutmac tpck-Tn
mutmac CT
del mugb
mugb tpck-
mugb CT

0.5 M MES to 10 mM MES, pH 6.0. A slurry of DEAE Sepharose Fast Flow was added to the protein solution, shaken gently at 37°C for 15 min, and poured into a column. The column was washed with 10 mM MES, pH 6.0 until the flow through was less than 0.1 OD at A_{280}. It was eluted with a gradient of 0 to 0.5 M NaCl. Immunoblot analysis identified the trypsin D189C containing fractions which were pooled and dialyzed against 50 mM glycine-HCl, pH 2.5 at 4°C.

The dialyzed trypsin pool was loaded on a CM Sepharose Fast Flow column, washed with loading buffer and eluted with a gradient of 0 to 0.5 M NaCl. Immunoblotting using both anti-trypsin and anti-ecotin[25] antibodies showed that the CM column run at low pH successfully separated these two proteins. Trypsin D189C fractions were dialyzed against 10 mM MES, pH 6.0 at 4°C.

This material was loaded onto a BPTI affinity column[26]. The column was washed with 10 mM MES, pH 6.0 with 0.5 M NaCl and eluted with 0.1 M citric acid. Immunoblot analysis showed that the vast majority of the trypsin D189C was in the salt wash with some in the flow through and virtually none in the acid fractions. The flow through and salt wash were pooled and dialyzed against 10 mM MES, pH 6.0. A second BPTI affinity column was made and similar results were obtained[27].

---

[25]Ecotin is an endogenous *E. coli* inhibitor that binds very tightly to trypsin and other serine proteases (Chung et al 1983). Its mechanism is apparently different than other serine protease inhibitors (M.McGrath, personal comunication, 1990).

[26]Prepared by immobilizing BPTI on Affigel-10 (Bio-Rad, Inc.) according to the manufacturer's instructions.

[27]It is likely that the failure to bind to this column lies in the steps that generated the affinity resin rather than the inability of trypsin D189C to bind to BPTI (basic pancreatic trypsin inhibitor). The dissociation constant of this inhibitor to trypsin is 1.6×10^{-13} (Lazdunski et al, 1974). Consequently, even

The flow through and salt wash from the second BPTI column were pooled and dialyzed against 10 mM Tris, pH 8.6 and cycled through a benzamidine column at 4°C for 4 days to maximize binding of the mutant trypsin. The column was washed with 10 mM Tris, pH 8.6, then with Tris pH 8.6, 0.5 M NaCl. Mutant protein was eluted with 0.1 M citric acid, 10 ml at a time with a 30 min wait for 4 collection periods. Aliquots of both washes and the elution fractions were examined by Coomassie staining and by immunoblot. Mutant protein was found in the salt wash and in the acid elution. Approximately 80% of the trypsin D189C eluted in the salt wash with other contaminants. Only the trypsin that eluted in the acid fractions was used for the kinetics experiments. These fractions were pooled and concentrated using an Amicon Centricon 10. *p*CMB assays were accomplished by adding 10 µl to a 1 ml cuvette containing the buffer and enzyme. A dilution series over 6 orders of magnitude, beginning with a final concentration of 250 µM *p*CMB, was examined.

---

a catastrophic drop in ability to bind lysine (the P1 amino acid) of 10^6, should still allow BPTI to bind to trypsin sufficiently well for chromatographic separation.

RESULTS

High Level Expression of Active Trypsin

Direct expression of trypsin in *E.coli* was achieved by fusing the signal peptide of alkaline phosphatase to the Ile16 of trypsin (Evnin and Craik 1988). The alkaline phosphatase promoter directs high level expression of its cognate gene although full induction occurs only upon phosphate limitation (Vásquez, Evnin et al. 1989). *PhoR* strains (no repressor expression) constitutively activate the *phoA* promoter to approximately 30% the levels observed for full induction (Torriani and Rothman 1961). Thus, the convenience of constitutive expression can only be realized at the expense of promoter power and at the loss of the freedom to readily switch strains. Alkaline phosphatase is expressed to high levels from its cognate promoter, but trypsin is not. Among the myriad of possible explanations for the differences in the expression levels include the differences in the primary and secondary structure of the RNA transcripts. Also, it is possible that a signal peptide will support different levels of secretion for different proteins.

The *Salmonella typhimurium* amino acid carrier protein hisJ is expressed to high levels in *E.coli* TA3997 which harbors pFA54 (K. Nikaido, personal communication). Transcription is controlled by the *tac* promoter, and upon IPTG induction, hisJ protein accumulates to approximately 25% of the total periplasmic protein (K. Nikaido, personal communication). In an effort to achieve higher levels of trypsin expression, pT3 was built from pBsTn2 and pFA54. In the final construct, trypsin expression is controlled by the *tac* promoter and *trpA* terminator, and the efficiency of

translation is mediated by the *tac* and *hisJ* Shine-Dalgarno sequences. Secretion of trypsin is directed by the hisJ signal peptide. The tac, hisJ, trypsin, and terminator (T3) expression cassette was spliced into pBR322 from the pUC based plasmid pBS in the anticipation that expression would be more tightly regulated in a lower copy number plasmid (Figure 2-2).

Trypsin was harvested from induced and uninduced cultures of four different *lacIq* strains each carrying pT3 (Figure 2-3). In all four strains tested, addition of IPTG induced high level trypsin expression. As is usual in *E. coli*, low level gene expression without induction is observed. *E. coli* X90 provided the highest yields of active trypsin followed by JA221I, SY903, and XL1-blue (Figure 2-3). Microplate analysis of trypsin activity indicated that X90 cultures generated 3-4 more activity than JA221I cultures upon constitutive trypsin production.

Immunoblot analysis of pT3 and pT3D189E expressed in X90 cells indicated constitutive expression in 1 mM IPTG yielded a greater than or equivalent amount of trypsin as was produced via induction (Figure 2-4). Microplate analysis indicated that two fold more active trypsin was obtained by constitutive expression. Figure 2-4 shows that X90 pT3 is more effective than the previous best expression strain, SM138 pTn2. High level expression of active trypsin in the periplasmic space of X90 cells did not appear to be detrimental to their growth. Active trypsin constituted approximately 0.5% of the total protein present in the purified extract as determined by measurement of the concentration of total

Figure 2-2. Plasmid pT3

The dappled box indicates sequences from the *tac* promoter. The vertically striped box indicates sequences of hisJ translation control and signal peptide sequences. The box with heavy slanted lines represents the trypsin coding sequence with the Eco RI site indicated for reference. The black box represents the *trpA* terminator. Important restriction sites are noted. The nucleotides and cognate amino acid sequence are indicated in the detail of the junction between the hisJ and trypsin sequences. The heavy arrow indicates the direction of trypsin. Lighter arrows indicate the direction of ß-lactamase (Amp^r) transcription and the orientation of the origin of replication. The "Z" line in the amino acid sequence of the detail section indicates the site of signal peptidase cleavage. The underlined bases in the detail section show the Pvu I restriction site.

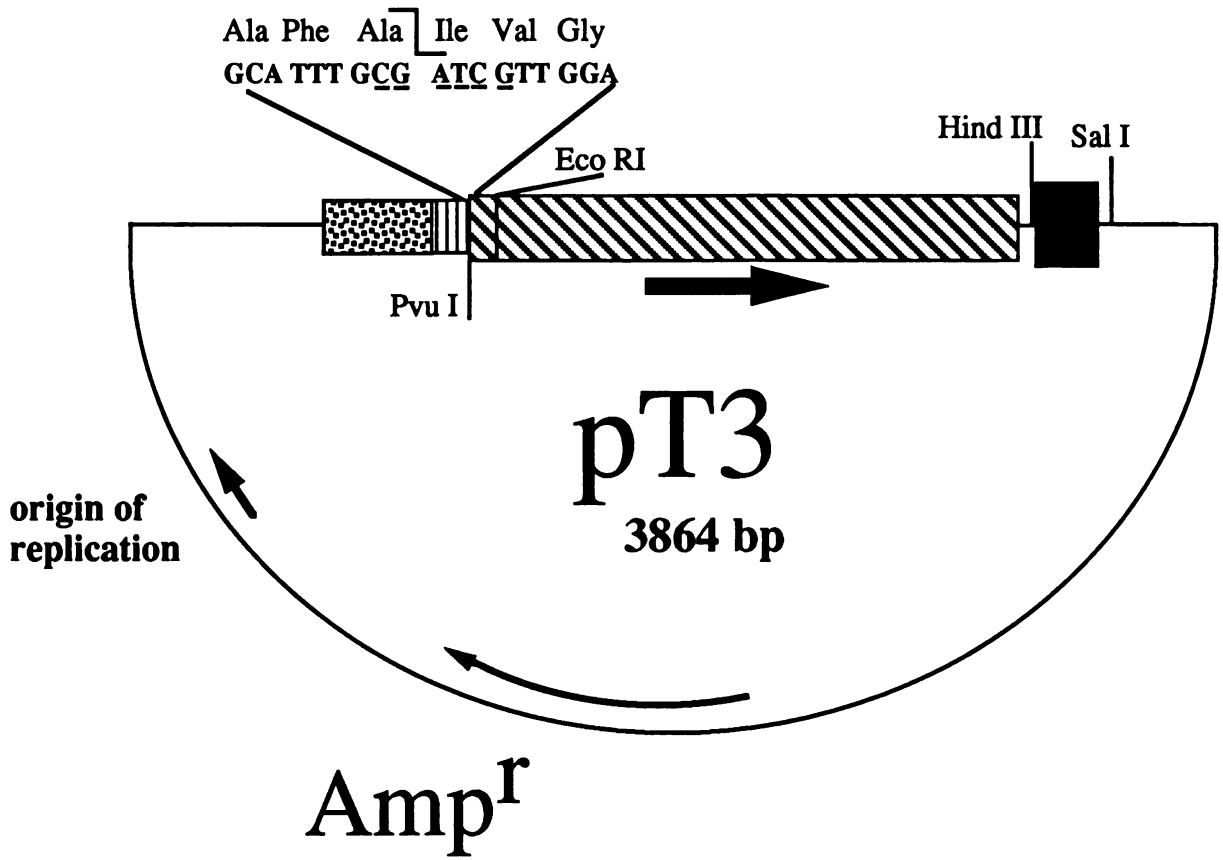

Ala Phe Ala Ile Val Gly
GCA TTT GCG ATC GTT GGA
Eco RI
Hind III
Sal I
Pvu I
pT3
3864 bp
origin of
replication
Amp^r

Figure 2-3. Expression of Trypsin from pT3 in Different Strains of *E. coli*

Immunoblot analysis of the amounts of trypsin harvested from 1.5 ml of a saturated culture of each of four $lacI^q$ strains carrying pT3. Induction of the *tac* promoter was accomplished by dilution of a saturated culture 1:10 into fresh media containing 1mM IPTG. The new culture was incubated for 4 hours and then both the uninduced and induced cultures were harvested. Each of the four strains is shown as induced (+ IPTG) and uninduced from left to right. The name of the strain in indicated above each lane. The trypsin band is indicated with an arrow.

Expression of pT3 in Different Strains

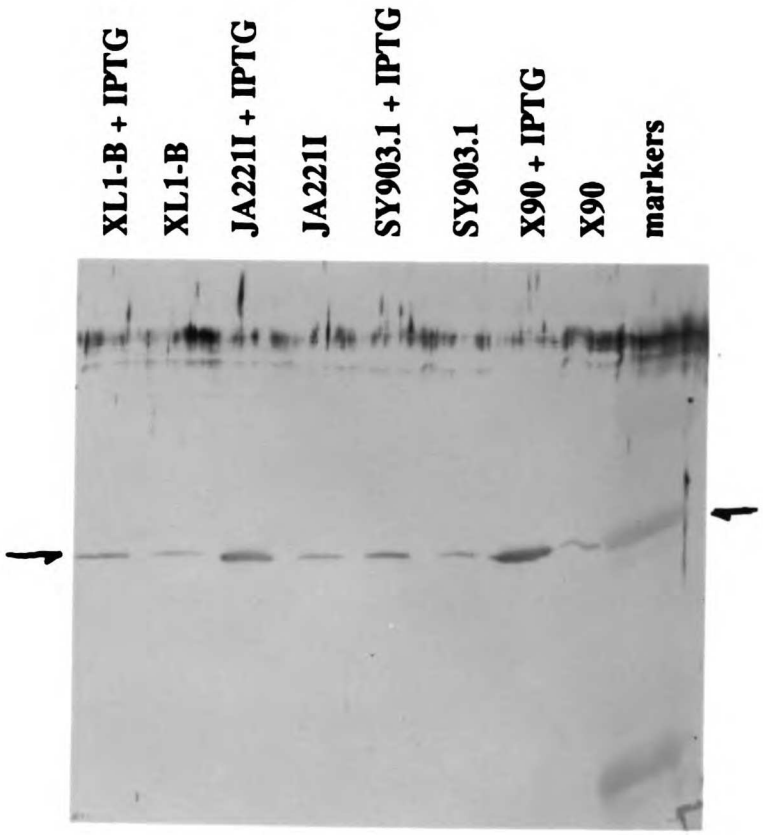
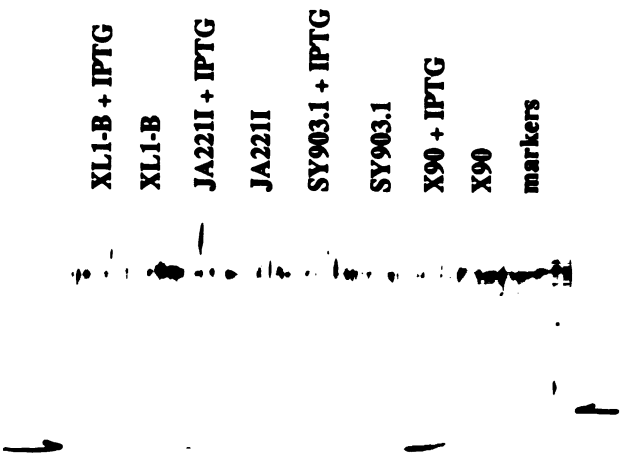


Figure 2-4. Constitutive and Induced Expression of pT3

Immunoblot analysis of trypsin and trypsin D189E produced in the X90 pT3 system and trypsin produced from the SM138 pTn2 system. Both a total protein and a periplasmic protein preparation were analyzed. Total protein was prepared by addition of the equivalent of 100 µl of saturated culture to 2x Laemmeli loading dye with 2-mercaptoethanol (Laemmli 1970). The sample was boiled for 10 min and then mixed vigorously. Periplasmic protein was prepared from 1.5 ml as described (Evnin and Craik 1988). Trypsin production was induced (+ IPTG) as in the cultures in Figure 2-3. Constitutive production (const) was enabled by addition of 1 mM IPTG to the growth media upon inoculation. Both the wild type trypsin which has aspartic acid 189 and the D189E (Asp189Glu) mutant were expressed. Both mutants are active although the mutant retains 0.01% the activity at pH 8.0. Constitutive expression of pTn2 is a result of the *phoR* mutation in SM138 (Vásquez, Evnin et al. 1989). The trypsin standard is purified rat trypsin. The trypsin band is indicated by the arrow.

Expression of pT3 in Different Strains



protein (Smith, Krohn et al. 1985) and the concentration of trypsin active sites.

Arginine Specificity of the Position 189 Trypsin Mutants

18 trypsin mutants differing only at position 189 at the base of the substrate binding pocket were subcloned from pBsTn2 to pT3. 15 of these mutants had been isolated from a library of position 189 mutants (Evnin and Craik 1988) and the other 3 had been specifically generated. Immunoblot analysis confirmed that all the mutants were expressed and secreted to the cytoplasm. The activity of the 18 trypsins was examined by activity gel analysis using the arginine substrate TAME. In descending order of catalytic potency, trypsin, trypsin D189E, D189N, D189S, and D189A were active. The same extracts were also examined by microplate analysis using tBOC-Leu-Gly-Arg-*p*-nitroanilide. The rates measured for trypsin D189C and D189E were 4.4% and 2.5% those of trypsin, respectivley. Trypsins D189N, A, and S were only detected in the gel overlay assay and the activity could not be quantitated. Apparently, SDS-PAGE of trypsin D189C reduced or destroyed its activity although the mechanism of the loss is not understood. The two most active mutants, trypsin D189E and D189C, are presumably negatively charged at pH 8.0. The pKa of the glutamate carboxylate is 4.0, and the pKa for the cysteine sulfur is 8.0. The oxidation state of the cysteine at position 189 is not known; the oxidized forms of the sulfhydryl are also negatively charged.

For further characterization, the active trypsins were partially purified. The endogenous *E. coli* inhibitor ecotin (Chung, Ives et al. 1983) was not separated from the trypsin or trypsin mutants in this process. Consequently, a fraction of the trypsin is inactivated; it is assumed that the mutants were similarly inactivated. Single turnover was observed for trypsin D189E using MUGB, but approximately linear rates of hydrolysis were observed for MUGB for the D189N, D189A, D189S mutants. The inability to form a stable acyl-enzyme complex may reflect the loss of the carboxylate at the base of the pocket. MUTMAC, a titrant for chymotrypsin (Jameson, Adams et al. 1973) (Figure 2-1), was also hydrolyzed rapidly by these mutants, as it is by trypsin. Consequently, the concentration of active trypsin for these mutants was estimated by immunoblot analysis of the periplasmic extract. This can be an accurate method of quantitation of the total trypsin content but not the concentration of active sites.

For trypsin D189E, Km is increased by 110 fold and kcat is decreased by 65 fold retaining 1.5% the rate of hydrolysis observed with trypsin (Table 2-1). It appears that the additional methylene group of the glutamate side chain reduces the stability of both the ground state complex and the transition state complex. Trypsin D189C on Tos-GPR-AMC retains 3.8% of the rate for trypsin, and exhibits a four fold increase in Km (Table 2-1). Apparently, the sulfur atom (or the oxidized derivatives) mimics the aspartic acid carboxylate with reasonable fidelity.

Trypsin D189N, D189A, and D189S retained similar, low levels of Arg-specific activity with observed kcat/Km constants of 0.01,

Table 2-1. Kinetic Analysis of Mutants at Position 189 on an Arginine Substrate

These values represent the average of at least two determinations of the Michaelis Menten constants for each trypsin. A determination comprised the rate of hydrolysis at five different concentrations of tosyl-Gly-Pro-Arg-AMC: 200, 100, 50, 25, and either 12.5 μM or 500 μM. The release of AMC was followed in a Perkin-Elmer LS-5B Luminesence Spectrometer calibrated with a standard solution of AMC at 460 nm with excitation at 380 nm. Enzyme was added to a final concentration of 2.5-150 nM (depending on enzyme activity) to 50 mM Tris-HCl pH 8.0, 100 mM NaCl. This buffer was maintained at 26°C. Velocities were obtained from the linear portion of the free AMC vs. time curve. The kinetic parameters Km and kcat were obtained using a linear regression analysis of the Eadie Hofstee plot V vs. V/{S}. The error in the determinations for trypsin, trypsin D189C, and trypsin D189E is less than 20% and in most cases less than 10%. Only the kcat/Km values could be obtained for trypsin D189N, D189A, and D189N/S190T because of their reduced activity. Furthermore, determination of the concentration of active sites for these enzymes could not be obtained with either 4-methylumbelliferyl-*p*-guanidinobenzoate or the chymotrypsin titrant 4-methylumbelliferyl-(*p*-N, N, N-trimethylammonium) cinnamate chloride (Jameson, Adams et al. 1973). For the purposes of comparison, immunoblot analysis was used to determine the concentration of trypsin in these samples. Consequently, the kcat/Km values are accurate to within an order of magnitude.

amino acid	kcat(/min)	Km(μM)	kcat/Km (/min/μM)	% kcat/Km
Asp	4450	4.8	927	100
Cys	170	13.5	12.6	1.36
Glu	69	530	0.13	0.014
Asn			0.02	0.002
Ser			0.07	0.008
Ala			0.01	0.001

0.02, and 0.07 respectively (Table 2-1). For trypsin D189S, activity loss of 1.3 x 10^4 is consistent with the 5.7 x 10^4 loss in activity observed previously (Gráf, Jancsó et al. 1988). The individual constants Km and kcat could not be determined for these weakly active mutants. Asparagine retains one hydrogen bond acceptor (Nitrogen can act only as a hydrogen bond donor). However, as alanine, serine and asparagine provide approximately similar substrate stabilization, any beneficial interactions such as hydrogen bonding between the oxygen of the asparagine amide and the arginine guanidinium group are apparently offset by the deleterious effects of the amide for carboxylate substitution. For example, hydrogen bonds to atoms in the protein and to the substrate maybe be absent or only partially satisfied.

Trypsin D189C

A large scale preparation of this protein was carried out and the enzyme was purified to homogeneity by affinity chromotography (Figure 2-5). Silver staining of a SDS polyacrylamide gel was used to quantitate the amount of trypsin present after purification (Figure 2-5). Kinetic parameters on Z-GPR-AMC were determined from pH 5.0 to 10.5 (Figure 2-6). At pH 8.1 and 26°C with buffers of 50 mM MOPS, and 50 mM TAPS containing 100 mM NaCl, 1 mM $CaCl_2$, trypsin achieved a kcat/Km of 5.3 and 4.9 $\mu M^{-1}min^{-1}$ respectively. The kcat was 170 min^{-1} and the Km constants were 32 and 25 μM respectively. For Tos-GPR-AMC at pH 8.0 and 26°C with 50 mM Tris-HCl containing 100 mM NaCl a kcat/Km of 12.6 $\mu M^{-1}min^{-1}$ observed. Although the kcat was identical at 170 min^{-1}, the

Figure 2-5. Purification of Trypsin D189C

From left to right, lanel is protein markers (200, 97.4, 68.0, 43.0, 29.0, 18.4, and 14.3 kD), lanes 2-5 are increasing amounts of purified rat trypsin (6 ng (1 µl), 30 ng (5 µl), 150 ng (25 µl), and 300 ng (50 µl) of a 250 nM stock solution), lanes 7-12 are increasing amounts of the stock trypsin D189C solution (10, 25, and 50 µl). Note that even the 50 µl trypsin D189C lane is clean except for the band at 24,000 kD which is trypsin D189C and other faint bands. Ecotin (16 kD) is absent. The sample is judged to be better than 95% pure. The concentration of the wild type trypsin was determined by active site titration and the minor high molecular weight contaminants would not affect this value. The A_{280} of the trypsin D189C material indicated that the concentration was 25 µM; active site titration was not possible using MUGB. This experiment indicates that the stock solution of trypsin D189C is approximately 2.5 µM since 10 µl of the stock solution contained approimately 600 ng of trypsin.

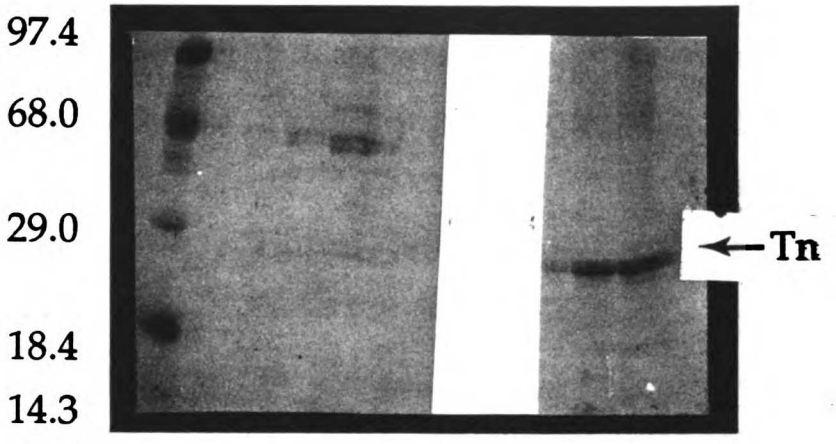
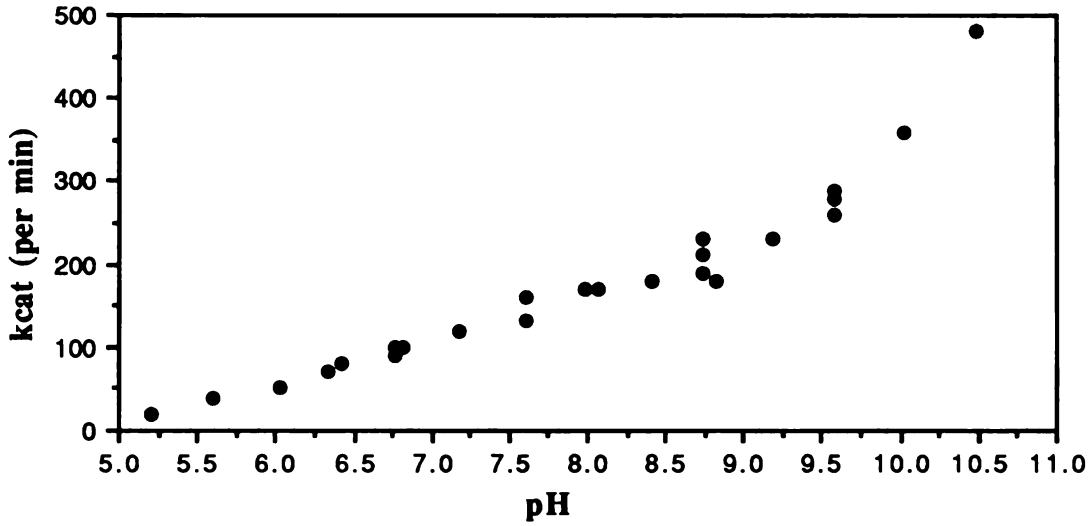
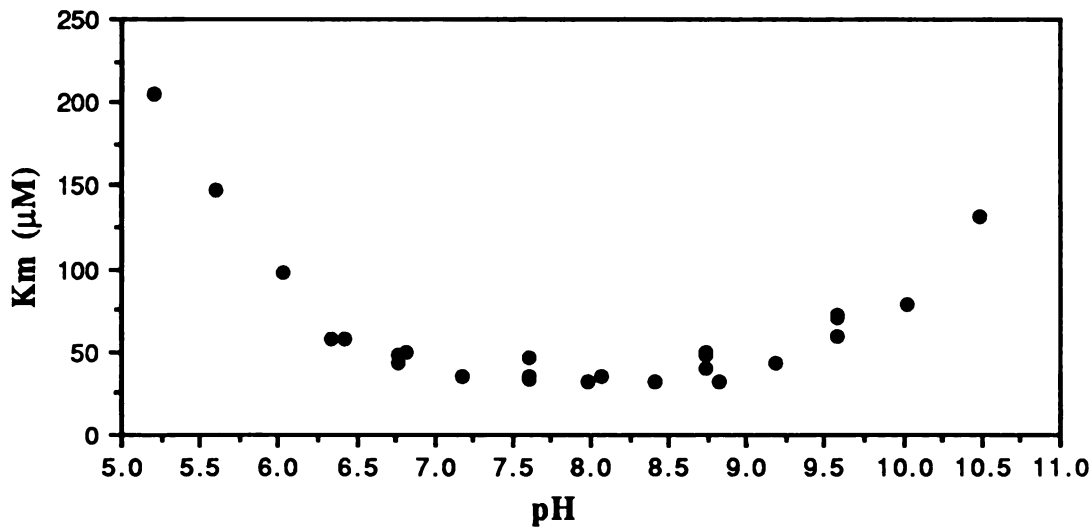
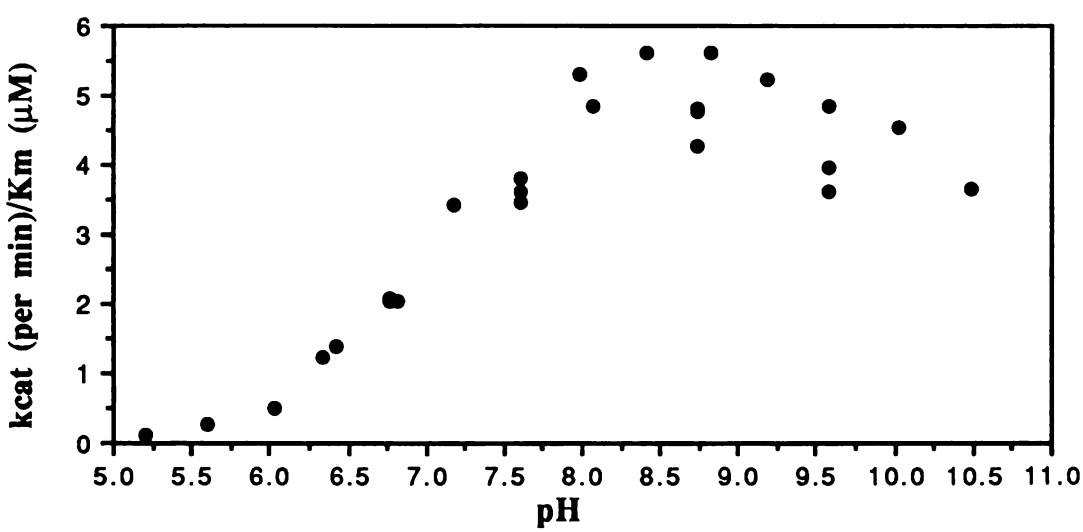

97.4
68.0
29.0
18.4
14.3
Tn

Figure 2-6. pH-activity Profile of Trypsin D189C

Each point represents a single determination of the kinetic parameters for trypsin D189C on carboxybenzyl-Gly-Pro-Arg-AMC (see Table 2-1 for a complete explanation of kinetic determination). The carboxybenzyl blocked substrate was used because a pH activity profile for trypsin had been completed (Vásquez and Craik 1990). The buffer system was as described (Craik, Roczniak et al. 1987; Vásquez and Craik 1990) and no buffer effects were noted. 25 data points determined with 18 different buffers define the trends over the pH range of 5.2- 10.5. The buffers, all at 50 mM with 100 mM NaCl, 1 mM $CaCl_2$ were MES at pH 5.2, 5.6, 6.0, 6.4, 6.8, MOPS 6.3, 6.8, 7.2, 7.6, 8.0, TAPS 8.1, 8.4, 8.8, and Glycine 8.7, 9.2, 9.6, 10.0, 10.5. The top graph is a pH-kcat profile; the middle graph is a pH-Km profile; and the bottom graph is the pH-kcat/Km profile. The scatter of the data in the kcat/Km profile is large, but all determinations at the same pH are within 20% of each other and most are within 10%. In cases in which the difference was larger than 10%, additional determinations were done and the value for that point is expressed as the average.


kcat (per min)
500
400
300
200
100
0
5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 11.0
pH


Km (μM)
250
200
150
100
50
0
5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 11.0
pH


kcat (per min)/Km (μM)
6
5
4
3
2
1
0
5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 11.0
pH

Km was only 13.5 μM. This decrease in Km with the tosyl substrate is consistent with that observed for trypsin (4.8 and 18 μM) and may reflect strong interactions at the P4 site for the tosyl group.

The pH-kcat/Km profile shows that at low pH, the enzyme activity is extinguished by a low rate and a high Km. At high pH, the increase in the Km dominates the rise in rate and the activity drops. Consequently, the pH activity profile is bell shaped. There appears to be one inflection point in the pH-activity profile at approximately pH 7.0. The Km is fairly constant from pH 6.9-8.9 across the range in which the sulfur might be expected to titrate. It is possible that the large increase in Km as the pH decreases from 6.9 is a function of the titration of the cysteine sulfur with a pKa that is two pH units lower than normal. There appears to be one inflection point in the kcat profile which may be related to the protonation of histidine 57 below pH 7.0 or to the protonation of the cysteine 189 or to both. The increase in rate at alkaline pH is not observed in the wild type enzyme and the mechanism of the effect for trypsin D189C is unknown.

However, it is critical to establish that the sulfur is reduced; the oxidized forms of cysteine are negatively charged but do not have a pKa in the appropriate pH range. Consequently, these derivatives may retain some catalytic activity but they would not titrate. *p*CMB modification of the enzyme was attempted although the primary specificity pocket may be too narrow for this reagent. We anticipated that the covalent modification of a reduced sulfhydryl atom would diminish activity as the volume of the pocket would be

altered and the side chain neutralized. After incubation of crude extract with 2.2 mM *p*CMB, trypsin D189C retained fully 25% of its activity while trypsin retained 80-90% of its activity as determined by microplate assay. *p*CMB modification of trypsin was also completed on affinity-purified trypsin D189C; no measurable kinetic effect on hydrolysis of Tos-GPR-AMC, even at 2.2 μM *p*CMB, was observed.

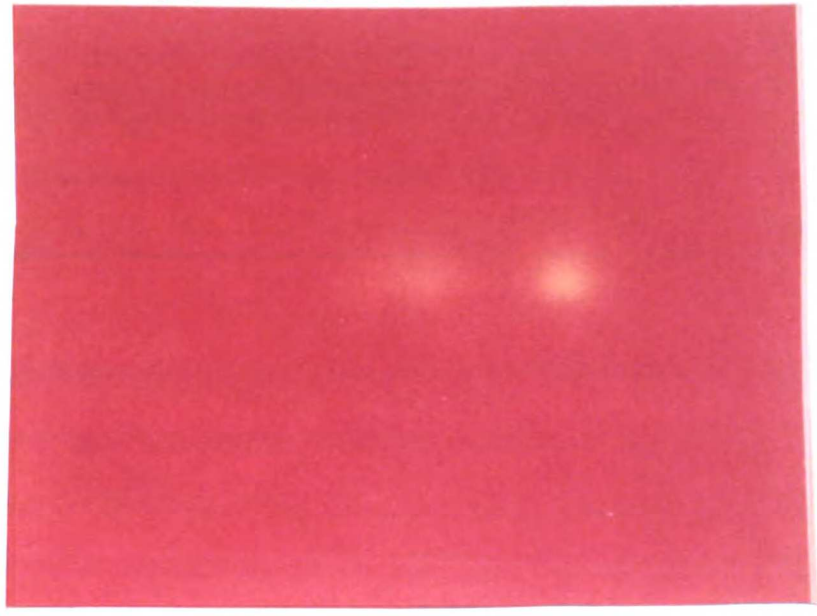
New Substrate Specificities

To evaluate the specificity of the pocket after substitution at position 189, the 18 trypsins were tested against 2 additional amide substrates with Ala and Phe at the P1 positions and 5 additional ester substrates with Ala, Phe, Leu, Ser, and Glu at the P1 positions. Trypsin was inactive on the seven substrates. However, trypsin D189A was active on the Phe ester and amide substrates as tested by activity gel analysis (Figure 2-7) and microplate analysis. Trypsin D189S was active on the Phe ester substrate (Figure 2-7). All other enzyme-substrate combinations were observed to be inactive. The assays with gels containing glutamic acid ester substrate were probably less sensitive than the other assays since the entire gel sandwich drifted to low pH over 3 hr. This pH drop is presumably due to the pKa of the glutamate side chain.

Figure 2-7. Gel Activity Analysis with a Phenylalanine Substrate

The specificity of 18 trypsins differing at position 189 were examined using activity gels based on 5 different single residue ester substrates- Ser, Leu, Glu, Ala, and Phe. The buffering capacity of the glutamate side chain towards its pKa of approximately 4.0 substantially diminished the color discrimination, and, therefore, the sensitivity in this application of the single substrate gel overlay activity assay. In the other gels, a sensitivity similar to that observed for Arg and Phe was anticipated (Chapter One). The data shown is for the phenylalanine activity gel which is the only substrate on which any of the trypsins were active. The yellow bands indicate activity. Only three samples were loaded on this gel although all had been tested. The leftmost band is the hydrolysis from the equivalent of 1.5 ml the trypsin D189A extract. The center lane is a negative control, trypsin D189K. In the rightmost lane, is the extract from 3 ml of trypsin D189S.

Phenylalanine Overlay Gel
Position 189 Mutants



DISCUSSION

We have investigated the function of the aspartate at position 189 at the base of the trypsin substrate binding pocket using a set of amino acid replacements. Among the trypsin mutants, the two most active mutants, D189C and D189E retained 1.4% and 0.01% of the activity of trypsin. Five Arg-specific trypsins were identified in all including trypsins D189N, D189A, and D189S.

Crystallographic solutions of the three dimensional structure of trypsin bound to an arginine analog include the solutions with benzamidine (Kreiger, Kay, et al. 1973) and with the K15R mutant of BPTI in the presence of the Val-Val dipeptide (Bode, Walter et al. 1984). Both complexes are approximate models of the ground state enzyme-Arg substrate interaction (see Figure i-4 for reference)[28]. The three dimensional structure solutions of trypsin indicate that hydrogen bonds are formed from the position 189 Oδ1 to serine Oγ and O, to glycine 216 O, and from the position 189 Oδ2 to glycine 219 O, and to glycine 216 O. This network is likely to be perturbed when the carboxylate is displaced (trypsin D189E), substituted (trypsin D189C), or eliminated as in the weakly active position 189 mutants (trypsins D189N, D189A, D189S). These changes would be expected to affect protein function and may be manifested in either or both the Km and kcat. Two direct hydrogen bonds, and two additional indirect ones, are formed between the D189 carboxylate and the Arg guanidinium. The Km increase that is observed for trypsin D189E may be related to interference with the primary

[28]The length of benzamidine is very similar to that of arginine as measured from the ß carbon (Mares-Guia et al, 1965)

enzyme-substrate hydrogen bond while the kcat decrease may interfere with the additional stabilization that is achieved as the reaction prgresses[29]. The magnitude of the effect on kcat/Km observed for trypsin D189E (0.01%) is generally consistent with the results obtained using a homoarginine ester substrate (0.10%) (Baines, Baird et al. 1964). Also, a 1 x 10^3 increase was observed in the Ki for acetamide relative to benzamidine (Mares-Guia and Shaw 1965). It should be noted that we assume the structural changes are limited to the substituted position and local perturbations. Additional structural and functional characterization is necessary to verify these assumptions.

The negative charge at position is critical for function. Loss of the charge leads to a loss of activity of 5-6 orders of magnitude. Also, the only two amino acids that are negatively charged in the active pH range do sustain activity. However, it is difficult to measure the contribution of the negative charge in the absence of steric considerations. Thus, the results of an asparagine substitution for aspartate are difficult to interpret. However, it appeared that trypsin D189C would provide a unique opportunity.

Cysteine is the only amino acid other than histidine that has a pKa in the range in which trypsin is active. Examination of the kinetic parameters across a range of pH might allow measurements of the activity of a single trypsin structure with both a charged and

---

[29]This is speculative. However, it is not generally understood how changes in kinetic constants can be interpreted in structural terms. In this situation, it may be that as the reaction progresses the angles of the hydrogen bonds change, strengthening the interaction, and that some of the stabilization of the transition state of the reaction comes from this reorientation in the binding pocket.

uncharged amino acid at the base of the S1 pocket. Based on the results with the substitution of uncharged amino acids at position 189, it was expected that the kcat would decrease and the Km increase at lower pH (where the sulfur is protonated). However, no additional inflection points appeared in the pH profiles suggesting that the pKa of the side chain was not 8.0. The cysteine pKa may be shifted to below that of the His57 and so the activity effects are masked. In the acidic limb of the Km profile, an increase is observed beginning at pH 6.8 which might be related to the protonation of this low pKa sulfur. Alternatively, the mutant enzyme may be unfolding at low pH. Results with trypsin D189K indicated that the environment at the base of the pocket has evolved to stabilize an anion. Furthermore, since cysteine 189 would be adjacent to the desolvated positive charge of the Arg in the enzyme substrate complex, ensconced in the low dielectric field of a protein, a large shift in the pKa is possible.

One curious feature of these studies is the low Km observed for trypsin D189C on the Tos-GPR-AMC (13.5 μM) and Z-GPR-AMC (35 μM) substrates at pH 8.0, 26°C which would suggest that the S1 pocket is fully functional (Km is a good approximate to Ks under standard assumptions about amide bond hydrolysis). Nonetheless, the enzyme binds 4-methylumbelliferyl-*p*-guanidino-benzoate (MUGB) poorly and is only weakly bound by benzamidine. The *p*-guanidino-benzoate is so tightly bound as to be a good active site titrant (the kcat for the ethyl ester is 0.0046 sec^{-1} which is 3000 fold reduced as compared to TAME (Mares-Guia, Shaw et al. 1967)) and the Ki for benzamidine inhibition of 210 μM benzoyl-arginine-

p-nitroanilide is 18.4 μM[30] at pH 8.0 and 15°C (Mares-Guia and Shaw 1965). Although a Ki is not related to substrate binding but is an independent constant describing the binding of the pseudo-substrate, the weak binding of trypsin D189C to these inhibitors is indicative of changes in the functioning of the S1 pocket. Additional structural and functional characterization is required to take advantage of the underappreciated chemical properties of cysteine.

Gráf and colleagues (Gráf, Hegyi et al. 1988) published kinetic experiments performed at pH 9.0 where they anticipated that a Lys substrate would be partially deprotonated and neutral. The elevated pH studies were novel trypsin experiments. Bovine cationic trypsin losses activity at high pH while rat anionic trypsin does not. Although significant effects were not reported for activity of trypsin, a 20-30 fold increase in the trypsin D189S rate at the elevated pH was noted. However, a single pH instead of a pH profile was examined. Consequently, the titration of the Lys amino group cannot be followed and it is unclear that the Lys substrate was in fact neutral. At pH 9.0, only half the lysine side chains are predicted to be deprotonated and the local environment of the base of the pocket clearly favors interaction with a cationic substrate. Furthermore, the experiments reported were also conducted with Arg substrates which would be positive across the pH range

---

[30] Recent experimental work over a wide range of benzamidine concentrations has shown that there is a second binding site in trypsin with high apparent Ki. The binding of the benzamidine molecules is cooperative and trypsin activity decreases in a non-linear manner as inhibitor concentration increases (Mares-Guia et al 1980).

examined and yet, similar increases in the rate of hydrolysis are observed at the higher pH. It appears that there is linear dependence of rate and hydroxide concentration (log k_{obs} vs. log $[H^+]$) indicating that the increase in the rate of hydrolysis may be non-enzymatic.

In α-lytic protease and subtilisin, two serine proteases that are not in the trypsin family of enzymes, the substrate specificity has been shown to be a malleable property (Estell, Graycar, et al. 1986, Wells, Cunningham, et al. 1987; Wells, Powers et al.1987). Indeed, the substitution M192A redirected the specificity of α-lytic protease to Met and away from Ala although the activity against a range of substrates was elevated for the mutant (Bone, Silen et al. 1989)[31]. Similarly, the pocket of subtilisin can be tailored to be specific for smaller hydrophobic groups by increasing the size of the amino acid at position 166. Furthermore, the pocket could be made specific for Lys by substituting a negatively charged amino at either position 156 or 166 or both (Wells, Powers et al.1987). Such results have led some to propose that there is flexibility in the geometry of a given site such that specificities can be readily programmed (Bone, Silen et al. 1989). Previous results (Gráf, Craik et al. 1987; Gráf, Jancsó et al. 1988) which have been confirmed and extended in our experiments suggest that the trypsin binding

---

[31] For α-lytic protease, the kcat/Km for a tetrapeptide Suc-Ala-Ala-Pro-Ala substrate is 21,000 $min^{-1}\mu M^{-1}$ and for a cognate Met substrate is 18000 $min^{-1}\mu M^{-1}$. For α-lytic protease M213A the activity toward the Ala substrate is 600 $min^{-1}\mu M^{-1}$ and 980 $min^{-1}\mu M^{-1}$ towards Met. For α-lytic protease M192A the activity toward Ala is 10,000 $min^{-1}\mu M^{-1}$ and 35,000 $min^{-1}\mu M^{-1}$ towards the Met substrate.

pocket cannot be readily re-directed to a new specificity. A critical determinant of specificity, Asp189, was changed to 17 other amino acids and the mutant proteins were examined for specificity toward five new substrates Leu, Ser, Glu, Ala, and Phe. Only weak activity was observed and only toward the Phe substrate.

Apparently, the binding pockets of these three serine proteases differ in important aspects that define the specificity. One possibility is that in contrast to trypsin, the broad hydrophobic clefts that define the α–lytic protease and subtilisin pockets may allow solvent access to the substrate. The deleterious effect of unsatisfied hydrogen bonds might then be masked. Also, these enzymes may not rely as trypsin does on strong binding of the substrate in the ground state. For example, α-lytic protease exhibits Km values larger than 1 mM even with a tetrapeptide substrate (Bone, Silen et al. 1989). It has been pointed out that enzymes evolve to higher Km given a constant kcat/Km provided that the substrate is available at a substrate concentration initially greater than Km (Fersht 1985).

Studies are in progress to identify a suppressor for the loss of the activity in the position 189 mutants. This might more fully explain the function of the native state. Additionally, the ability of the enzyme to utilize both Arg and Lys substrates is an unexamined issue and probably relates directly to the evolutionary choice of aspartate for position 189.

Chapter Three

Development and Application of a Selection for Protease Activity to the Study of Trypsin Substrate Specificity

ABSTRACT

Current efforts to understand the relationship of protein structure and function could be markedly enhanced with a means to examine a range of amino acid substitutions at a range of positions. To this end, we have developed a genetic selection for protease activity in *E. coli* and have applied it to the study of substrate specificity in the serine protease trypsin. The selection relies on tryptic hydrolysis of arginine-ß-naphthylamide to relieve arginine auxotrophy of the host. Cells expressing trypsin D189S (kcat/Km=0.007% of trypsin) survived while cells harboring trypsins still less active did not, indicating that the dynamic range of the selection spanned over four orders of magnitude in proteolytic activity. The amino acids at positions 189 and 190 which exert a very strong influence on the substrate specificity of trypsin were mutated to all 400 possible amino acid combinations by oligonucleotide-directed mutagenesis of the trypsin gene. Characterization of the library revealed that it was representative although there existed a skewed distribution of the mutants in the library. Three searches of this library, encompassing 90,000 transformants, yielded potential positives whose trypsin activity was confirmed by screening. Sequence analysis of these positives revealed 16 different trypsins including wild type trypsin. The collection of mutants indicates that a negative charge at one of the two positions is critical for specificity. The results also suggest that the character of the side chain at position 190 (serine in trypsin) is unimportant although the volume of that functional group is constrained.

INTRODUCTION

A genetic selection provides the means to search a large number of protein structures for one that satisfies a functional requirement (Helinski and Yanofsky 1963). The power of selections is routinely applied to the study of a host of biological issues. However, only recently has it been used to address the protein structure::function relationship in model systems in which the three dimensional structure of the protein is known. The molecular interactions between the *trpA* aporepressor and the tryptophan co-repressor as well as the interactions between the repressor and its cognate DNA operators have been assessed by genetic selection (Kelley and Yanofsky 1985). This information has been instrumental in interpreting the structural information that emerged from X-ray crystallographic analysis of both the aporepressor and the repressor (Schevitz, Otwinowski et al. 1985; Zhang, Joahimiak et al. 1987). A set of amino acids in the l repressor distant from the face that interacts with the DNA and a set of core residues were randomly mutated. A compilation of the functional replacements at these positions revealed that many amino acids are important only as place holders in the structure (Reidharr-Olson and Sauer 1988; Bowie, Reidharr-Olson et al. 1990). Recently, a genetic selection has been applied to triose phosphate isomerase in an attempt to find an alternate solution to the perfectly evolved arrangement of its catalytic amino acids (Hermes and Knowles 1987;Hermes, Blacklow et al. 1990).

Enzyme specificity results from interactions between the ligand and the enzyme including the main and the side chain atoms of the

enzyme, bound water molecules, and the cumulative electrostatic forces generated by the protein. Trypsin is somewhat more specific for Arg substrates, but efficiently cleaves after Lys substrates as befits a general digestive protease (Kraut 1977; Steitz and Shulman 1982). Kinetic analysis has revealed that the trypsin pocket comprises a hydrophobic slit adjacent to an anionic site (Mares-Guia and Shaw 1965). Crystallographic solution of the three dimensional structure of chymotrypsin (Matthews et al. 1967) and subsequently of trypsin (Kreiger, Kay et al. 1974; Ruhlmann, Kukla et al. 1973) identified the aspartic acid 189 as the anionic determinant of the pocket. Mutagenesis of this position has revealed that the aspartate is necessary for specificity toward Arg and Lys substrates. However, substitution at this position with a cationic amino acid (lysine) did not reverse the specificity, indicating that the amino acid at position 189 is not the only factor in the substrate specificity (Gráf, Craik et al. 1987; Gráf, Hegyi et al. 1988; Gráf, Jancsó et al. 1988). We seek an improved understanding of how substrate specificity is achieved in trypsin. To this end, a technique for selecting active proteases in *E. coli* has been developed. A library of mutations encoding the possible permutations of amino acids at positions 189 and 190 has been generated by site-directed mutagenesis and this library has been characterized. Using the genetic selection, 16 different trypsins that retain at least partial activity have been isolated in a search through this library.

MATERIALS AND METHODS

Materials

Restriction enzymes, T4 DNA ligase and T4 DNA polymerase were purchased from New England Biolabs Inc. (Beverly, MA) or Promega Inc. (Madison, WI). Modified T7 DNA polymerase was purchased from United States Biochemicals Corp. (Cleveland, OH). Isopropyl ß-D-thiogalactopyranoside (IPTG), Nα-*p*-tosyl-arginine methyl ester (TAME), and 4-methylumbelliferyl-*p*-guanidino benzoate (MUGB) were purchased from Sigma Chemical Company (St. Louis, MO). Arginine ß-naphthylamide, Nα–carboxybenzyl-Gly-Pro-Arg-7-amido-4-methyl coumarin, Nα-*p*-tosyl-Gly-Pro-Arg-7-amido-4-methyl coumarin, Nα-*p*-tosyl-Gly-Pro-Lys-7-amido-4-methyl coumarin, and *t*-butyloxycarbonyl-Leu-Gly-Arg-*p*-nitroanilide were purchased from Bachem Bioscience Inc. (Philadelphia, PA). 4-amino-7-methyl coumarin (AMC) was purchased from Enzyme Systems Products (Livermore, CA).

Bacterial strains and Plasmids

Escherichia coli strain X90 (F' lac I^q, lacZY, pro AB/ Δ(lac-pro), ara, nalA, argE(am), thi, rif^r) (Amann, Brosius et al. 1983) and SY903(F' lac I^q, lacZ::Tn5, proAB/ recA1, srl::Tn10, Δlac-pro, rif^r, nalA) were gifts from Dr. A. Vershon; CJ236 (F' CJ105/dut, ung, relA) was furnished by BioRad Laboratories Inc. (Richmond, CA). *E. coli* LE112 and LE113 (F' lac I^q, lacZ::Tn5, proAB/ dut1, ung1, relA1) contained the episome from strain SY903 and the chromosome from strain CJ236.

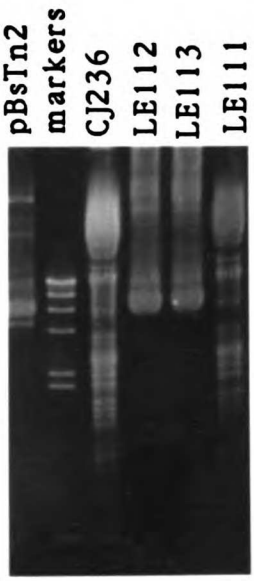
The creation of LE112(113) was required to retain the ability to mutagenize and express trypsin in the same vector; CJ236 could not generate uracil-laden single stranded p^mT3. Curing of pCJ105 from CJ236 was achieved by repeated growth to saturation without selection for the episome (chloramphenicol), and was confirmed by plating to LB and replica plating to LB plates containing 15 μg/ml chloramphenicol. The modified CJ236 and SY903 (F^+) strains were mated, and the new *dut, ung* strain was identified by the ability to grow on minimal media in the presence of 10 μg/ml kanamycin and to hydrolyze X-gal. The ability of LE112 (113) to produce uracil-laden single stranded DNA from a vector containing trypsin regulated by a *tac* promoter was confirmed (Figure 3-1). The episome from SY903 carries the $lacI^q$ marker which may be the important difference between CJ236 and LE112.

Three vectors were used in the bacterial expression and selection of trypsin, pT3, p^mT3, and pLIB. Overexpression of rat anionic trypsin (Craik, Choo et al. 1984) was achieved using a *tac* promoter and the *his J* signal peptide in pT3 (Higaki, Evnin et al. 1989). p^mT3 was created by replacing the 702 bp Sca I-Eco RV DNA fragment from pT3 with the 956 bp Sca I-Pvu II DNA fragment from pBluescript-minus (Stratagene Inc., La Jolla, CA) which contains the M13 origin of replication. pLIB was constructed from p^mT3 by eliminating the unique SphI site and deleting eight bases in the coding sequence fo trypsin. The terminal two bases of the position 188 codon and all six bases encoding the amino acids at positions 189 and 190 were deleted using the oligonucleotide 5'-GAGGGAAGGCATGCCAGGGCG-3'. The two underlined bases flank the

Figure 3- 1. Isolation of Uracil-laden Single Stranded DNA

A new strain with the *dut* and *ung* markers was created to preserve the ability to mutagenize and express in the same vector as CJ236 proved to be inadequate for this job with the new expression vector. Isolates of the strain obtained from mating SY903 to CJ236 were numbered LE111, LE112, and LE113. The vector containing the *tac* promoter and a fragment of the hisJ structural gene upstream of the alkaline phosphatase, trypsin, *trpA* terminator expression cassette was used in these trials (pBsTn2/*tac* is larger by 430 bases than the parental pBsTn2 vector). In the control lane (leftmost), single stranded pBsTn2 from strain DG98 is shown and the closed covalent single stranded DNA co-migrates with linear double stranded DNA that is 6.6 kb in length. Markers are formed from Hind III-digested double stranded λ DNA (23, 9.3, 6.6, 4.4, 2.3, 2.1, and 0.6 kb). The DNA isolated from strain CJ236 pBsTn2/*tac* is shown, and note that intact single stranded plasmid is <u>not</u> present. The products of strains LE112, LE113, and LE111 are displayed in the labelled lanes. From strains LE112 and LE113 intact plasmid DNA can be isolated. LE111 did not yield intact plasmid DNA, and the strain is probably an abberant product of the mating event.

Isolation of single stranded DNA
in tac Vector



deletion. This step incorporated a unique SphI site at the location of the deletion. pLIB does not express trypsin as determined by immunoblot analysis.

For yeast expression, two vectors were needed, pRP179, which contains the glucose regulated hybrid ADH2/GAPDH promoter and the α–factor signal peptide, and pBS24.1, an *E.coli*-yeast shuttle vector. Both were generous gifts of Dr. S. Pichuantes (Chiron Corporation). The oligonucleotide 5'-GTGACAAAAGCCCTCTAGATAAAATCGTTGGAGG-3' was used for site directed mutagenesis of both native and trypsin D189P to reconfigure the nucleotide sequence at the 5' end of trypsin to allow Xba I, Sal I cloning of trypsin and trypsin D189P into pRP179 (Pichuantes, Babé et al. 1989). Trypsin D189P contained an internal Xba I site necessitating partial digestion of this DNA. The Bam HI-Sal I 2032 bp expression cassette including the ADH2/GAPDH promoter, α-factor leader, the trypsin gene, and the α-factor transcriptional terminator from each of three intermediate constructs (two for trypsin and one for trypsin D189P) were cloned into the 12.8 kb Bam HI-Sal I digested pBS24.1. Transformation and expression were performed in the protease deficient *Saccharomyces cerevisiae* strain AB110 (MATa, leu2-3 112 (defective allele), ura3-52, pep4-3, his4-580, [cir^0]) as described (Barr, Cousens et al. 1988).

pE318 vector, a derivative of the pEZZ318 plasmid was a gift of Dr. B. Nilsson (Royal Institute of Technology, Sweden). A synthetic Eco RI- Hind III adaptor was synthesized to allow insertion of unique Hind III and Sac I sites while destroying the existing Hind

III site (sense strand was 5'-AATTAAGCTTCTCGAGGAACGAGCTC-3', and the anti-sense strand was 5'-AGCTGAGCTCGTTCCTCGAGAAGCTT-3'). The trypsinogen gene present in a Hind III-Sac I DNA fragment from pGAP-Tg (Evnin and Craik 1988) would be cloned into the modified pE318. The amino terminus of the excreted trypsinogen will be non-native, but the protein will retain the enterokinase cleavage site (Asp)4-Lys preceding Ile16. Site-directed mutagenesis of the single stranded form of this pE318-trypsinogen vector could be used to eliminate the propeptide sequence allowing direct expression of trypsin.

Library Construction and Selection

Mutagenesis of pLIB with a pool of oligonucleotides comprising all possible nucleotide combinations at the positions encoding amino acids 189 and 190 (5' CTA GAG GGA GGC AAG NNN NNN TGC CAG GGC GAC TCT GGT GGG CC 3') was as described (Evnin and Craik 1988) except that uracil-laden single-stranded DNA was isolated from strain LE112. Electrotransformation into a concentrate of *E. coli* X90 cells was achieved with narrow-gapped cuvettes (0.2 cm) and the Bio-Rad (Richmond, CA) Gene Pulser equipped with a Pulse Controller (Dower, Miller et al. 1988). The pulse is an exponential wave form which decays as a function of the resistance and capacitance. With settings of 2.5 kV, 25 μF, and 100 or 200 ohms, transformation efficiencies of 1 x 10^7/ μg of supercoiled DNA were achieved with X90 cells. Electro-competent cells were prepared as described (Dower, Miller et al. 1988). In brief, one liter of X90 cells were harvested in the logarithmic stage of growth (A_{600}= 0.5), then

concentrated and washed twice by centrifugation and resuspension in ice cold water. The cells were washed in 10% ice cold glycerol and finally resuspended in 1 ml of ice cold glycerol. The concentrate was stored in 100 µl aliquots at -70°C.

Four µl of the mutagenesis reaction was incubated with 80 µl of concentrated cells for 60 sec in a sterile cuvette and the mixture was electrocuted. The electroshocked cells were agitated in an air shaker at 37°C for 45 min in SOB media (LB, 20 mM glucose, 10 mM $MgSO_4$). The transformed cells were plated onto nitrocellulose discs overlaid on LB plates containing 50 mg/L ampicillin. Plasmid DNA was harvested by an alkaline lysis miniprep procedure (Birnboim and Doly 1979) from 2.5 x 10^5 colonies. This DNA encoded the library of trypsin mutants. Prior to transformation into X90 cells, the DNA was digested with Sph I to reduce background (unmutated plasmids retain the Sph I site as and they are linearized in this step). However, the Sph I digestion proved to be incomplete. Consequently, for the third and final search of the library, a new pool of DNA was generated using the original library DNA. The DNA isolation procedure was identical to that used initially. However, the source of the DNA for the transformation was the original library and not the mutagenized template. Also, this second DNA preparation was purified by density gradient centrifugation. Sph I digestion of this purified DNA was more readily achieved and appeared to be more complete.

In all cases, selection was initiated by electro-transformation which helped ensure a clonal population of each mutant in a colony. Colonies obtained directly from mutagenesis can be impure because

the mutant heteroduplex may be replicated prior to its repair. Also, two colonies may be superimposed. One potential problem that has been encountered elsewhere (D. Santi, personal communication) is the formation of concatemers of the parental and mutant plasmid. In this instance, subsequent generations of colonies would be impure in spite of retransformation.

For selection, a density of 5-10 x 10^3 transformants per 15 x 100 mm petri dish was desired. To achieve this, the transformation required dilution using 0.1 M $MgSO_4$. Dilution in LB results in the addition of sufficiently large source of arginine to defeat the selection.

Screening

Microplate screen (Chapter Two). Assay of the trypsin activity in periplasmic extracts was done directly in a microplate. The release of *p*-nitroaniline from tBOC-Leu-Gly-Arg-*p*-nitroanilide was monitored at 405 nm (Chapter three). The activity of endogenous *E.coli* proteases was determined using an extract from cells harboring pLIB. Low levels of activity were confirmed by following *p*-nitroaniline release over 1-2 hr.

Activity gel screen (Chapter One). Trypsin activity from putative positive isolates was assayed using a single substrate overlay gel as described (Evnin and Craik 1988). In brief, the proteins of the periplasmic extract are separated by SDS-PAGE, the gel is treated to remove the SDS and the Tris-HCl buffer and trypsin activity is monitored by observing a color change in a polyacrylamide gel overlay impregnated with an ester substrate, *N*α-*p*-tosyl-arginine

methyl ester (TAME), and the pH indicator phenol red. Endogenous esterase activity is undetectable in this assay.

RESULTS

Selection for Trypsin Activity

The genetic selection relies on the demonstrated ability of *E. coli* to secrete active trypsin into the periplasm (Evnin and Craik 1988), and on the inability of *E. coli* to metabolize a ß-naphthylamide derivative of an amino acid (Miller, Green et al. 1982). Consequently, Arg-ß-naphthylamide cannot be an anabolic source of arginine for *E. coli*. X90 cells are auxotrophic for Arg since they lack one of the enzymes required for Arg biosynthesis. Trypsin readily cleaves the arginyl naphthylamide bond to release arginine. Thus, when X90 cells are grown on a medium containing Arg-ß-naphthylamide as the only source of Arg, they will survive only if they express active, Arg-specific trypsin (Figure 3- 2).

Examination of the survival of X90 pT3 cells on M9 (Miller 1972) and MOPS (Neidhart, Bloch et al. 1974) minimal medias revealed that X90 cells grew somewhat better on MOPS. The presence of the transition state metals Fe, Cu, Co, Mn, and Zn and the particular sources of phosphate, sulfur, and nitrogen may be important. Optimal discrimination was achieved in the absence of arginine supplementation although the time required to achieve discrimination over background growth was reduced from roughly 72 to 48 hr by addition of 1 µM Arg to the plates. The long growth times required were partly attributable to the adaptation the cells must make to growing on minimal media. Free Arg supplementation at levels greater than 5 µM prevented discrimination between background colonies expressing inactive trypsin and survivors expressing active trypsin. Although greater

Figure 3-2. Selection for Active Trypsin Mutants

The trypsin is secreted to the periplasm {1}. The amino acid required for survival, arginine, is present in the solid media as the ß-naphthylamide blocked derivative. It is presumably at equilibrium between the cytoplasm and the periplasm as it can cross the bilayer. Trypsin attacks the scissile bond of the Arg-X which is Arg-ß-naphthylamide {2}. Hydrolysis releases the amino acid which is carried to cytoplasm by specific carrier proteins or by diffusion across the inner membrane {3}. Once inside the cell, the arginine relieves the auxotrophic bacteriostasis.

CYTOPLASM
PERIPLASM
Arg--X
Arg--X
Arg--X
Arg--X
Arg--X
Arg--X
① Trypsin
② Trypsin
③ Arg

sensitivity could theoretically be obtained at higher concentrations of Arg-ß-naphthylamide, the compound retards *E. coli* growth. For cells producing wild type trypsin, it was observed that 300 µM Arg-ß-naphthylamide led to optimal discrimination. Relative to incubation at 37°C, incubation of the plates at 32°C substantially enhanced the discrimination.

Arg-Arg-ß-naphthylamide was also demonstrated to be an effective substrate for selecting Arg-specific trypsins. The Arg-Arg dipeptide is a good source of Arg for *E. coli.* In principle, twice as much Arg should be released for the cell using the Arg-Arg-ß-naphthylamide substrate which would allow discrimination at a lower concentration of the substrate relative to the Arg-ß-naphthylamide. However, the background of free Arg is likely to be twice as high and the Arg-Arg dipeptide might provide a competitive substrate for proteolysis of the Arg-Arg-ß-naphthylamide to the free dipeptide. In any case, as with the Arg-ß-naphthylamide, 300 µM appeared to be the optimal concentration for selection of X90 pT3 using the Arg-Arg-ß-naphthylamide substrate. The use of dipeptide ß-naphthylamide could, in principle, allow a single strain to be utilized in the selection of a variety of substrate specificities. Thus Arg-Phe-ß-naphthylamide would be used to search for a Phe specific trypsin but the free dipeptide would relieve the Arg auxotrophy. However, Phe-Arg-ß-naphthylamide, Pro-Arg-ß-naphthylamide, and Gly-Arg-ß-naphthylamide failed to support the growth of X90 pT3. The size and solubility of these molecules may prevent ready access to the periplasm. Consequently, the single residue substrates

coupled to the appropriate auxotrophy seem to be the best choice for selecting a particluar proteolytic specificity.

Other Selection Concepts

The substrate poly-Arg is too large to enter the *E.coli* periplasm. However, a portion of the cells present in any colony are dead and have lysed. The trypsin liberated upon cell death may be sufficient to rescue the auxotrophy by digesting the poly-Arg. Supplementation would be a critical feature of this selection. However, preliminary indications were that poly-Arg is toxic to *E.coli.*

Another approach to selection is to employ trypsin to destroy a toxin (e.g. ß-lactamase protects *E.coli* against the toxin ampicillin). Leu and Val containing di- and tri-peptides are toxic to *E.coli* because of the negative feedback inhibition on the operon that encodes the enzymes for Ile biosynthesis. The free amino acids Leu and Val do not have the same effect. As a result, the activity of trypsin towards dipeptides such a Arg-Leu and Lys-Leu could be used as a selection for trypsin activity. Bacteriostasis of SM138 pTn2 using Arg-Leu has been demonstrated (J. Vásquez, personal communication). However, the concentration appropriate for discriminating between trypsin positive and trypsin negative colonies was not identified.

CBZ-Leu-Leu had been successfully used to select for mutants in yeast that secreted carboxypeptidase Y (Rothman and Stevens, 1986). It is possible that a similar selection could function in bacteria for carboxypeptidase or a different protease. *E. coli* fail to

transport dipeptides across the cytoplasmic (inner) membrane if the amino terminus is neutral or negatively charged (Payne and Gilvarg 1971). Consequently, substrates such as Nα-carboxybenzyl-Arg-Phe fail to relieve Arg or Phe auxotrophy. However, as blocked amino acids are utilizable (Payne and Gilvarg 1971), trypsin could generate an Arg metabolite by cleavage of the Arg-Phe bond. Although, relative to the free amino acid, the blocked amino acid is an inferior source of the amino acid. Alternatively, the free Phe could be utilized. The blocked dipeptide is just small enough to enter into the periplasm through the non-specific pores of the outer plasma membrane (600 D< (Decad and Nikaido 1976)) although insolubility may inhibit transport through these aqueous channels or may dictate entry across the lipid bilayer. However, no success has been achieved using CBZ-Arg-Phe to select for active trypsins. Initial attempts with the pTn plasmid in a strain auxotrophic for Phe were unsuccessful. However, the expression levels from the strain were low, and this expression plasmid directed relatively meager expression of trypsin. Also, it should be noted that relative to the amino acid ß-naphthylamide bond the peptide bond is approximately 20 fold more stable. Subsequent attempts involving X90 pT3 relied on utilization of CBZ-Arg as a source of Arg. In these experiments, differential growth of cells expressing trypsin was not observed. Further efforts were eclipsed by the success achieved with Arg-ß-naphthylamide[32].

---

[32]A derivative of the X90 strain that is auxotrophic for Phe has been constructed by P1 transduction. This strain in conjunction with pT3 or $p^{m}T3$ might be the best candidate for subsequent efforts with this system.

A yeast selection for trypsin activity was envisioned based on the release of a required amino acid from an Nα-carboxybenzyl blocked dipeptide; a selection for carboxypeptidase Y activity in yeast utilizing a CBZ-blocked dipeptide exists (Rothman and Stevens 1986). To express trypsin in yeast, the Ile16 of trypsin and trypsin D189P, a dead mutant, were precisely fused to the α-factor leader sequence (Kurjan and Herskowitz 1982) at the terminus of the pro peptide sequences and placed adjacent to the yeast ADH2/GAPDH fusion promoter (Gardell, Craik et al. 1985; Barr, Cousens et al. 1988; Pichuantes, Babé et al. 1989). However, only the inactive trypsin D189P mutant yielded transformants; the two wild type trypsin constructs failed to transform *Saccharomyces cerevisiae* AB110. Selection using both Leu$^-$ and Ura$^-$ sorbitol media was used in an effort to isolate the wild type construction (Barr, Cousens et al. 1988). Apparently, the active trypsin protein is lethal to yeast when expressed in this fashion in spite of the α-factor leader sequence which would render the enzyme inactive. Perhaps a small portion of the trypsin is being fortuitously converted to the active protease by action of endogenous proteases, and this active trypsin is lethal. This result confirms earlier observations using a similar yeast expression system for trypsinogen (C.S. Craik, personal communication)[33].

Expression of trypsin D189P was followed for 48 hr after induction. Concentration and examination of the media revealed

---

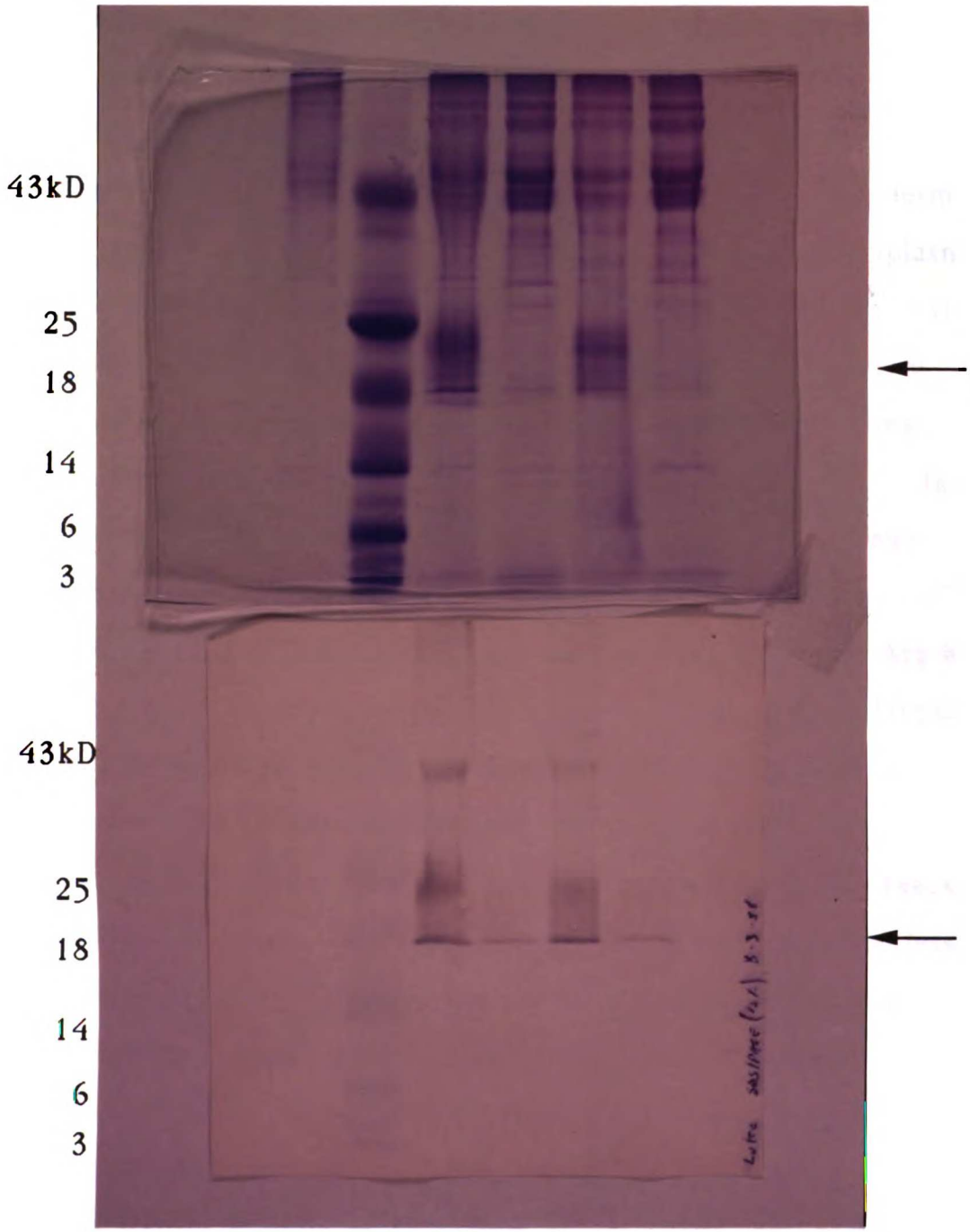
[33]Apparently, high levels of trypsinogen can be expressed from yeast using an acid phosphatase promoter (Laszlo Szilágyi, personal communication). This promoter system has generated 100 mg/ L of culture (10 OD of cells) of carboxypeptidase A (M. Phillips, personal communication).

that trypsin was readily identified by immunoblot analysis and was a visible band on a Commassie-stained SDS-polyacrylamide (Figure 3-3). Apparently, the KEX2 processing site which forms the 3' end of the leader peptide (Leu-Asp-Lys-Arg-) is appropriately recognized when trypsin is downstream. The trypsin band appears at approximately MW 19,000 because the mutant trypsinD189P migrates more quickly than trypsin (MW 24,000) in SDS-PAGE.

An alternative selection concept involves a system in which trypsin is excreted (directed to the media) in order to utilize substrates too large to diffuse into the periplasm (MW > 600 (Decad and Nikaido 1976)) (for a review of excretion from *E. coli* see Pugsley and Schwartz 1985). Purification of trypsin from the culture media may be more readily accomplished than purification from periplasmic extracts. Also, the *E. coli* inhibitor ecotin (Chung, Ives et al. 1983) may be avoided if the trypsin passed directly through inner and outer membranes to the culture media. Excretion of proteins from *E.coli* has been shown to depend on a leader peptide (Ikemura, Tagaki et al. 1987; Silen, Frank et al. 1989), perhaps in an analogous fashion to the requirement observed for secretion in the Gram positive organism *Bacillus subtilis* (Carter and Wells 1987). A general vector for excretion in *E. coli* hosts has been developed based on the properties of *Staphylococcus aureus* protein A (Abrahmsen, Moks et al. 1986; Moks, Abrahmsen et al. 1987). Bovine basic pancreatic trypsin inhibitor has also been expressed in this system by growth at 30°C (J. Altman, personal communication). The gene of interest is positioned behind the promoter, signal sequence, and the first six

Figure 3-3. Yeast Expression of Trypsin D189P

Trypsin D189P expression is controlled by the ADH2/GAPDH promoter and the α-factor terminator. Excretion of the protein is directed by the α-factor leader. Media from the cultures of two independent clones were grown in Leu- /glucose broth for 24 hr and this culture was used to inoculate a Leu^-, Ura^- fermentor medium (Pichuantes, Babé et al. 1989). After 24 and 48 hr, a portion of the two cultures were harvested and the proteins from 10 ml of media were TCA precipitated. A portion of the four sample was run through 15% SDS-polyacrylamide gels. One gel was stained by Commassie blue and the other gel was blotted for immunological detection of trypsin. Markers were 43, 15, 18, 14, 6, and 3 kD. From right to left, clone 1 at 24 hr, clone 1 at 48 hr, clone 2 at 24 hr, clone 2 at 48 hr, molecular weight markers, cells carrying pBS24.1 (plasmid without a trypsin gene).


43kD
25
18
14
6
3
43kD
25
18
14
6
3

amino acids of the E domain of proteinA. Although the construct was not completed, the vectors and oligonucleotides requisite for construction were obtained (see Materials and Methods).

Dynamic Range of the Selection

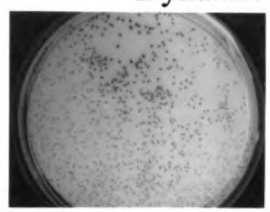
Trypsin is expressed from vector $p^{m}T3$ with an amino terminal signal peptide that acts both to direct trypsin into the periplasm and to serve as a propeptide (Evnin and Craik 1988; Higaki, Evnin et al. 1989; Vásquez, Evnin et al. 1989). To assess the level of trypsin activity required to confer survival, *E. coli* expressing trypsin or a mutant trypsin were plated on selective media. In addition to trypsin, the trypsin mutants D189E, D189C, D189S, D189N, and D189A (nomenclature is described in Knowles 1987) were sufficiently active to scavenge enough Arg from the Arg-ß-naphthylamide to allow growth of the host (Figure 3-4). Trypsin D189C has 1.4% the activity of trypsin, trypsin D189E (not shown in Figure 3-4) has 0.01%, while trypsin D189S is reduced approximately 5 orders of magnitude in its kcat/Km (Gráf, Jancsó et al. 1988). The mutants with H, I, T, or Q at position 189 are less active than trypsin D189S as determined by activity gel analysis, and failed to support growth. These tests indicated that the dynamic range of the selection spans five orders of magnitude of proteolytic activity.

The trypsin activity is necessary and sufficient for selection; no additional selection for the plasmid, such as ampicillin, is needed. *E.coli* X90 cells (400-500) expressing active trypsins (wild type and S190P) were mixed with cells (40,000) carrying pLIB (dead) and

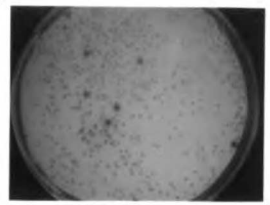
Figure 3-4. Dynamic Range of the Genetic Selection

The dynamic range was assessed by plating a mixture of cells expressing a trypsin mutant with previously defined kinetic parameters and cells carrying pLIB which cannot express trypsin. Sufficiently active trypsins allow survival of the host. Five µl of 10^{-6} dilution (roughly 25 cells) of a saturated culture of the mutant was mixed with 50 µl of a 10^{-5} dilution (roughly 2500 cells) of a saturated pLIB culture. The mixture was plated on selective plates and incubated at 32°C for 72 hr. For photographic purposes, the contrast between the colonies and the media was enhanced by spraying the plates with X-gal and allowing the action of ß-galactosidase to render the colonies blue. The large colonies indicate those cells that are expressing an active trypsin except when there is contamination or reversion of the ArgE marker (note 2 colonies on Ile189 plate, 1 on Thr189, and 1 on the His189 plate). The identity of revertants can be readily established by streaking to minimal media without any supplementation. The revertants will grow, but colonies that are surviving by expressing active trypsin will not. A shorthand nomenclature is employed here in which the amino acid at position 189 is indicated without reference to the wild type case (which is aspartate at 189 and serine at 190). In the upper left is a plate with the pLIB culture plated; here the nomenclature emphasizes that pLIB is the plasmid which has a deletion in the trypsin gene spanning positions 189 and 190. The kcat/Km, a measure of activity, of each mutant on the substrate Tos-GPR-AMC is indicated as a percentage of the wild type kcat/Km; wild type trypsin has 100% of the kcat/Km.

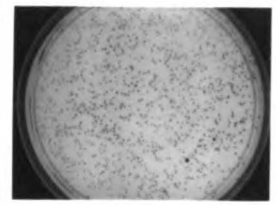
Dynamic Range of Genetic Selection for Trypsin



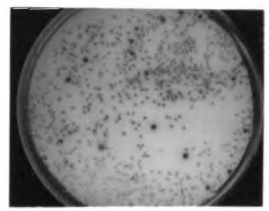
DEL (188-190)
kcat/Km= 0



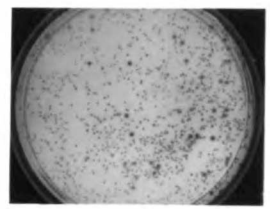
Asp 189 (wild type)
kcat/Km= 100%



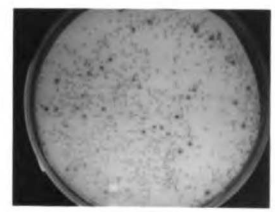
Ile 189
kcat/Km< .001%



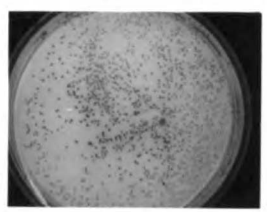
Ser 189
kcat/Km= 0.001%



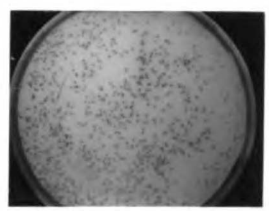
Asn 189
kcat/Km= 0.001%



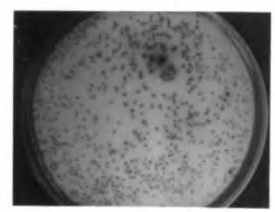
Cys 189
kcat/Km= 0.1% 1.4%



Thr 189
kcat/Km< .001%



Gln 189
kcat/Km< .001%



His 189
kcat/Km< .001%

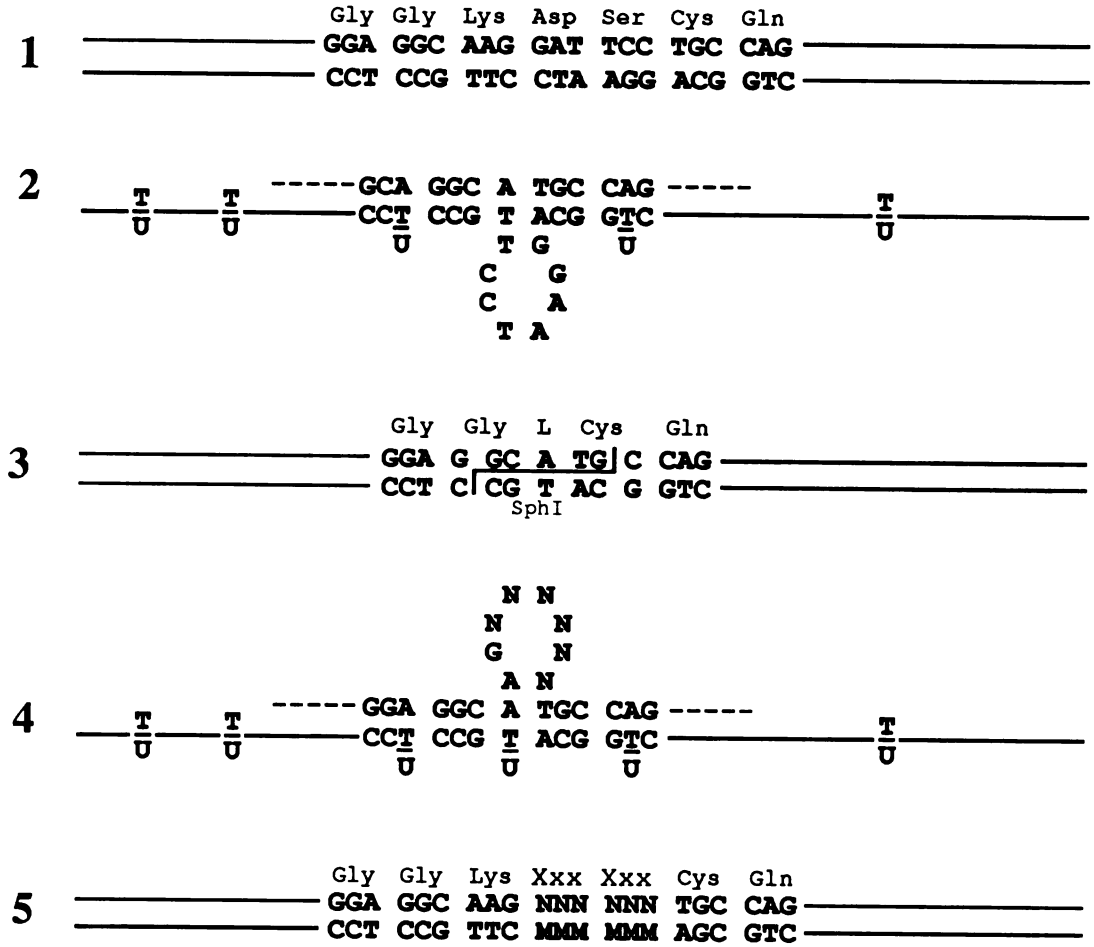
were plated to selective media with and without ampicillin. The same (+/- 20%) number of colonies of trypsin and of trypsin S190P were observed. Antibiotic is added to the selective media to prevent the growth of cells that have no plasmid but have sustained a reversion of *ArgE*(am) to a functional gene.

Generation of a Library of Mutant Trypsins

A library encoding the set of trypsins substituted at position 189 and 190 was constructed by oligonucleotide directed mutagenesis (Evnin and Craik 1988). A derivative of p^{m}T3 called pLIB was created to facilitate this work. It was anticipated that each of the oligonucleotides in the pool for randomization of positions 189 and 190 would anneal with roughly equal efficiency to the pLIB template since the oligonucleotides had an equal number of bases complementary to the template (Figure 3-5). This precaution was taken because the library of trypsins mutated at position 189 was believed to have been skewed due to bias against purine-purine base pairing in the annealing of certain oligonucleotide primers (Chapter One). The pool of oligonucleotides employed was a mixture of 4096 possible DNA combinations differing at the six bases that encoded the position 189 and 190 amino acids. These combinations encode all 400 possible amino acid combinations at the 189 and 190 pair. The mutagenesis reaction was transformed into X90 cells, plated to LB, 50 mg/L ampicillin, and the plasmid DNA from these cells was obtained. This DNA constituted the library DNA (see Materials and Methods).

Figure 3-5. Generation of a Library at Positions 189 and 190

Step one indicates the starting pT3 sequence with the wild type trypsin sequence at positions 186-192. Step two depicts mutagenesis to delete 8 bases encoding part of position 188 and all of positions 189 and 190. Only a portion of the oligonucleotide is displayed in the interest of clarity. The result of the mutagenesis is depicted in step 3. An Sph I site was created at the new junction. As the gene is out of frame, no immunologically detectable protein is synthesized from this plasmid. Mutagenesis of this vector with the pool of oligonucleotides repaired the deletion, put the gene in frame, and eliminated the Sph I site (step 4). As the nucleotides in the template stand that would have been complementary to the randomized were deleted, it was anticipated that each of the oligonucleotides in the pool would anneal with roughly equal efficiency to the template and would generate a library with an even distribution of mutants (step 5).


1
Gly Gly Lys Asp Ser Cys Gln
GGA GGC AAG GAT TCC TGC CAG
CCT CCG TTC CTA AGG ACG GTC
2
GCA GGC A TGC CAG
CCT CCG T ACG GTC
T G
C G
C A
T A
3
Gly Gly L Cys Gln
GGA G GC A TG C CAG
CCT C CG T AC G GTC
SphI
4
N N
N N
G N
A N
GGA GGC A TGC CAG
CCT CCG T ACG GTC
5
Gly Gly Lys Xxx Xxx Cys Gln
GGA GGC AAG NNN NNN TGC CAG
CCT CCG TTC MMM MMM AGC GTC

Retransformation into X90 cells and plating to LB, 50 mg/L ampicillin allowed characterization of the library.

Immunoblot analysis of 12 isolates from the rich media showed that only seven of them made trypsin (Figure 3-6). It is likely that some sets of amino acids at positions 189 and 190 will preclude expression, and, in some instances, missense mutations will be introduced. However, some of the remaining five produced no protein because they harbored the pLIB plasmid and not a mutant. (Sequence analysis of random isolates did confirm that roughly 40% of the colonies in the library harbored pLIB). Apparently, the Sph I digestion of the library DNA was incomplete otherwise all pLIB molecules would have been digested and, thus, inactivated[34]. Among the seven trypsins, several exhibit altered electrophoretic mobility (Figure 3-6). This behavior is consistent with that observed for the position 189 mutants; apparently, mutations at the base of the pocket alter the migration of the denatured protein even in 0.1% SDS (Laemmli 1970).

Although the efficiency of the mutagenesis was lower than had been previously achieved (Evnin and Craik 1988), it was adequate, and a more intensive characterization of the library was pursued. The positions 189 and 190 from 47 isolates chosen randomly from the portion of the library plated to LB, 50 mg/L ampicillin were sequenced (Sanger, Nicklen et al. 1977; Tabor and Richardson 1987) (22 were pLIB trypsin). This experiment demonstrated that a skewed library had been obtained from this mutagenesis reaction.

---

[34]The DNA for the third search was re-isolated and purified to allow complete digestion.

Figure 3-6. Immunoblot analysis of Random Isolates from the Library of Position 189 and 190 Mutants

The periplasm from 12 random isolates from the library of trypsins mutated at positions 189 and 190 was harvested and subjected to immunoblot analysis (Burnette 1981). The trypsin band is indicated. A purified rat trypsin standard was run in the rightmost lane as a marker and control. In the adjacent lane is the trypsin from a culture of wild type trypsin cells. In the other 12 lanes, the 12 random isolates were run. The bottom band in most lanes is the endogenous inhibitor ecotin. The identity of the upper band is unknown.

Tn
Ec

From statistics complied from the sequences of the 25 mutants, the nucleotides G and A are overrepresented at each position (Figure 3-7). The stability of the loop that must form in the oligonucleotide upon annealing (Figure 3-6) may be stabilized by purine bases. Using a deletion to avoid preferential annealing failed in this case. Other methods for obtaining a random library have been successfully implemented and may be useful (Reidharr-Olson and Sauer 1988; Hermes, Blacklow et al. 1990). One explanation is that the bias of unequal annealing was avoided but the biology of replication introduced an unanticipated bias. Perhaps the loops formed by a subset of the oligonucleotides were repaired prior to the repair occurring in the uracil-laden template strand of the duplex. Alternatively, the mix of four nucleotides inserted into the six randomized positions in the oligonucleotides may have been unequal.

From a statistical analysis of these results, an estimate of the likelihood of a given base at a position can be calculated. These approximate probabilities indicate the prevalence of an amino acid at a particular position (189 or 190). In a library of even distribution, wild type would be present 29 times in 10,000 mutants [$GA^{T}/_{C} * (TCN + AG^{T}/_{C})$]= [0.25*0.25*0.5* (0.25*0.25 *1 + 0.25*0.25*0.5)]. With the skewed distribution obtained, it would be present approximately 1 in 10,000. The prevalence of trypsin S190T was estimated as 0.3 in 10,000 while the prevalence of trypsin D189S/S190D and S190E were estimated as 9.4 and 42.4 per 10,000 (Table 3-1).

Figure 3-7. Oligonucleotide-directed Randomization of the Bases Encoding Positions 189 and 190.

The library was transformed into X90 cells and plated to rich media. Forty seven colonies were selected randomly and the DNA around the mutation was sequenced. Twenty five were mutants, the remainder was background. The identity of the bases at positions 189 and 190 is displayed and the amino acids encoded are translated. The number of occurrences of each base at each position were tabulated and are displayed in the "Base Composition" table as % of the total 25 occurrences. Apparently, there was a strong bias for or against certain bases or base combinations as adenosine (A) is present at least 33% of the time at all six randomized positions and 58% in the first and the fifth positions. (Although cytosine (C) did not appear at all at positions 2 and 6, it is not absent from these positions, only rare, as mutants were ultimately recovered with a C in these positions).

	position 189 190	amino acids position 189 190
1	AAG GCA	Lys Ala
2	AAA TAG	Lys Stop
3	AAA GAA	Lys Glu
4	AGT GAA	Ser Glu
5	AGA GGA	Arg Gly
6	TTA GAA	Leu Gly
7	GGA ATG	Gly Met
8	GGG AGA	Gly Arg
9	AGG AAA	Arg Lys
10	TGC ATA	Cys Ile
11	ATC AAA	Ile Lys
12	AGC GGA	Ser Gly
13	ATG GAG	Met Glu
14	CGA GAA	Arg Glu
15	AAC GAA	Asn Glu
16	AAA GAA	Lys Glu
17	AAA AGG	Lys Arg
18	CGA AAA	Arg Lys
19	TAG AAG	Stop Lys
20	GAA GCA	Glu Ala
21	TAG GGA	Stop Arg
22	AGC GAT	Ser Asp
23	GGA ATG	Gly Met
24	AGA AAA	Arg Lys
25	ATT AAA	Ile Lys

Base Composition at Each Positon

	1	2	3	4	5	6
G	17	50	25	46	21	25
A	58	33	46	46	58	71
T	17	17	8	4	13	4
C	8	0	21	4	8	0

Table 3-1. Calculated Distribution of Mutations in the Position 189 and 190 Library

Based on the number of times a base was found at a given position, the number of times that a combination of bases would be found can be calculated. In the cases where no C was found (position 2 and 6) it was assumed that C would be present 1% of the time. When the combinations are examined by codon composition, the frequency that any amino acid should be isolated, given no selection pressure, can be calculated. For example, glutamic acid is encoded by GAA and GAG and Ala is encoded by GCN. At the 189 position, GAA/G are present at a frequency of 0.17*0.33*(0.25+ 0.46) or 0.4% of all position 189 amino acids. Ala at position 190 is present at frequency of 0.46*0.08*1.0 or 3.6% of all amino acids at position 190.

position 189	calc'd frequency	**position 190**	calc'd frequency
Ala	0.002	Ala	0.097
Arg	0.246	Arg	0.269
Asn	0.056	Asn	0.060
Asp	0.016	Asp	0.036
Cys	0.025	Cys	0.005
Gln	0.187	Gln	0.093
Glu	0.039	Glu	0.013
Gly	0.085	Gly	0.256
His	0.008	His	0.037
Ile	0.074	Ile	0.015
Leu	0.034	Leu	0.045
Lys	0.136	Lys	0.006
Met	0.025	Met	0.002
Phe	0.008	Phe	0.001
Pro	0.001	Pro	0.017
Ser	0.101	Ser	0.009
stop	0.078	stop	0.014
Thr	0.006	Thr	0.001
Trp	0.021	Trp	0.022
Tyr	0.016	Tyr	0.003
Val	0.028	Val	0.005

Identification of Active Trypsins

Electro-transformation of this preparation into *E.coli* X90 cells and plating to selective media allowed the library to be searched. Three separate searches of the library, encompassing 90,000 isolates (excluding isolates that contained the unmutated, parental plasmid), were completed. Out of these, 275 were judged active after restreaking onto selective media, and the trypsin activity in each was evaluated using the microplate assay. In 111, activity was observed and the isolate was restreaked on rich media. Activity gel analysis of these re-purified positives confirmed the activity of 68 isolates. The nucleotide sequence of the mutations was determined for 55 of 68 mutants in which the trypsin activity was confirmed; 16 trypsins were identified (Table 3-2).

Of the 68 active mutants, 38 were obtained in the third of the three searches in which conditions had been optimized. In this particular search, 11,000 transformants were initially searched to generate 195 potential positives. Of these, 121 isolates were confirmed by microplate screen analysis and 38 were further confirmed by activity gel analysis. Sequence analysis of 28 of the 38 positives revealed 13 trypsins. In the two previous searches, a total of 10 trypsins had been identified. Among these 23 trypsins, there were 16 different enzymes (Figure 3-8).

One of the changes made in the optimization was to plate the transformation directly after recovery. In the initial searches, the transformed cells were stored overnight while a test of the transformation efficiency of the library DNA was completed.

Table 3-2. Trypsins Recovered from the Position 189, 190 Library

The active mutants recovered from the library are listed along with the number of times the mutant was identified. Both the activity of the mutant and its distribution in the library presumably relate to the number of times that the mutant was isolated. Assuming that trypsin D189V, G, and P have similar activities on Arg, the number of times isolated directly relates to their estimated prevalence in the library. However, it is unclear how levels of activity relate to identification since trypsin and the very active trypsin S190T were isolated a single time while the very weakly active trypsin D189N was isolated 7 times.

position 189 and 190	number of times isolated
Asp Ser	1
Asp Cys	1
Asp Thr	1
Asp Val	2
Asp Gly	6
Asp Pro	1
Glu Ser	4
Glu Thr	1
Glu Ala	7
Ser Asp	5
Ser Glu	1
Gly Asp	2
Gly Glu	7
Asn Ser	7
Asn Thr	7
Ala Ser	1

Figure 3-8. Searching the Library: A Flow Chart

The flow chart illustrates how mutants were isolated from the library. Initially, the library was plated to selective media. Large colonies were identified after a 72 hr incubation at 32°C and were streaked to selective media to confirm their ability to grow and to purify the isolate. Single colonies were picked to 2 ml LB, 50 mg/L ampicillin, 1 mM IPTG. Trypsin was harvested from these cultures and the activity was examined by the microplate assay. Potential positives were streaked to plates containing LB, 50 mg/L ampicillin plates to ensure purity. Single colonies were amplified in liquid culture and the trypsin activity was examined by the activity gel. Plasmid DNA from each of the mutants that appeared positive in this assay was harvested and the mutation was identified by dideoxy sequence analysis. The number in parentheses following each step refers to the total number of the original transformants that was handled at each step.

Searching the Library

SELECT: Isolation of active mutants (90,000)

SCREEN: Preliminary examination of activity with microplate assay (275)

RE-SCREEN: Initial characterization of protein by immunoblot and Single Substrate Gel Overlay (111)

SEQUENCE MUTANT DNA (68)

CHARACTERIZE UNIQUE TRYPSINS (16)

As several of the isolates were identified only a single time, it is clear that the library has not been searched exhaustively. Indeed, trypsin S190A which is present in trypsin homologs such as thrombin, is absent although it seems that alanine at position 190 is present in the library (trypsin D189E/S190A was isolated). A trend that emerges is that a negatively charged amino acid at position 189 or 190 is important for activity (Table 3-2). The preferred amino acid is D. The partner to the D or E amino acid is from a limited set of amino acids. The principal limitation is probably the volume available at the base of the pocket as D189 and S190 are close packed in trypsin (Kreiger, Kay et al. 1974). Three of the mutants lack a negative charge at either position, trypsin D189A, D189N, and D189N/S190T.

Confirmation of the Selection

A large number of isolates were screened. One might argue that the selection was ineffective and that the screening was the effective tool in isolating the partially active mutants. To investigate this possibility, each mutant was tested in the same manner as the position 189 mutants had been; a small number of mutants was mixed with 1000 inactive (trypsin Δ188-190) mutants and plated to selective media. Large colonies appeared on all plates; two from each were picked and the trypsin activity was assessed by activity gel analysis. All picks exhibited an activity commensurate with that previously observed.

CONCLUSIONS

From examination of the classes of mutants that have been isolated, preliminary conclusions can be advanced. The presence of a negative charge at the base of the pocket is important although either aspartate or glutamate at positions 189 or 190 suffice. Mutational analysis of positions 216 and 226 indicated that volume of the pocket was constrained. The collection of active mutants confirms this conclusion. Additional functional and structural characterization is necessary in order to dissect the important interactions that comprise substrate specificity.

It is important to note that the nature of biological systems does not enable one to draw conclusions from negative results in a screen or selection. In an ideal system, all possible mutants would be available. However, several factors contribute to a less than ideal situation including the bias of the genetic code, the dificlutly in obtaining equal substitution of the mutant nucleotide bases, and the bias introduced by the biological system (selection for or against one or more mutants). Consequently, it would be incorrect to say that the solutions identified to this functional problem of Arg and Lys binding in trypsin are the only solutions, even after exhaustive search of a complete library.

Although the library was skewed, mutants that were estimated to be quite rare were identified. However, mutants such as trypsin S190A that were expected based on the binding pockets of trypsin homologs such as kallikrein or thrombin, were not identified. Also, active mutants were not identified with more facility than inactive ones. Thus, the efficiency of the selection technique is enigmatic.

Nonetheless, the selection does succeed and should be general for any protease activity.

However, it will be necessary in all cases to identify a strain that both expresses the protease to high levels and has the appropriate auxotrophy. It is not clear whether all strains will be suitable for selection although there were no known special properties (i.e markers) of X90 that enabled the trypsin selection. P1 transduction can modify a chosen strain (such as X90) as required although a suitable donor strain must be identified. Also, the desired properties of the parent (recipient) strain must be confirmed after the transduction.

Chapter Four

An Investigation of the Substrate Specificity of Trypsin, Aided by a Genetic Selection

ABSTRACT

The structural determinants of the primary substrate specificity of rat anionic trypsin were examined using oligonucleotide-directed mutagenesis coupled to a novel genetic selection. A library was created that encoded trypsins substituted at amino acid positions 189 and 190 at the base of the substrate binding pocket. A genetic selection, with a dynamic range of 5 orders of proteolytic activity, was used to search 90,000 transformants of the library. Rapid screens for arginyl amidolysis and esterolysis confirmed the activity of the purified isolates. Trypsin and 15 mutant trypsins with partially preserved function were identified and characterized kinetically on arginyl and lysyl peptide substrates. New arrangements of amino acids in the substrate binding pocket sustained efficient catalysis. A negative charge at amino acid positions 189 or 190 was shown to be essential for high level catalysis. With the favored aspartate at position 189, several amino acids could substitute for serine at position 190. Modulation of the specificity for arginine and lysine substrates was shown to depend on the amino acid at position 190. The regulatory effect of the amino acid side chain at position 190 on the substrate specificity is also reflected in substrate binding pockets of naturally occurring trypsin homologs.

INTRODUCTION

The substrate specificity of an enzyme is determined by its capacity to form a stable complex with a particular ligand in both the ground state and the transition state. Steric constraints, intermolecular forces, and the hydrophobic effect each contribute to the stability of a given complex. Examination of the three-dimensional structures of the enzyme-inhibitor and enzyme-pseudosubstrate complexes can reveal important hydrogen bonds, van der Waal's contacts and electrostatic interactions without indicating the magnitude of these effects in a specific reaction. Furthermore, determination of how and why a particular configuration of amino acids specifies a function is not currently possible from three dimensional structure analysis alone. Critical components of substrate specificity in several enzymes have been studied by making amino acid substitutions in the structure and characterizing the functional consequences of the alterations (Craik, Largman et al. 1985; Estell, Graycar et al. 1986; Cronin, Malcolm et al. 1987; Wilks, Hart et al. 1988; Bone, Silen et al. 1989; Scrutton, Berry et al. 1990). Though powerful, this approach generally requires a prior assessment of the importance of the substituted amino acid. A genetic selection can significantly expedite analysis as it provides a means to search a large number of protein structures for those that satisfy a functional requirement making it unnecessary to economize on the number of positions or the number of mutants that may be sampled (Reidharr-Olson and Sauer 1988; Hermes, Blacklow et al. 1990).

Trypsin is a paradigm for the family of serine proteases that have evolved to cleave peptide bonds after arginine and lysine amino acids (Stroud 1974; Kraut 1977; Steitz and Shulman 1982). The substrate specificity exhibited by this enzyme family toward these two structurally disparate, positively charged amino acids is defined by the topography of the enzyme, particularly the primary specificity pocket. Crystallographic analysis of trypsin-inhibitor complexes (Kreiger, Kay et al. 1974; Ruhlmann, Kukla et al. 1973; Bode, Schwager et al. 1978) and mutagenesis of amino acids that comprise the substrate binding pocket (Craik, Largman et al. 1985; Gráf, Craik et al. 1987; Gráf, Jancsó et al. 1988) indicated that two positions, 189 and 190[35], were critical in defining the substrate specificity of the enzyme. The remainder of the interactions between the enzyme and the substrate side chain are mediated by main chain atoms of the three ß strands that define the pocket (refer to the Introduction and Figures i-1, i-3, and i-4).

We wished to examine the functional contribution to arginine and lysine specificity of the two amino acids at the base of the substrate binding pocket. A library encoding trypsins substituted at positions 189 and 190 was constructed and a genetic selection was employed to search through this collection of mutants. A set of mutant trypsins with partially preserved function was isolated and kinetically characterized to investigate the components of Arg and Lys specificity.

---

[35] Amino acids in trypsin are numbered according to convention based on the numbering of chymotrypsinogen (Hartley et al. 1967)

MATERIALS AND METHODS

Materials

Restriction enzymes, T4 DNA ligase and T4 DNA polymerase were purchased from New England Biolabs Inc. (Beverly, MA) or Promega Inc. (Madison, WI). Modified T7 DNA polymerase was purchased from United States Biochemicals Corp. (Cleveland, OH), isopropyl ß-D-thiogalactopyranoside (IPTG), *N*α-*p*-tosyl-L-arginine methyl ester (TAME), 4-methylumbelliferyl-*p*-guanidinobenzoate, and 4-methylumbelliferyl-*p*-(N,N,N-trimethylammonium) cinnamate chloride were purchased from Sigma Chemical Company (St. Louis, MO). 7-amino-4-methyl-coumarin (AMC) was purchased from Enzyme Systems Products (Livermore, CA). Arginine-ß-naphthylamide, *N*α-benzyloxycarbonyl-Gly-Pro-Arg-7-amido-4-methyl coumarin, tosyl-Gly-Pro-Arg-7-amido-4-methyl coumarin (Tos-GPR-AMC), tosyl-Gly-Pro-Lys-7-amido-4-methyl-coumarin (Tos-GPK-AMC), *t*-butyloxycarbonyl-Leu-Gly-Arg-*p*-nitroanilide, Z-Arg-Phe, Arg-Arg-ß-naphthylamide, and Pro-Arg-ß-naphthylamide were purchased from Bachem Bioscience Inc. (Philadelphia, PA). Purified *Streptomyces griseus* trypsin was a generous gift of Dr. M. N. G. James (University of Alberta).

Strains and Plasmids

Escherichia coli X90 (F' lac I^q , lacZY, pro AB/ Δ(lac-pro), ara, nalA, argE(am), thi, rif^r) (Amann, Brosius et al. 1983) and *E. coli* SY903 (F' lac I^q, lacZ::Tn5, proAB/ recA1, srl::Tn10, Δlac-pro, rif^r, nalA) (Sauer, Smith et al. 1988) were gifts from Dr. A. Vershon (University of California- San Francisco). *E. coli* CJ236 (F'

CJ105/dut1, ung1, relA1) was a gift of BioRad Laboratories Inc. (Richmond, CA). *E. coli* LE112 (F' lac I^q, lacZ::Tn5, proAB/ dut1, ung1, relA1) contained the episome from strain SY903 and the chromosome from strain CJ236 (Chapter Three).

All mutants were created and expressed in pLIB which is a derivative of pT3. In this vector, trypsin expression is controlled by the *tac* promoter and the *trpA* terminator and the secretion of the protein is directed by the hisJ signal peptide (Chapter Three).

Selection (see Chapter Three)

Selection for trypsin activity was optimal on MOPS minimal media (Neidhart, Bloch et al. 1974) using 0.4% glucose as a carbon source, 0.01% vitamin B1, 1 mM IPTG, 300 µM Arg-ß-naphthylamide at 32°C. Addition of ampicillin at 50 mg/L was not required for selection but was included to prevent the growth of untransformed ArgE revertants. Optimal discrimination between background colonies expressing inactive trypsin and survivors expressing active trypsin was achieved in the absence of free Arg supplementation although the time required to achieve the discrimination was reduced from roughly 72 to 48 hr by addition of 1 µM Arg to the selective media.

Screening

Trypsin was harvested from 1.5 ml of an *E. coli* X90 culture grown to saturation in LB, 50 mg/L ampicillin, 1mM IPTG (Evnin and Craik 1988). For the microplate assay, 20 µl of the 40µL of extract were added to 180 µl of 100mM Tris-HCl pH 8.0, 0.4 mM *t*-

BOC-Leu-Gly-Arg-*p*-nitroanilide in one well of a 96-well microplate (see Chapter Two). Trypsin activity was measured in duplicate by following the release of *p*-nitroaniline at 405 nm. For the single substrate gel overlay screen, the activity was assayed as described (Chapter One) (Evnin and Craik 1988).

Enzyme Purification

E. coli X90 cells containing the desired mutant plasmid were inoculated into 500 ml of LB, 50 mg/L ampicillin, 1 mM IPTG and cultured for 12-16 hr. The periplasmic proteins of these cells were harvested in 25% sucrose, 10mM Tris-HCl pH 8.0, 5 mM EDTA as described (Vásquez, Evnin et al. 1989), and was dialyzed exhaustively at 4°C in 10 mM sodium citrate, pH 2.8. The acid-precipitate was separated by centrifugation for 20 min at 12,000 x g, and the supernatant containing trypsin was concentrated by ultrafiltration in an Amicon (Danvers, MA) Centriprep 10 (10,000 MW cutoff). From approximately 20 ml of dialysate, 0.5 ml of concentrate was obtained. This trypsin solution was spun at 14,000 x g for 5 min and the precipitate discarded.

The concentration of active sites was determined by following the release of 4-methylumbelliferone from 4-methylumbelliferyl-*p*-guanidinobenzoate (MUGB) (Jameson, Adams et al. 1973). No burst was observed with an extract from cells carrying pLIB (Figure 2-1). Active trypsin constituted approximately 0.5% of the total protein present in the purified extract as determined by measurement of the concentration of total protein (Smith, Krohn et al. 1985) and the concentration of trypsin active sites.

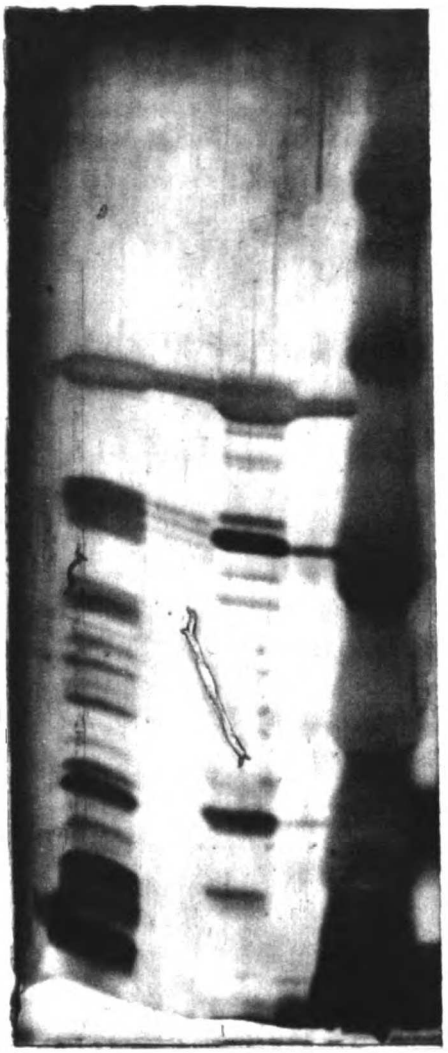
Michaelis-Menten parameters were obtained from a kinetic analysis of acid-precipitated extract and of affinity purified trypsin and trypsin S190G. As explained in more detail in Chapter two, Materials and Methods, the kcat for the crude extract was higher than that observed for the affinity purified by roughly two fold. The Arg/Lys ratio of the two forms were indistinguishable. The hydrolysis rate of Z-Gly-Pro-Arg-AMC by extracts from cells containing pLIB was negligible.

Streptomyces griseus Trypsin

Lyophilized *Streptomyces griseus* trypsin (SGT) was dissolved in 1mM HCl to a concentration of 1 mg/mL for a calculated concentration of 4.35×10^{-5} M. The A_{280} of the solution was measured as 1.53×10^{-3} indicating a molar concentration of 4.12×10^{-5} assuming $\Sigma = 3.71 \times 10^{-5}$ M (Hatanaka, Tsunematsu et al. 1985). The integrity of the enzyme was examined by silver staining a 15% polyacrylamide/0.8% bis gel following SDS-PAGE for three hours at 250 volts (Figure 4-1). The SGT was judged to be 90% intact. The purity was sufficient for crystallization (Read and James 1988) suggesting that the additional lower molecular weight bands observed by silver staining were probably products of autolysis that has occured during storage of the protein. The active site concentration of the SGT was 38.4 μM as determined using 4-methylumbelliferyl-*p*-guanidinobenzoate (MUGB) or 0.88 mg/ml.

Figure 4-1. *Streptomyces griseus* Trypsin

The Bio-Rad silver stain kit was used according to the supplier's protocol. Markers are 43, 25.5, 17.9, 14.5, 8, and 3 kD. The four right-most lanes are relevant. From right to left, these lanes are affinity purified rat trypsin 1 and 10 µl of a 0.4 mg/mL solution (15.7 µM by active site titration) followed by 1 and 10 uls of the SGT solution (1 mg/mL). The gel was mounted between sheets of dialysis membrane after soaking in 10% glycerol and was dried. The reproduction is achieved by photocopying the dried gel.


SGTJTn

The estimate from silver staining was that 0.9 mg/mL of intact protein were present.

RESULTS

Selection of Active Trypsins

E. coli X90 cells are auxotrophic for Arg since they lack one of the enzymes required for Arg biosynthesis, and, as with most *E. coli* K12 strains, they lack proteases that can release free Arg from Arg-ß-naphthylamide (Miller, Green et al. 1982). Trypsin readily cleaves the arginyl naphthylamide bond. Consequently, growth of *E. coli* X90 pT3 cells on a medium containing Arg-ß-naphthylamide as the only source of Arg depends on the secretion of active, Arg-specific trypsin (see Chapter Three).

Kinetic Analysis of Trypsins

The trypsin mutants were characterized by analysis of the kinetic parameters kcat, Km and kcat/Km for comparable tripeptide Arg and Lys substrates (Figure 4-2). Relative to trypsin, all mutants except trypsin S190G have smaller values for kcat than trypsin (Figure 4-2a). Except for trypsin S190T, all mutants exhibit elevated Km values (Figure 4-2b). All mutants exhibit a reduction in kcat/Km for both Arg and Lys substrates (Figure 4-2c). Only the second order rate constant kcat/Km could be obtained for the weakly active trypsins D189N, D189N/S190T, and D189A. These three mutants were the only ones isolated that lacked a negative charge at both position 189 and 190.

Figure 4-2. Kinetic Parameters of Trypsins Mutated at Position 189, 190

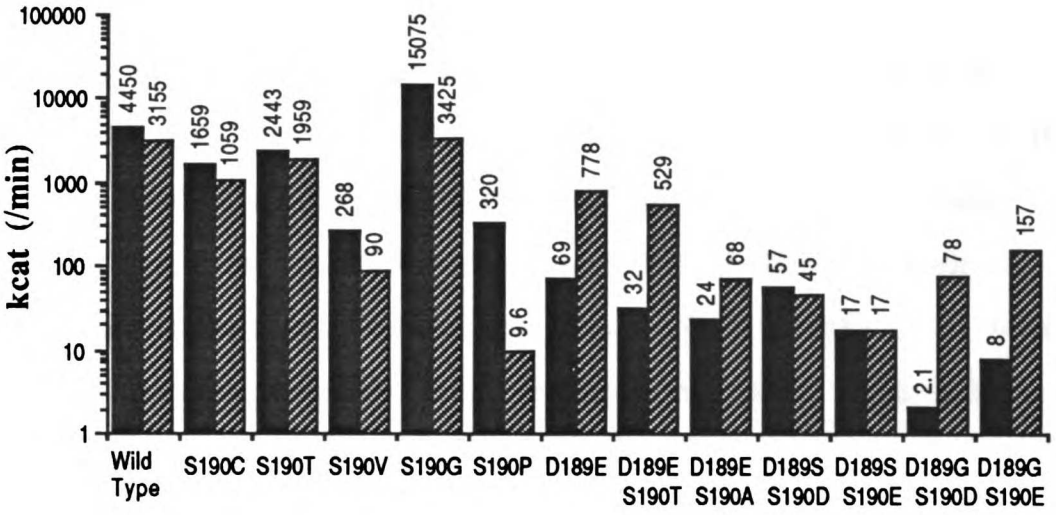
Semi-log plots of the data obtained from kinetic analysis of the purified mutants and trypsin. Each bar represents the average of two determinations on the substrates Tos-GPR-AMC (solid bars) or Tos-GPK-AMC (hatched bars). Numerical values are noted above each bar. Five substrate concentrations were employed per determination: 200, 100, 50, 25, and either 12.5 μM or 500 μM. Substrates were dissolved in DMF, and the final concentration of DMF in the reaction was 1%. AMC release was followed in a Perkin-Elmer LS-5B Luminesence Spectrometer, which was calibrated with a standard solution of AMC, at 460 nm with excitation at 380 nm. Enzyme was added to a final concentration of 2.5-150 nM (depending on the activity of the enzyme). The buffer, 50 mM Tris-HCl pH 8.0, 100 mM NaCl was maintained at 26°C. Velocities were obtained from the linear portion of the free AMC vs. time curve. The kinetic parameters Km and kcat were obtained using a linear regression analysis of the Eadie Hofstee plot V vs. V/{S}. The error in these determinations is less than 20% and in most cases less than 10%. Only the kcat/Km values could be obtained for trypsin D189N, D189A, and D189N/S190T because of their reduced activity. Furthermore, determination of the concentration of active sites for these enzymes could not be obtained with either 4-methylumbelliferyl-*p*-guanidinobenzoate or the chymotrypsin titrant 4-methylumbelliferyl-*p*-N,N,N-trimethylammonium) cinnamate chloride (Jameson, Adams et al. 1973). For the purposes

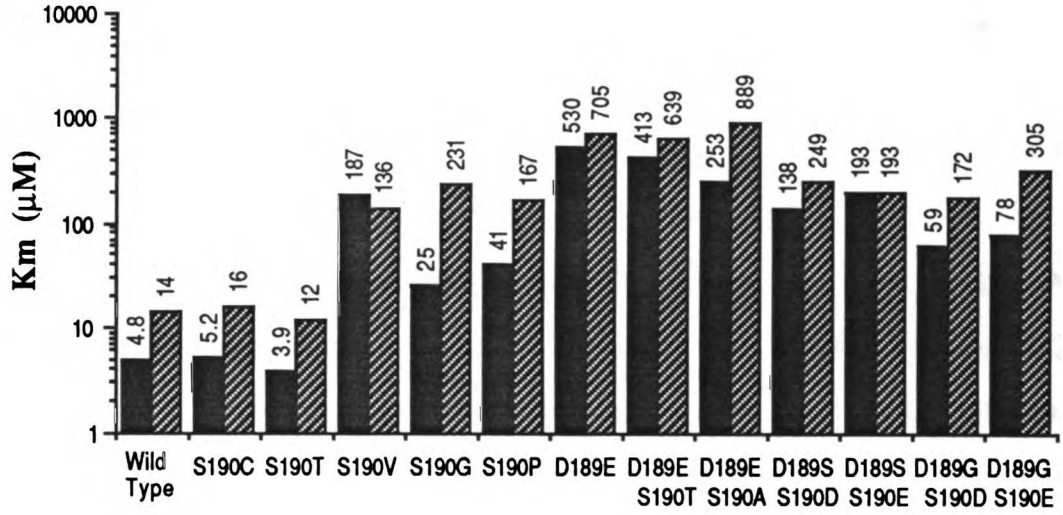
of comparison, immunoblot analysis was used to determine the concentration of trypsin in these samples.

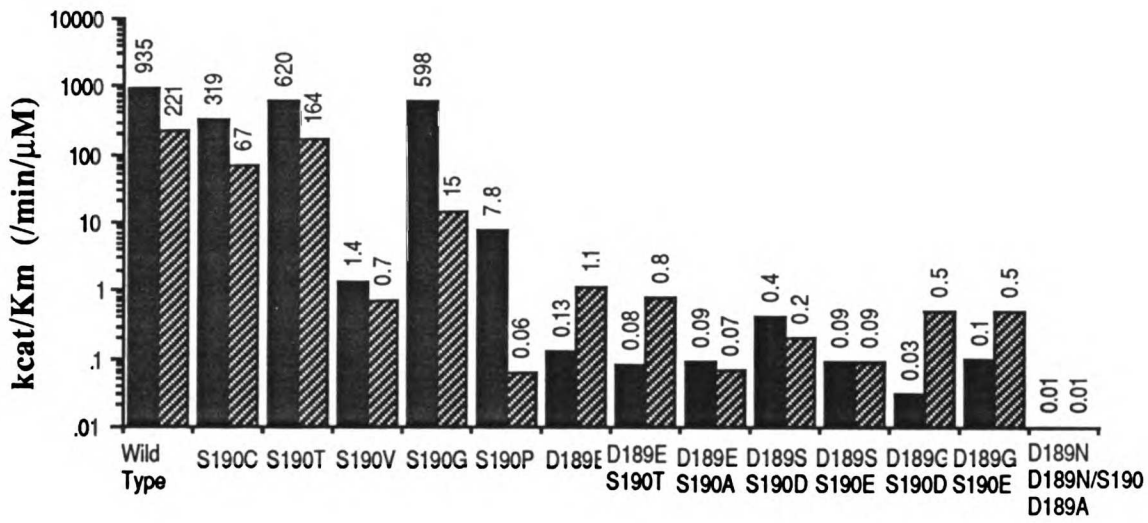
A.) Observed kcat values. Units are min^{-1}.

B.) Observed Km values. Units are μM.

C.) Observed kcat/Km values. Units are $min^{-1}\mu M^{-1}$.


kcat (/min)
100000
10000
1000
100
10
1
4450
3155
1659
1059
2443
1959
268
90
15075
3425
320
9.6
69
778
32
529
24
68
57
45
17
17
2.1
78
8
157
Wild Type
S190C
S190T
S190V
S190G
S190P
D189E
D189E S190T
D189E S190A
D189S S190D
D189S S190E
D189G S190D
D189G S190E


Km (μM)
10000
1000
100
10
1
4.8
14
5.2
16
3.9
12
187
136
25
231
41
167
530
705
413
639
253
889
138
249
193
193
59
172
78
305
Wild Type
S190C
S190T
S190V
S190G
S190P
D189E
D189E S190T
D189E S190A
D189S S190D
D189S S190E
D189G S190D
D189G S190E


kcat/Km (/min/μM)
10000
1000
100
10
1
.1
.01
935
221
319
67
620
164
1.4
0.7
598
15
7.8
0.06
0.13
1.1
0.08
0.8
0.09
0.07
0.4
0.2
0.09
0.09
0.03
0.5
0.1
0.5
0.01
0.01
Wild Type
S190C
S190T
S190V
S190G
S190P
D189E
D189E S190T
D189E S190A
D189S S190D
D189S S190E
D189G S190D
D189G S190E
D189N
D189N/S190 D189A

The very high kcat observed for trypsin S190G on Tos-GPR-AMC is probably a unique property of the interaction of this tripeptide with rat anionic trypsin[36]. Kinetic measurement of the purified extract with Z-GPR-AMC and of the affinity-purified preparations with Tos-GPR-AMC and Z-GPR-AMC with both tripeptide substrates demonstrated that trypsin S190G is a faster enzyme than trypsin with this peptide (Table 4-1). For trypsin S190G, the Z blocked substrate was a slightly worse substrate. The single residue substrate shows trypsin S190G as a slower enzyme than wild type. However, the kcat/Km values differ by only two fold for Z-R-AMC and Z-GPR-AMC; trypsin S190G retains 11% instead of 24% the activity of trypsin.

Substitution of Glutamate for Aspartate

Trypsins D189E and D189E/S190T preferred the Lys to the Arg substrate by approximately 40 fold relative to trypsin (Table 4-2; Figure 4-3). Apparently, the additional methylene group separating the D189E carboxylate from the protein backbone was a handicap in the catalysis of an Arg substrate, as compared with the Lys side chain, leading to the reversal of the relative preference for the Arg substrate. Increases in relative Lys specificity for the D189E and D189E/S190T mutants are largely a result of the differential effects on kcat, and presumably reflect selective stabilization of the transition state

---

[36] A detailed characterization of the Tosyl and Z blocked derivatives of GPR-AMC substrates indicated that the rate determining step in catalysis for the wild type enzyme might be product release instead of formation of the acyl enzyme (J. Vásquez, personal communication).

Table 4-1. Kinetic Analysis of Trypsin S190G

This tabulation represents the data from several different experiments comparing trypsin and trypsin S190G. Pure enzyme refers to affinity purified trypsin while extract precipitated periplasmic extract that has been purified by acid precipitation (see Materials and Methods). The observed values for kcat, Km and kcat/Km were obtained by a linear regression of V against V/S according to Eadie-Hofstee. Determinations were in 50 mM Tris pH 8.0, 100 mM NaCl at 26°C (with 20 mM $CaCl_2$ added in the determination of the single residue data). In all cases, the values reported are the average of at least two determinations. Multiple determinations differed from each other by less than 20%. For the tripeptide substrates, the five substrate concentrations employed were 12.5, 25, 50, 100, and 200 μM. For the single residue with the extracts, the substrate concentration were 62.5, 125, 250, 500 and 1000 μM.

enzyme	purity	substrate	kcat (/min)	Km (μM)	kcat/Km (/min/μM)	kcat/Km %
Wild Type	extract	Tos-GPR-AMC	4450	4.8	927	100
S190G	extract	Tos-GPR-AMC	15075	25	603	65
Wild Type	extract	Tos-GPK-AMC	3155	14	225	100
S190G	extract	Tos-GPK-AMC	3425	231	15	7
Wild Type	extract	Z-R-AMC	27	27.5	1.0	100
S190G	extract	Z-R-AMC	13	63	0.2	21
Wild Type	pure	Z-GPR-AMC	5300	20	265	100
S190G	pure	Z-GPR-AMC	6740	106	64	24
Wild Type	pure	Z-R-AMC	89	57	1.6	100
S190G	pure	Z-R-AMC	57.5	344	0.2	11
Wild Type	pure	Tos-GPR-AMC	2389	4.8	502	100
S190G	pure	Tos-GPR-AMC	4785	22.7	211	42
Wild Type	pure	Tos-GPK-AMC	2045	15.6	131	100
S190G	pure	Tos-GPK-AMC	1662	368	5	3

Figure 4-3. Relative Arg/Lys Substrate Preferences of Trypsins

The value shown is the log of the ratio of the kcat/Km measured for the Tos-GPR-AMC substrate and the kcat/Km measured for the Tos-GPK-AMC substrate. Therefore, negative numbers indicate a preference for Lys, 0 indicates no preference, and positive numbers indicate a preference for the Arg substrate.

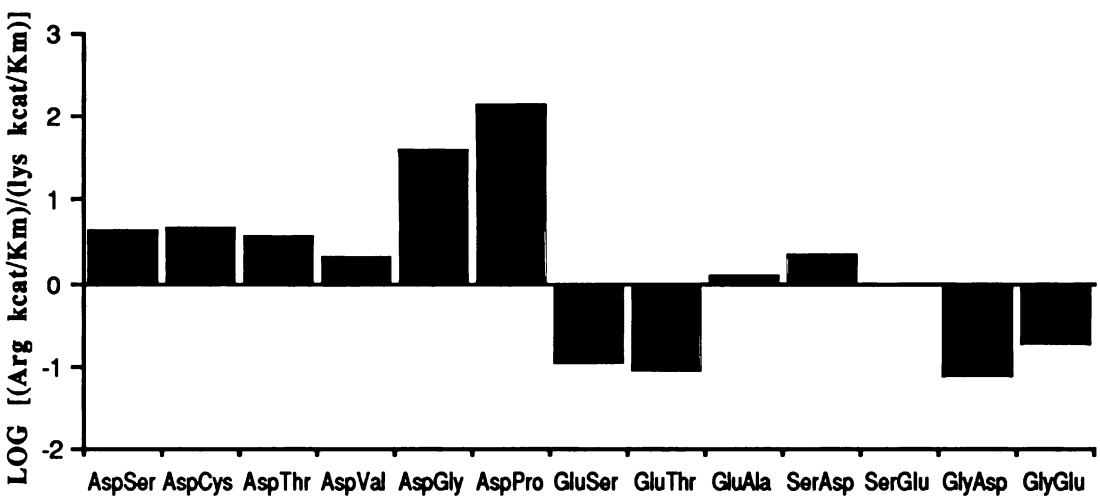

LOG [(Arg kcat/Km)/(lys kcat/Km)]
3
2
1
0
-1
-2
AspSer AspCys AspThr AspVal AspGly AspPro GluSer GluThr GluAla SerAsp SerGlu GlyAsp GlyGlu

Table 4-2. Differential Arg/Lys Specificity of Trypsins Mutated at Positions 189 and190

A kinetic characterization of the 15 mutants and trypsin was completed on Tos-GPR-AMC and Tos-GPK-AMC. The Arg/Lys preference represents the value obtained when the kcat/Km measured for hydrolysis of the Tos-GPR-AMC substrate is divided by the kcat/Km measured for the Tos-GPK-AMC substrate.

Amino Acid Positions 189 190	Arg/Lys Preference
Asp Ser	4.2
Asp Cys	4.7
Asp Thr	3.8
Asp Val	2.2
Asp Gly	40.4
Asp Pro	135.5
Glu Ser	0.12
Glu Thr	0.10
Glu Ala	1.3
Ser Asp	2.28
Ser Glu	1.00
Gly Asp	0.08
Gly Glu	0.20
Asn Ser	-
Asn Thr	-
Ala Ser	-

involving the Lys substrate. Indeed, the kcat values of both enzymes are an order of magnitude greater for the Lys substrate. Our results are consistent with the investigation of trypsin's kinetic parameters using a homoarginine ester substrate (Baines, Baird et al. 1964). The kcat/Km for Nα–benzoyl-homoarginine methyl ester is 0.16% that for cognate arginine substrate at pH 8.4 in 100 mM NaCl, 2 mM $CaCl_2$. The kcat value for homoarginine is 5% the kcat for arginine with a Km that is 51 fold greater than the one for the arginine substrate (Table i-1). Our results differ from those previously obtained with trypsin D189E in which the observed kcat value was 7.7×10^{-5} that of trypsin and the observed Km was lower by 0.6 fold as compared to trypsin (Gráf, Hegyi et al. 1988). However, this discrepancy may be a reflection of the different substrates and conditions utilized.

In contrast to the results at position 189, D and E amino acids functioned similarly at position 190. The Km, kcat, and kcat/Km of the D189S mutants were very similar with D or E at position 190. The D189G/S190D and D189G/S190E mutants also exhibited very similar kinetic parameters (Figure 4-2). Presumably, the hydrogen bonding arrangement between enzyme and substrate has been modified for all four enzymes although the anionic character of the base of the pocket has been preserved.

The substrate binding pocket can operate at a minimal efficiency on positively charged substrates without any negative charge at its base. Trypsins D189N, D189N/S190T and D189A retain sufficient activity to be identified in searches through the library. However, kinetic analysis revealed that these mutants are

reduced by approximately five orders of magnitude in their kcat/Km (Table 4-2).

Trypsin S190T and *S. griseus* Trypsin

The kinetic constants on both the Lys and Arg substrate for the S190T and D189E/S190T enzymes are within two-fold of the cognate enzymes containing S190. Computer modelling of the substitution of threonine for serine indicates that the Oγ oxygens of serine and threonine can be superimposed by introducing only one contact less than 3Å between the threonine Cγ methyl group and neighboring atoms in the binding pocket. Rotation of isoleucine 138 which resides in the shell of amino acids that surrounds the amino acids of the S1 pocket could relieve this short contact. *Streptomyces griseus* trypsin (SGT) is a bacterial homolog of trypsin with threonine at position 190. The three dimensional structure of SGT at high resolution (Read and James 1988) supports a direct comparison of the binding pockets of the two enzymes. Kinetic analysis of purified *S. griseus* trypsin at pH 7.0, 26°C in 50 mM Tris-HCl, 100 mM NaCl with 1 nM enzyme showed that SGT is approximately 7 fold more active than rat trypsin (Table 4-3). The Arg/Lys preference, which is a sensitive measure of equivalent function, of both enzymes is four fold. This correspondance supports the evidence derived from model building that the S190T substitution in the P1 binding pocket occurs without functional disruption.

On the Tos-GPR-AMC substrate, SGT is approaching diffusion control. Knowles and colleagues have shown that a kinetic reaction

Table 4-3. Kinetic Analysis of *S. griseus* Trypsin

Kinetic analysis of tryspin and *S. griseus* trypsin on the tripeptide substrates was achieved using 1 nM enzyme in 50 mM Tris-HCl, 100 mM NaCl buffer maintained at 26°C. The values are averages of two determinations of the parameters. The analysis was performed as with the trypsin mutants (Figure 4-2) except that the concentration of the substrates was adjusted; substrate concentrations of 1.05, 3.1, 6.25, 12.5, and 20 μM were employed. The results using the single residue substrates are not mine. The data with S. griseus trypsin was obtained at 25°C (Hatanaka, Tsunematsu et al. 1985) at pH 7.85. For the ester substrate the enzyme concentration was 3.59-3.86 x 10^{-8} M and the substrate concentrations ranged from 15.6 mM to 1.23 M. The buffer was 10 mM Tris-HCl, 10 mM KCl, 5 mM $CaCl_2$. The hydrolysis of the *p*-nitroanilide substrates was measured spectrophotometrically at 410 nM. The enzyme concentration was 63.9-121 nM. The substrate concentration ranges were 12.3 μM to 2.3 mM. The single residue data for trypsin was obtained with the bovine enzyme (Tsunematsu, Nishimura et al. 1985). The ester data was obtained as with *S. griseus* trypsin except that the enzyme concentration was 88.5 nM. The hydrolysis of the *p*-nitroanilide substrate was followed at pH 8.0.

	S. griseus Trypsin			Trypsin		
Substrate	kcat(per s)	Km (μM)	kcat/Km (per μM per s)	kcat(per s)	Km (μM)	kcat/Km (per μM per s)
Tos-GPR-AMC	4500	1.8	2500	1000	2.7	370
Tos-GPK-AMC	3100	4.7	660	110	13	8
Bz-Arg-OEt	61.5	5	12100	31.4	61	512
Bz-Arg-pNA	22.8	101	225	1.67	461	3.61
Z-Arg-pNA	7.44	58	129	3.82	291	13.1
Ac-Arg-pNA	11.9	308	38.5	3.14	510	6.16

that is characterized by a kcat/Km 4 x 10^8 $s^{-1}M^{-1}$ is diffusion controlled (Albery and Knowles 1976; Albery and Knowles 1977; Hermes, Blacklow et al. 1987). The catalysis of N-(methoxy-carbonyl)-tryptophan-*p*-nitrophenyl ester by chymotrypsin with a kcat/Km of 3.5 x 10^7 $M^{-1}s^{-1}$ was calculated to be 40% diffusion controlled (Brouwer and Kirsch 1982). At pH 7.0, the kcat/Km measured for SGT is 4.2 x 10^7 $M^{-1}s^{-1}$. Thus, the reaction observed at pH 7.0 may be partially diffusion limited, and diffusion control would probably be observed at higher pH (at pH 8.0, SGT is more active) and for reactive substrates such as the thiobenzyl ester derivatives (Harper, Cooke et al. 1984).

Other substrates were employed to extend the analysis of SGT. However, both Arg-AMC, Lys-AMC were ineffective substrates presumably because the amino terminus acts as a competitive inhibitor for the side chain. Kinetic analysis of Z-Arg-AMC was also attempted. Curiously, this Z- blocked single amino acid substrate also showed competitive inhibition. Previous characterization of SGT demonstrated that it exhibited a higher kcat and was more active (kcat/Km) and exhibited a lower Km than trypsin on single residue ester and amide Arg substrates (Hatanaka, Tsunematsu et al. 1985; Olafson and Smillie 1975; Tsunematsu, Nishimura et al. 1985) (Table 4-3).

Modulation of Arg/Lys Substrate Specificity

With the tripeptide substrates Tos-GPR-AMC and Tos-GPK-AMC, trypsin maintains a four-fold Arg/Lys preference (Table 4-2; Figure 4-3). For trypsins S190G and S190P, the Arg substrate is preferred

40 and 135 fold over the Lys substrate. For these mutants, 300% and 1% of the kcat on the Arg substrate are retained. The substantial decrease in trypsin S190P activity may be due to steric clashes between the proline ring and the side chains of Y228 and D189. Apparently, deletion of the hydroxyl group from the side chain of position 190 has caused a substantial modulation of the Arg and Lys substrate selectivity. Similarly, trypsin D189E and D189E/S190T strongly prefer the Lys substrate while trypsin D189E/S190A prefers the Arg substrate by 1.3 fold over the Lys substrate.

The Arg preference exhibited by trypsin S190V is weaker than that of trypsin S190P or S190G and also is weaker than that of wild type trypsin. This result appears to contradict the proposed role of the serine hydroxyl as important for Lys but not Arg catalysis. However, it appears that there are two competing tendencies. Assuming the magnitude of kcat reflects stabilization of the transition state, preferential stabilization of the transition state involving the Arg substrate is observed for this mutant. But, the S190V mutant is the only one in which the Km for the Lys substrate is lower than that of the Arg substrate. The smaller and more flexible Lys residue can move to accommodate the introduction of the Cγ methyl group at position 190, and may even retain a stabilizing van der Waal's contact with that group. The larger more rigid Arg residue is unable to efficiently adjust; it must sacrifice some of the strength of its ground state interaction with Asp189 to accomodate the Val Cγ methyl group. When the kcat/Km constants of trypsin S190V catalysis on the Arg and Lys substrates

are compared, the differentially large increase in Km for Arg masks the differential decrease in the kcat for the Lys substrate.

No known serine protease has cysteine at position 190. However, the activity of trypsin S190C is reduced only 2.5 fold relative to trypsin. The Km is essentially unchanged. Apparently, cysteine can substitute for this serine with impunity in spite of the difference in hydrophobicity and the larger atomic radius of the sulfur atom (1.85 Å versus 1.4 Å). Interaction between the position 190 sulfhydryl and the sulfhydryls comprising the adjacent disulfide bond (positions 191 and 220; Figure i-3, i-4) were not detected. The sulfhydryl is believed to be uncharged at pH 8.0. Its proximity to the aspartate would strongly disfavor its deprotonation.

DISCUSSION

Aside from their identity in formal charge, Arg and Lys side chains are structurally disparate. The shape, length, torsional variability, volume, hydrophobicity, polarizability, distribution of the full positive charge, and hydrogen bonding pattern all differ. Thus, an enzyme that has evolved to cleave both Arg and Lys substrates must take advantage of the similarity and yet have additional features that allow it to be specific for each side chain. The similarity lies in the interaction between the cation of the substrate and the anion at the base of the trypsin binding pocket. The importance of this anionic field is underscored by the impotency of the D189N and D189N/S190T trypsins.

However, the lysine amino group can only interact with this anion through a structural water molecule. Nonetheless, the lysine substrate is nearly as a good a substrate as the arginine. I suggest that the anion at the base of the pocket helps to define an anionic field that emanates from the base of the pocket. This propagation of the charge helps to explain how lysine can be an excellent substrate. The character of the trypsin electrostatic field has been calculated using a description of the electrostatic contribution of all the atoms in the protein and water molecules in the adjacent solvent (Soman, Yang et al. 1989).

The results of experiments in which reversal of the polarity of ion pair has been attempted (Gráf, Craik et al. 1987; Cronin and Kirsch 1988) indicate that this functional change requires multiple changes in the protein structure. Single changes, even at the positions that are apparently critical for the charge character of the

pocket are insufficient. Consequently, the organization of the electrostatic field may be dictated by the protein as a whole and not simply a single amino acid in the binding pocket (Hwang and Warshel 1988).

Trypsin is responsible for a rate enhancement for proteolysis of roughly 10^{10} fold over the rate observed in neutral water (Kahne and Still 1988). The amino acids G, A, T, C, V, and P can substitute for S without substantial effect on this rate enhancement. Apparently, the sole constraint on substitution at position 190 is volume, and it must be concluded that the chemical characteristics of the side chain at position 190 are relatively unimportant for the catalysis. However, a higher resolution analysis of the kinetic parameters of the trypsin S190X substitutions reveal that the absence of a substantial prejudice for Arg or Lys is a carefully crafted feature of trypsin.

Presumably, the substantial decreases in Lys-specific activity of mutants S190G and S190P (and of mutant D189E/S190A relative to D189E) result from destruction of the critical hydrogen bond triad among the crystallographic water molecule (W416 in Figure i-4b), S190 Oγ, and substrate Lys Nζ. Arg specificity is unperturbed because the critical cyclic network of hydrogen bonds between the guanidinium group of Arg and the carboxylate of the D189 is maintained and is the dominant feature of Arg substrate specificity. The differential effects on Lys and Arg specificity can be observed in both Km and kcat. However, unappreciated conformational changes elsewhere in the molecule may play a role in these effects. Clearly, additional structural and functional characterization of

these enzymes is required for a more complete understanding of the functional consequences of these single and double mutations and such analysis is currently underway.

Our conclusions are supported by an analysis of the preference for Arg or Lys substrates exhibited by five trypsin homologs (Figure 4-4). *S. griseus* trypsin exhibits a 3.8 fold preference for Tos-GPR-AMC relative to Tos-GPK-AMC which is nearly identical to that exhibited by rat trypsin. Thrombin and plasma kallikrein exhibit a substantial Arg to Lys substrate preference relative to trypsin (Ascenzi, Menegatti et al. 1982). On a set of single residue ester and amide substrates, thrombin preferred Arg substrates 7-10 fold; kallikrein preferred Arg substrates 15-16 fold. As with the S190P, S190G and D189E/S190A, both thrombin and kallikrein do not have an γ position hydroxyl group at position 190; both have Ala at position 190. The X-ray crystal structure solutions for trypsin (Kreiger, Kay et al. 1974), thrombin (Bode, Mayr et al. 1989), and a kallikrein (Chen and Bode 1983) show that the main chain atoms of the substrate binding pocket are virtually superimposable. Furthermore, except for the change noted, all the amino acids in the first shell of the binding pocket are identical. The known Lys specific trypsin homologs urokinase (Ascenzi, Menegatti et al. 1982) and plasmin (Weinstein and Doolittle 1972) have retained the Asp at position 189 and the Ser at position 190. Urokinase prefers Lys substrates appoximately 100 fold although trypsin has 1000 fold the activity of this homolog on Lys substrates (Ascenzi, Menegatti et al. 1982). The kinetic parameters for urokinase more closely match those of trypsin D189E than trypsin. The Michaelis-Menten

Figure 4-4. Alignment of Trypsin with Natural Homologs
Amino acid alignment of trypsin and enzymes that have a similar specificity and seem to have the structural motifs that indicate that the enzyme is trypsin-like in three dimensions. The amino acids are aligned using the three dimensional structure to help define the comparable sequences. The amino acids shown are the only residues at the bottom of the pocket to have direct contacts with the substrate. As would be expected, the amino acids are highly related. The differences are likely to be of functional importance.

	p189		p191	p214		p216	p219		p226		p228	Preference for Arg(R) / Lys (K)
bov Trypsin	D	S	C	S	W	G	G	C	G	V	Y	R > K
bov Factor XI	D	A	C	S	W	G	G	C	G	V	Y	
rat Kallikrein	D	A	C	S	W	G	G	C	G	V	Y	R > K
hu Plasmin	D	S	C	S	W	G	G	C	G	V	Y	K >> R
hu Plasminogen actv	D	A	C	S	W	G	G	C	G	V	Y	
hu Factor XII	D	A	C	S	W	G	G	C	G	V	Y	
hu Urokinase	D	S	C	S	W	G	G	C	G	V	Y	K > R
hu Factor VII	D	S	C	S	W	G	G	C	G	V	Y	
hu Factor IX	D	S	C	S	W	G	E	C	G	I	Y	
hu Factor X	D	A	C	S	W	G	G	C	G	V	Y	
hu protein C	D	A	C	S	W	G	G	C	G	V	Y	
bov Thrombin	D	A	C	S	W	G	G	C	G	F	Y	R > K
bov Tryptase	D	S	C	S	W	G	G	C	G	I	Y	
S. griseus Trypsin	D	T	C	S	W	G	G	C	G	V	Y	R > K

constants for plasmin have yet to be described. Currently, no biophysical characterization of the three-dimensional structure is available for either Lys-specific homolog.

Based on this comparison, it seems likely that trypsin S190A is an active enzyme. Further work with the library, which has yet to be exhaustively searched, may allow isolation of this mutant. Also, it may be possible to obtain an enzyme in which the aspartic acid is replaced by cysteine at position 189.

This selection provides a novel tool for dissecting the elements of substrate specificity, and, also, an avenue for the isolation of enzymes with a prescribed function without prior description of the critical structural elements. Ultimately, enzymes of a desired specificity differing from any known natural specificity could be designed using existing enzyme scaffolds if a complete understanding of the structural determinants of the forces involved was extant.

Conclusions

Summary and Future Directions

The work in this dissertation was directed toward understanding the structural basis for the exclusive substrate specificity that trypsin exhibits for lysine and arginine amino acids. The tools that will lead to a fuller understanding and more complete answers of the questions have been created, and partial answers to these questions have been identified.

Several technical improvements were essential for the success of this investigation including a bacterial expression system that produces mature trypsin directly (Chapter One and Chapter Three). The expression levels are at least 10 fold greater than those previously achieved in bacteria with pTRAP (Gráf, Jancsó et al. 1988). The yield of purified trypsin is 0.1-0.5 mg/L of culture[37] from a total yield of approximately 1 mg/L of culture. While this yield is low relative to heterologous bacterial expression of bacterial proteins such as ecotin or hisJ, the trypsin product is probably lethal at some level and the $p^{m}T3$ expression system may be close to the maximal level of active trypsin production that can be achieved. Nonetheless, the search for higher levels of expression is ongoing; more potent vectors are being developed for trypsin expression.

Also, the cloning of ecotin is a promising development (M. Mcgrath, personal communication, 1990). With the clone, it should be possible to generate a bacterial strain that can no longer produce the endogenous inhibitor, or to generate one that over-expresses

---

[37]The yield from culture is remarkably consistent on a per liter basis. Whether the cells are grown in an air shaker to roughly 3.0 OD or to 40 OD in a high density fermentor run, the trypsin yield is 1 mg (crude)/L of culture.

the inhibitor. Although expression of trypsin in a strain that produces no ecotin may be lethal, the potential advantages of this strain are important at both the level of the selection of trypsin mutants and for large-scale purification of the enzyme. Alternatively, as trypsin expression may be linked to ecotin expression, over-expression of the inhibitor may lead to increased trypsin expression.

Two different screens were developed that were specific, sensitive, and sufficiently rapid to be useful in screening large numbers of clones; a single substrate gel overlay assay (Chapter One) and a microplate assay (Chapter Two). The single substrate overlay activity gel can detect mutants reduced 10^5 in activity. The number of samples is limited only by the preparation of the periplasm extracts. The microplate assay is less sensitive because it relies on cleavage of an amide bond and because the detection is spectrophotometric in a potentially cloudy solution (due to precipitation of some of the proteins in the crude extract). It is however faster than the activity gel assay and has the added advantage of providing a quantitative expression of the activity for comparison to other extracts. The development of a screen for trypsin within a colony on a plate was not successful despite substantial efforts (Chapter One). This remains a goal of future research.

A rapid way to determine kinetic constants with accuracy was developed (Chapter Two and Chapter Four). Acid-precipitation of the periplasmic extract via dialysis proved to be the only step required to achieve a purity sufficient to determine

fluorometrically the concentration of trypsin active sites in the sample. Also, kinetic analysis could be obtained with these samples. Because the procedure is so rapid, many proteins can be examined in this fashion.

The crowning technical achievement of this work was the development of a genetic selection for trypsin in bacteria (Chapter Three). The selection, at least for an arginine substrate, has a dynamic range of 5 orders of magnitude. Thus, weakly active mutants can be identified. The mutants that were identified in the search of the library of position 189 and 190 mutants spanned the activity range from 60% to 0.002% the activity of trypsin. The gratifying result was that the selection could identify all these mutants. The unsettling result was that the frequency of isolation of a given mutant did not seem to be connected to the activity of the mutant, as might be expected. More extensive use of the selection will undoubtedly aid in understanding its parameters. Also, the selection sheme is reasonably versatile. To search for an enzyme of the desired specificity (i.e. Phe), the strain must be made auxotrophic for that amino acid (i.e. a Phe auxotroph). Considering the ease of P1 transduction, this should not prove to be an excessively high hurdle although identification of the appropriate donor strain may be time consuming. Also, strains other than X90 with the desired markers may express acceptably high levels of trypsin.

There were two levels in this investigation of the specificity of trypsin. In the first, the role of the aspartic acid at position 189 was examined in detail (Chapter One and Chapter Two). In the second,

positions 189 and 190 were examined in detail (Chapter Three and Four). The importance of the aspartic acid at position 189 in defining the character of the binding pocket was affirmed in detail. The best replacements for the aspartic acid are the (presumably[38]) negatively charged amino acids cysteine and glutamic acid which retain 1.4% and 0.01% the activity of trypsin. Trypsin D189N is reduced in activity by 10^5-10^6 fold. In addition, trypsins mutated at position 189 failed to acquire a substantial specificity toward other amino acids. Even neutral substitutions such as leucine or valine did not salvage any activity on a neutral substrate (Leu). The enzyme apparently has evolved to stabilize a negative charge at the base of the pocket. Also, the pocket may be pre-polarized by other elements in the protein to stabilize a specific enzyme-substrate ion pair and to destabilize any other arrangement of functional groups. This ion-pair involves a negative charge in the protein and a positive charge from the substrate.

The study of the pair of positions, 189 and 190, was initiated by searching a library of the possible trypsin mutants at these positions for active mutants. The genetic selection was critical in this step; fifteen mutants were identified in the three searches of the library. Consistent with the dominant role of the negative charge in the function of the pocket, the serine side chain was shown to be completely dispensable for specificity. Comparison of trypsin and trypsin S190G revealed that the serine side chain

---

[38]The pKa of the cysteine sulfur is 8.0. However, neither the ionization state nor the oxidation state of the sulfur were conclusively established.

contributes only 2 fold to arginine specificity and 15 fold to lysine specificity.

However, the serine 190 is the single amino acid that is intimately involved with the preservation of the balance of Arg and Lys specificities in the enzyme. Note that the relatively Arg specific mutants of trypsin have lost the postion 190 hydroxyl group and exhibit compromised Lys specificity. The serine 190 is the only amino acid side chain that tethers the Lys side chain to the protein via a hydrogen bond. The Lys side chain, in contrast to that of Arg, does not interact directly with aspartic acid 189. Natural homologs of trypsin that are Arg specific such as thrombin and urokinase are enzymes that lack the functional group necessary, such as the hydroxyl, to form a hydrogen bond between the side chain of amino acid at position 190 and the enzyme. Both enzymes have alanine at position 190.

One of the intriguing questions that remains is the nature and strength of the electrostatic field and its relationship to specificity. The negative charge of the aspartic acid probably organizes an anionic electrostatic field. The presence of the field helps explain how trypsin (and homologs) can be specific for lysine even though the lysine side chain is not long enough to interact with the aspartic acid carboxylate. Also, the presence of an electrostatic field helps explain how the substitution of glutamic acid for serine at position 190 rescues the activity of an enzyme with serine at postion 189 (trypsin D189S/S190E).

Kinetic analysis of the mutant trypsins revealed changes in both kcat and Km. While it is possible to rationalize changes in the

Km using three dimensional model of trypsin, changes in kcat are more difficult to rationalize. An understanding of how the relationship between the substrate and the S1 binding pocket change as the reaction approaches the transition state is necessary before the changes in the substrate binding pocket can be correlated to changes in kcat. It is important to note that structural analysis of kinetic data has proved elusive in most systems.

Immediate extensions of the work in the S1 pocket include mutation of amino acid positions 189, 190, 216, and 226 simultaneously. All four side chains could be important in modifying the specificity of the pocket. This larger library could be searched for clones that possess arginine-specific activity. By analogy with the results at position 190, an arginine specific enzyme may be created by placing a negative charge at 216 or 226.

An aggressive experimental approach using the selection would be to reverse the polarity of the ion pair. Directed mutagenesis of a trypsin mutant primed for this reversal may be a requisite first step. Alignment of the sequences of trypsin-like viral proteases specific for anionic amino acids identified a set of amino acids that might be critically important in the polarity reversal (F. Bazan, personal communication). Guided by these results, a triple mutant was generated, D189S/V213H/W215A[39]. Random mutagenesis of this template may generate a suitable library for a polarity reversal search.

---

[39]The activity of this mutant was tested against TAME using the activity gel assay (Chapter One) and none was found. This result is not surprising since the D189S mutant is at the limit of detection. No work on the specificity towards anionic substrates was completed.

The underappreciated qualities of cysteine as an amino acid that can replace aspartic acid leave avenues for several important studies. The cysteine sulfur has a larger non-bonded atomic radius than carbon, 1.8 versus 1.35 Å, and it has a pKa of 8.0. Consequently, the cysteine should mimic the aspartic acid above pH 8.0, but not below pH 8.0 (assuming the negative charge is important). Trypsin D189C was made and the protein was characterized (Chapter Two). Trypsin D189C was the most active mutant among the 17 mutants studied at position 189, retaining 3.8% of the kcat and 1.4% of kcat/Km on Tos-GPR-AMC at pH 8.0. While the pH-activity profile failed to identify a pKa near 8.0 that could be associated with the titration of the cysteine sulfur, it seems likely that the sulfur in this mutant was oxidized in the purification and storage of the enzyme. The oxidized forms of the enzyme are negatively charged and may be active, but they will not have the appropriate pKa. Additional work with this mutant is necessary in order to understand these results.

Other critical aspartic acids in trypsin that could be probed by cysteine substitution include those at position 102 (catalytic triad) and 194 (zymogen activation, oxyanion hole). Aspartic acids in other enzymes are also amenable to study with this replacement. Candidate enzymes include the aspartyl proteases which have two aspartic acids that interact for delivery of a hydroxide ion to the carbonyl carbon of the scissile bond. The realization of the isosteric properties of these two amino acids might allow for effective substitution in other systems in which the sulfur is a critical functionality. Such enzymes include the cysteine proteases.

The genetic selection could be useful in a variety of studies of trypsin and, presumably, other proteases. The three amino acids of the catalytic triad could be mutated simultaneously (and amino acids that are in intimate contact with these amino acids such as the serine214) to examine whether any other combination of amino acids could be arranged on the trypsin scaffold. Positive results in this search should expand the data base on why evolution has arrived at the Asp-His-Ser arrangement of amino acids at least two different times (trypsin and subtilisin class). The use of histidine as a base for polarizing either serine and cysteine for nucleophilic attack may also be examined in this experiment.

The application of protein engineering to a system where the three dimensional structure is known and a genetic selection is established has apparently been undervalued in the study of ß-lactamase. The importance of the catalytic serine 70 has been studied in the R-TEM enzyme, the class A enzyme from *E. coli*, by saturation mutagenesis (Schultz and Richards 1986). Several X-ray crystal structures of the ß-lactamases exist including a 2.5 Å resolution structure of the class A *S. aureus* enzyme (Herzberg and Moult 1987), and a 2.0 Å resolution structure of a class C enzyme from *Citrobacter freundii* (Oefner, D'Arcy et al. 1990). Consequently, there is good three dimensional structure information and there is an excellent selection for ß-lactamases using ampicillin. However, no integrated study of this enzyme has been completed. By analogy to the catalytic triad of trypsin, it has been proposed that tyrosine may act as a general base to activate the serine 70 (Herzberg and Moult 1987). Alternatively, several

lysine residues may be catalytically important. A more far-flung experiment is one that addresses the possibility that the lactamases and serine proteases are structurally related. This concept could be explored using mutagenesis and selection for a ß-lactamase that has become a protease (or vice-versa).

It is becoming clear that protein engineering, which was a term that was perhaps coined prematurely, will indeed deserve its title. With the aid of X-ray crystallography, NMR spectroscopy, and molecular biology, the structure::function (structure::activity) relationship in proteins is being dissected. In particular, biochemical and biophysical characterization of proteins carrying single and multiple mutations has proved to be essential in elucidating the function of the amino acid at the mutated position. In addition to understanding the specific interactions mediated by an amino acid at a given position in the protein, the data base generated by this research will allow more sophisticated *ab initio* calculation of the role of amino acids at other positions. Ultimately, our understanding of enzymes will allow alteration of existing enzymes such that an enzyme with a desired function can be generated.

Appendices

Bibliography

Abrahmsen, L., T. Moks, B. Nilsson and M. Uhlen (1986). "Secretion of Heterologous Gene Products to the Culture Medium of *Escherchia coli*." Nuc. Acids Res. **14**: 7487-7500.

Alber, T., S. Dao-ping, K. Wilson, J. Wozniak, S. P. and B. Matthews (1987). "Contribution to Hydrogen Bonds of Thr157." Nature **330**: 41-46.

Albery, W. J. and J. R. Knowles (1976). "Evolution of Enzyme Function and the Development of Catalytic Efficiency." Biochemistry **15**: 5631-5636.

Albery, W. J. and J. R. Knowles (1977). "Efficiency and Evolution of Enzyme Catalysis." Angew. Chem. Int. Ed. Eng. **16**: 285-293.

Amann, E., J. Brosius and M. Ptashne (1983). "Vector Bearing a Hybrid *trp-lac* Promoter Useful for Regulated Expression of Cloned Genes in *Escherichia coli*." Gene **25**: 167-178.

Ames, G. F.-L., K. Nikaido, J. Groarke and J. Petithory (1989). "Reconstitution of Periplasmic Transport in Inside-out Membrane Vesicles." J. Biol. Chem. **264**: 3998-4002.

Ascenzi, P., E. Menegatti, M. Guarneri, F. Bortolotti and E. Antonini (1982). "Catalytic Properties of Serine Proteases. **2**. Comparison between human urinary Kallikrein and human Urokinase, bovine ß-Trypsin, bovine Thrombin and bovine α-Chymotrypsin." Biochemistry **21**: 2483-2490.

Baines, N. J., J. B. Baird and D. T. Elmore (1964). "The Kinetics of Hydrolysis of Derivatives of Arginine, Homoarginine and Ornithine by Trypsin." Biochem. J. **90**: 470-476.

Baird, J. B., E. F. Curragh and D. T. Elmore (1965). "Kinetics and Mechanism of Catalysis by Proteolytic Enzymes: The Kinetics of Hydrolysis of Esters of γ-Guanidino-L-Toluene-p-Sulphonamidobutyric Acid by Bovine Trypsin and Thrombin." Biochem. J. **96**: 733-738.

Bauer, C.-A. (1978). "Active Centers of *Streptomyces griseus* Protease 1, Protease 3, and α-Chymotrypsin: Enzyme-Substrate Interactions." Biochemistry **17**: 375-380.

Bauer, C.-A., G. D. Brayer, A. R. Sielecki and M. N. G. James (1981). "Active Site of α-Lytic Protease." Eur. J. Biochem. **120**: 289-294.

Bauer, C.-A., R. C. Thompson and E. R. Blount (1976a). "The Active Center of *Streptomyces griseus* Protease 3 and α-Chymotrypsin: Enzyme-Substrate Interactions Remote from the Scissile Bond." Biochemistry **15**: 1291-1295.

Bauer, C.-A., R. C. Thompson and E. R. Blount (1976b). "The Active Center of *Streptomyces griseus* Protease 3 and α-Chymotrypsin: Enzyme-Substrate Interactions Close to the Scissile Bond." Biochemistry **15**: 1296-1301.

Birnboim, H. C. and J. Doly (1979). "A Rapid Alkaline Extraction Procedure for Screeing Recombinant Plasmid DNA." Nuc. Acids Res. **7**: 1513-1515.

Bode, W., J. Walter, R. Huber, H. R. Wenzel, and H. Tschesche (1984). "The Refined 2.2Å X-ray Crystal Structure of the Ternary Complex Formed by Bovine Trypsinogen, Val-Val, and the Arginine 15 Analog of Bovine Pancreatic Trypsin Inhibitor." Eur. J. Biochem. **144**: 185-190

Bode, W., I. Mayr, U. Baumann, R. Huber, S. R. Stone and J. Hofsteenge (1989). "The Refined 1.9Å Crystal Structure of Human α-Thrombin: Interaction with D-Phe-Pro-Arg Chloromethylketone and Significance of the Tyr-Pro-Pro-Trp Insertion." EMBO **8**: 3467-3475.

Bone, R., J. L. Silen and D. A. Agard (1989). "Structural Plasticity Broadens the Specificity of an Engineered Protease." Nature **339**: 191-195.

Bowie, J. U., J. F. Reidharr-Olson, W. A. Lim and R. T. Sauer (1990). "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions." Science **247**: 1306-1310.

Brayer, G. D., L. T. J. Delbaere and M. N. G. James (1978). "Molecular Structure of Crystalline *Streptomyces griseus* Protease A at 2.8Å Resolution." J. Mol. Biol. **124**: 261-283.

Brosius, J. and A. Holy (1986). "Regulation of Ribosomal RNA Promoters with a Synthetic *lac* Operator." Proc. Natl. Acad. Sci. **81**: 6929-6933.

Brouwer, A. C. and J. F. Kirsch (1982). "Investigation of Diffusion Limited Rate of Chymotrypsin by Viscosity Variation." Biochemistry **21**: 1302-1307.

Bryan, P., M. Rollence, M. Pantoliano, J. Wood, B. Finzel, G. Gilliland, A. Howard and T. Poulos (1986). "Proteases of Enhanced Stability: Characterization of Enhanced Thermal Stability." Proteins **1**: 326-334.

Burnette, W. (1981). ""Western blotting:" Electrophoretic Transfer of Proteins from Sodium Dodecyl Sulphate-Polyacrylamide Gels to Unmodified Nitrocellulose and Radiographic Detection with Antibody and Radioiodinated Protein A." Anal. Biochem. **112**: 195-203.

Carter, P. and J. A. Wells (1987). "Engineering Enzyme Specificity by 'Substrate-Assisted Catalysis.'" Science **237**: 394-399.

Chen, Z. and W. Bode (1983). "Refined 2.5Å X-ray Crystal Structure of the Complex Formed by Porcine Kallikrein A and the Bovine Pancreatic Trypsin Inhibitor." J. Mol. Biol. **164**: 283-311.

Cheng, Y.-S. and D. Zipser (1979). "Purification and Characterization of Protease III from *Escherchia coli*." J. Biol. Chem. **254**: 4698-4706.

Christie, G., P. Farnham and T. Platt (1981). "Synthetic Sites for Transcription Termination and a Functional Comparison with Tryptophan Operon Termination Sites *In vitro*." Proc. Natl. Acad. Sci. **78**: 4180-4184.

Chung, C. H., H. V. Ives, S. Almeda and A. L. Goldberg (1983). "Purification from *Escherichia coli* of a Periplasmic Protein that is a Potent Inhibitor of Pancreatic Proteases." J. Biol. Chem. **258**: 11032-11038.

Clark, M. R., S. Mel, F. Lupu and D. S. Friend (1987). "Influence of the Membrane Undercoat on Filipin Perturbation of Red Blood Cell Membranes." Expt. Cell Res. **171**: 321-330.

Craik, C. S., C. Largman, T. Fletcher, S. Roczniak, P. J. Barr, R. Fletterick and W. J. Rutter (1985). "Redesigning Trypsin: Alteration of Substrate Specificity." Science **228**: 291-297.

Craik, C. S., Q.-L. Choo, G. Swift, C. Quinto, R. MacDonald and W. Rutter (1984). "Structure of Two Related Rat Pancreatic Trypsin Genes." J. Biol. Chem. **239**: 14255-14264.

Craik, C. S., S. Roczniak, C. Largman and W. J. Rutter (1987). "The Catalytic Role of the Active Site Aspartic Acid in Serine Proteases." Science **237**: 909-913.

Creighton, T. E. (1983). Proteins: Structures and Molecular Properties. New York, W. H. Freeman.

Cronin, C. N. and J. F. Kirsch (1988). "The Role of Arg292 in the Substrate Specificity of Aspartate Aminotransferase as Examined by Site-Directed Mutagenesis." Biochemistry **27**: 4572.

Cronin, C. N., B. A. Malcolm and J. F. Kirsch (1987). "Reversal of Substrate Charge Specificity by Site-Directed Mutagenesis of Aspartate Aminotransferase." J. Am. Chem. Soc. **109**: 2222-2223.

deBoer, H. A., L. J. Comstock and M. Vasser (1983). "The *tac* promoter: A Functional Hybrid Derived From the *trp* and *lac* Promoters." Proc. Natl. Acad. Sci. **80**: 21-25.

Decad, G. M. and H. Nikaido (1976). "Outer Membrane of Gram-Negative Bacteria XII: Molecular Seiving Function of Cell Wall." J. Bacteriol. **128**: 325-336.

Delbaere, L. T. J., W. L. B. Hutchinson, M. N. G. James and W. E. Thiessen (1975). "Tertiary Structural Differences between Microbial Serine Proteases and Pancreatic Serine Enzymes." Nature **257**: 758-762.

Delbruck, M. and J. J. Weigle (1951). "Mutual Exclusion Between an Infecting Phage and a Carried Phage." J. Bacteriol. **62**: 301.

Dente, L., G. Cesareni and R. Cortese (1983). "pEMBL: A New Family of Single-Stranded Plasmids." Nuc. Acids Res. **11**: 1645-1655.

Dohet, C., R. Wagner and M. Radman (1985). "Repair of Defined Base-pair Mismatches in *E. coli*." Proc. Natl. Acad. Sci. **82**: 503-505.

Dotto, G. P. and N. D. Zinder (1981). "Functional Analysis of the Bacteriophage f1 Intergenic Region." Virology **114**: 463-473.

Doolittle, R. F. and D. F. Feng (1987). "Reconstructing the Evolution of Vertebrate Blood Coagulation from a Consideration of the Amino Acid Sequences of Clotting Proteins." Cold Spring Harbor Symp. Quant. Biol. **52**: 869-874.

Dower, W. J., J. F. Miller and C. W. Ragsdale (1988). "High Efficiency Transformation of *E.coli* by High Voltage Electroporation." Nuc. Acids Res. **16**: 6127-6145.

Engler, M., R. Lechner and C. Richardson (1983). "Two Forms of the DNA Polymerase of Bacteriophage T7." J. Biol. Chem. **258**: 11165-11173.

Estell, D. A., T. P. Graycar, J. V. Miller, D. B. Powers, J. P. Burnier, P. N. Ng and J. A. Wells (1986). "Probing Steric and Hydrophobic Effects on Enzyme-Substrate Interactions by Protein Engineering." Science **233**: 659-663.

Evnin, L. B. and C. S. Craik (1988). "Development of an Efficient Method for Generating and Screening Active Trypsin and Trypsin Variants." Ann. N. Y. Acad. Sci. **542**: 61-74.

Evnin, L. B., J. R. Vásquez and C. S. Craik (1990). "Substrate Specificity of Trypsin Investigated by Using a Genetic Selection." Proc. Natl. Acad. Sci. **In press.**

Fersht, A. (1985). "Hydrogen Bonding in Enzymatic Catalysis Analyzed by Protein Engineering." Nature **316**: 656-657.

Fersht, A. R. (1985). Enzyme Structure and Mechanism. San Francisco, W. H. Freeman and Company.

Fersht, A. R., J.-P. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. Waye and G. Winter (1985). "Hydrogen Bonding and Biological Specificity Analyzed by Protein Engineering." Nature **314**: 235238.

Fiedler, F. (1987). "Effects of Secondary Interactions on the Kinetics of Peptide and Peptide Ester Hydrolysis by Tissue Kallikrein and Trypsin." Eur. J. Biochem. **163**: 303-312.

Fletcher, T. S., M. Alhadeff, C. S. Craik and C. Largman (1987). "Isolation and Characterization of a cDNA Encoding Rat Cationic Trypsinogen." Biochemistry **26**:

Fujinaga, M. and M. N. G. James (1987). "Rat Submaxillary Gland Serine Protease, Tonin." J. Mol. Biol. **195**: 373-396.

Gardell, S. J., C. S. Craik, D. Hilvert, M. S. Urdea and W. J. Rutter (1985). "Site-Directed Mutagenesis Shows the Tyrosine 248 of Carboxypeptidase A Does Not Play A Crucial Role in Catalysis." Nature **317**: 551-555.

Geisselsoder, J., F. Witney and P. Yuckenberg (1987). "Efficient Site-Directed *In Vitro* Mutagenesis." BioTechniques. **5** 786-791.

Gráf, L., A. Jancsó, L. Szilágyi, G. Hegyi, K. Pintér, G. Náray-Szabó, J. Hepp, K. Medzihradszky and W. J. Rutter (1988). "Electrostatic Complementarity within the Substrate-Binding Pocket of Trypsin." Proc. Natl. Acad. Sci. **85**: 4961-4965.

Gráf, L., C. S. Craik, A. Patthy, S. Roczniak, R. J. Fletterick and W. J. Rutter (1987). "Selective Alteration of Substrate Specificity by Relacement of Aspartic Acid-189 with Lysine in the Binding Pocket of Trypsin." Biochemistry **26**: 2616-2623.

Gráf, L., G. Hegyi, I. Likó, J. Hepp, K. Medzihradszky, C. S. Craik and W. J. Rutter (1988). "Structural and Functional Integrity of Specificity and Catalytic Sites of Trypsin." Int. J. Peptide Res. **32**: 512-518.

Hanahan, D. (1985). Techniques for Transformation of *E.coli*, in DNA Cloning, a practical approach. D.M. Glover, ed., Washington D.C., IRL Press, pgs. 109-135.

Handbook of Chemistry and Physics (1978). R. Weast, ed., CRC Press Inc., Cleveland, Ohio, pgs D136-139 (Acid Base Indicators).

Harper, J. W., R. R. Cooke, C. J. Roberts, B. J. McLaughlin and J. C. Powers (1984). "Active Site Mapping of the Serine Proteases Human Leukocyte Elastase, Cathepsin G, Porcine Pancreatic Elastase, Rat Mast Cell Protease I and II, Bovine Chymotrypsin A, and Staphylococcus Aureus Protease V8 Using Tripeptide Thiobenzyl Ester Substrates." Biochemistry **23**: 2995-3002.

Hatanaka, Y., H. Tsunematsu, K. Mizusaki and S. Makisumi (1985). "Interactions of Derivatives of Guanidinophenylalanine and Guanidinophenylglycine with *Streptomyces griseus* Trypsin." Biochem. Biophys. Acta **832**: 274-279.

Hartley, B. S., J. R. Brown, D. L. Kauffman and L. B. Smillie (1965). "Evolutionary Similarities Between Pancreatic Proteolytic Enzymes." Nature **207**: 1157-1159.

Helinski, D. R. and C. Yanofsky (1963). "A Genetic and Biochemical Analysis of Second-Site Reversion." J. Biol. Chem. **238**: 1043-1048.

Hermes, J. D. and J. R. Knowles (1987). Design of Enzymes and Enzyme Models. Robert A. Welch Foundation XXXI. 207-211.

Hermes, J. D., S. C. Blacklow and J. R. Knowles (1987). "The Development of Enzyme Catalytic Efficiency: An Experimental Approach." Cold Spring Harbor Symp. Quant. Biol. **52**: 597-602.

Hermes, J. D., S. C. Blacklow and J. R. Knowles (1990). "Searching Sequence Space by Definably Random Mutagenesis: Increasing the Catalytic Potency of an Enzyme." Proc. Natl. Acad. Sci. **87**: 696-700.

Herzberg, O. and J. Moult (1987). "Bacterial Resistance to ß-Lactam Antibiotics: Crystal Structure of ß-Lactamase from *Staphylococcus aureus* PC1 at 2.5Å Resolution." Science **236**: 694-701.

Higaki, J. N., B. Gibson and C. S. Craik (1987). Evolution of Catalysis in the Serine Proteases. Cold Spring Harbor Symp. Quant. Biol. Cold Spring Harbor, Cold Spring Harbor Laboratory, pgs 615-621.

Higaki, J. N., L. B. Evnin and C. S. Craik (1989). "Introduction of a Cysteine Protease Active Site into Trypsin." Biochemistry **28**: 9256-9263.

Hummel, B. C. W. (1959). "A Modified Spectrophotometric Determination of Chymotrypsin, Trypsin, and Thrombin." Canad. J. Biochem. Phys. **12**: 1393-1399.

Hwang, J.-K. and A. Warshel (1988). "Why Ion Pair Reversal by Protein Engineering is Unlikely to Succeed." Nature **334**: 270-272.

Ikemura, H., H. Tagaki and M. Inouye (1987). "Requirement of Pro-sequence for the Production of Active Subtilisin E in *Escherichia coli*." J. Biol. Chem. **262**: 7859-7864.

Inagami, T. and T. Mitsuda (1964). "Mechanism of Trypsin Specificity II: Comparison of Trypsin and a-Chymotrypsin in the Non-Specific Catalysis of Acetylglycine Ethyl Ester." J. Biol. Chem. **239**: 1388-1394.

Inouye, H., W. Barnes and J. Beckwith (1982). "Signal Sequence of Alkaline Phosphatase of *Escherichia coli*." J. Bacteriol. **149**: 434-439.

James, M. N. G., A. R. Sielecki, G. D. Brayer and L. T. J. Delbaere (1980). "Structures of Product and Inhibitor Complexes of *Streptomyces griseus* Protease A at 1.8 Å Resolution." J. Mol. Biol. **144**: 43-88.

Jameson, G. W., D. V. R. R. W. Adams, W. S. A. Kyle and D. T. Elmore (1973). "Determination of the Operational Molarity of Solutions of Bovine α-Chymotrypsin, Trypsin, Thrombin, and Factor Xa by Spectrofluorimetric Titration." Biochem. J. **131**: 107-117.

Kahne, D. and W. C. Still (1988). "Hydrolysis of a Peptide Bond in Neutral Water." J. Am. Chem. Soc. **110**: 7529-7534.

Keil-Dlouhá, V., N. Zylber, J.-M. Imhoff and N.-T. Tong (1971a). "Proteolytic Activity of Pseudotrypsin." FEBS Lett. **16**: 291-294.

Keil-Dlouhá, V., N. Zylber, N.-T. Tong and B. Keil (1971b). "Cleavage of Glucagon by α- and ß-Trypsin." FEBS Lett. **16**: 287-290.

Kelley, R. L. and C. Yanofsky (1985). "Mutational Studies with the *trp* Represor of *Escherichia coli* Support the Helix-Turn-Helix Model of Operator Recognition." Proc. Natl. Acad. Sci. **82**: 483-487.

King, P. V. and R. P. Blakesley (1986). "Optimizing DNA Ligations for Transformation." Focus **8**: 1-3.

Kitagawa, K. and N. Izumiya (1959). "Action of Trypsin and Papain on Derivatives of α-Amino-γ-Guanidinobutyric Acid and Homoarginine." J. Biochem. **46**: 1159-1165.

Knowles, J. R. (1987). "Tinkering with Enzymes: What Are We Learning?" Science **236**: 1252-1258.

Kramer, B., W. Kramer and H.-J. Fritz (1984). "Different Base/Base Mismatches are Corrected with Different Efficiencies by the Methyl Directed DNA Mismatch Repair System of *E. coli*." Cell **38**: 879-887.

Kraut, J. (1977). "Serine Proteases: Structure and Mechanism of Catalysis." Ann. Rev. Biochem. **46**: 331-358.

Kreiger, M., L. M. Kay and R. M. Stroud (1974). "Structure and Specific Binding of Trypsin: Comparison of Inhibited Derivatives and a Model for Substrate Binding." J. Mol. Biol. **83**: 209-230.

Kunkel, T. (1985). "Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection." Proc. Natl. Acad. Sci. **82**: 488-492.

Kunkel, T. A., J. D. Roberts and R. A. Zakour (1987). Rapid and Efficient Site-Specific Mutagenesis without Phenotypic

Selection. Meth. Enzym. San Diego, Academic Press Inc., pgs 367-381.

Kurjan, J. and I. Herskowitz (1982). "Structure of a Yeast Pheromone Gene (MFα): A Putative α-Factor Precursor Contains Four Tandem Copies of Mature α-Factor." Cell **30**: 633-943.

Laemmli, U. (1970). "Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4." Nature **227**: 680-685.

Lazdunski, M., J.-P. Vincent, H. Schweitz, M. Péron-Renner and J. Pudles (1974). "The Mechanism of Association of Trypsin (or Chymotrypsin) with PT Inhibitors (Kunitz and Kazal): Kinetics and Thermodynamics of the Interaction" in Proteinase Inhibitors, eds H. Fritz, H. Tschesche, L.J. Greene, and E. Truscheit. Springer Verlag, Berlin, pgs. 420-431.

Lechner, R., M. Engler and C. Richardson (1983). "Characterization of Strand Displacement Catalyzed by Bacteriophage T7 DNA Polymerase." J. Biol. Chem. **258**: 11174-11184.

Lederberg, E. M. (1951). "Lysogenicity in *E. coli* K-12." Genetics **36**: 560.

Lo, K., S. Jones, N. Hackett and H. Khorana (1984). "Specific Amino Acid Substitutions in Bacterioopsin: Replacement of a Restriction Fragment in the Structural Gene by Synthetic DNA Fragments Containing Altered Codons." Proc. Natl. Acad. Sci. **81**: 2285-2289.

MacDonald, R., S. Stary and G. Swift (1982). "Two Similar but Nonallelic Rat Pancreatic Trypsinogens." J. Biol. Chem. **257**: 9724-9732.

MacGregor, R. D. and C. A. Hunt (1986). A New Functionally Improved Model of the Cell Membrane. Beth Collins, Masters Thesis, University of California, School of Pharmacy.

Mandel, M. and A. Higa (1970). "Calcium Dependent Bacteriophage DNA Infection." J. Mol. Biol. **53**: 154.

Maniatis, T., E. F. Fritsch and J. Sambrool (1982). Molecular Cloning. Cold Spring Harbor, Cold Spring Harbor Press.

Mares-Guia, M. and E. Shaw (1965). "Studies on the Active Center of Trypsin." J. Biol. Chem. **240**: 1579-1585.

Mares-Guia, M. and E. Shaw (1967). "The Specific Inactivation of Trypsin by Ethyl *p*-Guanidinobenzoate." J. Biol. Chem. **242**: 5785-5788.

Mares-Guia, M., E. Shaw and W. Cohen (1967). "Studies on the Active Center of Trypsin II." J. Biol. Chem. **242**: 5777-5781.

Marquart, M., J. Walter, J. Deisenhofer, W. Bode and R. Huber (1983). "The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen, and Its Complexes with Inhibitors." Acta Cryst. **B39**: 480-490.

Matthews, B. W., P. B. Sigler, R. Henderson, and D. M. Blow (1967). "Three-dimensional Structure of Tosyl-a-Chymotrypsin." Nature **214**: 652-656.

McClary, J. A., F. Witney and J. Geiselsoder (1989). "Efficient Site-Directed *In Vitro* Mutagenesis Using Phagemid Vectors." BioTechniques **7**: 282-289.

McGrath, M. E., M. E. Wilke, J. N. Higaki, C. S. Craik and R. J. Fletterick (1989). "Structures of Two Trypsins with a Cysteine Protease Active Site." Biochemistry **28**: 9264-9271.

Messing, J. (1983). Meth. Enzymol. **101**: 20-78.

Miller, C. G. and K. MacKinnon (1974). "Peptidase Mutants of *Salmonella typhimurium*." J. Bacteriol. **120**: 355-363.

Miller, C. G., L. Green and R. Schultz (1982). "*Salmonella typhimurium* Mutations Affecting Utilization of L-Leucine ß-Naphthylamide." Mol. Gen. Genet. **186**: 228-234.

Miller, J. H. (1972). Experiments in Molecular Genetics. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory.

Moks, T., L. Abrahmsen, E. Holmgren, M. Bilich, A. Olsson, M. Uhlen, G. Pohl, C. Sterky, H. Hultberg, S. Josephson, A. Homgren, H. Jornvall and B. Nilsson (1987). "Expresion of Human Insulin-like Growth Factor I in Bacteria: Use of Optimized Gene Fusion Vectors to Facilitate Protein Purification." Biochemistry **26**: 5239-5244.

Nakamura, K., Y. Masumai and M. Inouye (1982). J. Mol. Appl. Gen. **1**: 289-299.

Nakata, A., H. Shinagawa and M. Amemura (1982). "Cloning of the Alkaline Phosphatase Isozyme Gene (IAP) of *E.coli.*" Gene **19**: 313-319.

Neidhart, F. C., P. L. Bloch and D. F. Smith (1974). "Culture Medium for Enterobacteria." J. Bacteriol. **119**: 736-747.

Neurath, H., K. A. Walsh and W. P. Winter (1967). "Evolution of Structure and Function of Proteases." Science **158**: 1638-1644.

Nossal, N. (1974). "DNA Synthesis on a Double-Standed DNA Template by the T4 Bacteriophage DNA Polymerase and T4 gene 32 DNA Unwinding Protein." J. Biol. Chem. **249**: 5668-5676.

Oefner, C., A. D'Arcy, J. J. Daly, K. Gubernator, R. L. Charnas, I. Heizne, C. Hubschwerlen and F. K. Winkler (1990). "Refined Crystal Structure of ß-Lactamase from *Citrobacter freundii* Indicates a Mechanism for ß-Lactam Hydrolysis." Nature **343**: 284-288.

Olafson, R. W. and L. B. Smillie (1975). "Enzymic and Physicochemical Properties of *Streptomyces griseus* Trypsin." Biochemistry **14**: 1161-1167.

Payne, G. and C. Gilvarg (1971). "Peptide Transport." Adv. Enzymol. **35**: 187-244.

Pheiffer, B. and S. Zimmerman (1983). "Polymer-Stimulated Ligation: Enhance Blunt- or Cohesive-End Ligation of DNA or Deoxyribonucleotides by T4 DNA Ligase in Polymer Solutions." Nuc. Acids Res. **11**: 7853-7864.

Pichuantes, S., L. M. Babé, P. J. Barr and C. S. Craik (1989). "Recombinant HIV1 Protease Secreted by *Saccharomyces cerevisiae* Correctly Processes Myristylated *gag* Polyprotein." Proteins **6**: 324-337.

Pugsley, A. P. and M. Schwartz (1985). "Export and Secretion of Proteins by Bacteria." FEMS Microbiol. Res. **32**: 3-38.

Read, R. J. and M. N. G. James (1988). "Refined Crystal Structure of *Streptomyces griseus* Trypsin at 1.8 Å Resolution." J. Mol Biol. **200**: 523-551.

Reidharr-Olson, J. F. and R. T. Sauer (1988). "Combinatorial Cassette Mutagenesis as a Probe for the Informational Content of Protein Sequences." Science **241**: 53-57.

Rothman, J. H. and T. H. Stevens (1986). "Protein Sorting in Yeast: Mutants Defective in Vacuole Biogenesis Mislocalize Vacuolar Proteins into the Late Secretory Pathway." Cell **47**: 1041-1049.

Ruhlmann, A., D. Kukla, P. Schwager, K. Bartels and R. Huber (1973). "Structure of the Complex Formed by Bovine Trypsin and Bovine Pancreatic Trypsin Inhibitor." J. Mol. Biol. **77**: 417-436.

Russell, M., S. Kidd and M. R. Kelley (1986). "An Improved Filamentous Helper Phage for Generating Single-Stranded Plasmid DNA." Gene. **45**: 333-338.

Saenger, W. (1984). Principles of Nucleic Acid Structure. Advanced Texts in Chemistry. New York, Springer Verlag.

Sanborn, B. M. and G. E. Hein (1968). "The Interaction of Trypsin with Neutral Substrates and Modifiers." Biochemistry **7**: 3616-3624.

Sanger, F., F. Nicklen and A. Coulson (1977). "DNA Sequencing with Chain-Terminating Inhibitors." Proc. Natl. Acad. Sci. **74**: 5463-5467.

Sarath, G., R. S. de la. Motte and F. W. Wagner (1989). "Protease Assay Methods" in Proteolytic Enzymes: A Practical Approach. eds. R. J. Beynon and J.S. Bond. New York, IRL Press, pgs 25-42.

Sauer, R. T., D. L. Smith and A. D. Johnson (1988). "Flexibility of the Yeast α2 Repressor Enables It to Occupy the Ends of Its Operator, Leaving the Center Free." Genes and Development **2**: 807-816.

Sawyer, L., D. M. Shotton, J. W. Campbell, P. L. Wendell, H. Muirhead, H. C. Watson, R. Diamond and R. C. Ladner (1978). "The Atomic Structure of Crystalline Porcine Pancreatic Elastase at 2.5Å Resolution: Comparisons with the Structure of α-Chymotrypsin." J. Mol. Biol. **118**: 137-208.

Schechter, I. and A. Berger (1967). "On the Size of the Active Site in Proteases: I. Papain." Biochem. Biophys. Res. Comm. **27**: 157-162.

Schevitz, R. W., Z. Otwinowski, A. Joahimiak, C. L. Lawson and P. B. Sigler (1985). "The Three-Dimensional Structure of *trp* Repressor." Nature **317**: 782-786.

Schultz, C. and J. Richards (1986). "Site-Saturation Studies of ß-Lactamase: Production and Characterization of Mutant ß-Lactamases with All Possible Amino Acids Substitutions at Residue 71." Proc. Natl. Acad. Sci. **83**: 1588-1592.

Scrutton, N. S., A. Berry and R. N. Perham (1990). "Redesign of the Coenzyme Specificity of a Dehydrogenase by Protein Engineering." Nature **343**: 38-43.

Shortle, D. (1983). "A Genetic System for Analysis of Staphylococcal Nuclease." Gene **22**: 181-189.

Shortle, D. and B. Lin (1985). "Genetic Analysis of Staphylococcal Nuclease: Identification of Three Intragenic "Global" Supressors of Nuclease-Minus Mutations." Genetics **110**:

Silen, J. L., D. Frank, A. Fujishige, R. Bone and D. A. Agard (1989). "Analysis of Prepro-α-Lytic Protease Expression in *Escherichia coli* Reveals that the Pro Region Is Required for Activity." J. Bacteriol. **171**: 1320-1324.

Silhavy, T. J., M. L. Berman and L. W. Enquist (1984). Experiments with Gene Fusions. Cold Spring Harbor, Cold Spring Harbor Laboratory.

Singer, S. J. and D. G. Nicolson (1972). "The Fluid Mosiac Model of the Structure of Cell Membranes." Science **175**: 720-731.

Slots, J (1987). "Detection of Colonies of *Bacteriodes gingivalis* by a Rapid Fluorescence Assay for Trypsin-Like Activity." Oral Microbiol. Immunol. **2**: 139-141.

Smith, P. K., R. I. Krohn, G. T. Hermanson, A. F. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson and D. C. Klenk (1985). "Measurement of Protein Using Bicinchoninic Acid." Anal. Biochem. **150**: 76-85.

Soman, K., A.-S. Yang, B. Honig and R. Fletterick (1989). "Electrical Potentials in Trypsin Isozymes." Biochemistry **28**: 9918-9923.

Sprang, S., T. Standing, R. J. Fletterick, R. M. Stroud, J. Finer-Moore, N.-H. Xuong, R. Hamlin, W. J. Rutter and C. S. Craik (1987). "The Three-Dimensional Structure of Asn102 Mutant of Trypsin: Role of Asp102 in Serine Protease Catalysis." Science **237**: 905-909.

Steitz, T. and R. Shulman (1982). "Crystallographic and NMR studies of the Serine Proteases." Ann. Rev. Biophys. Bioeng. **11**: 419-444.

Stroud, R. M. (1974). "A Family of Protein-cutting Proteins." Sci. Am. **231**: 74-88.

Tabor, S. and C. C. Richardson (1987). "DNA Sequence Analysis with a Modified Bacteriophage T7 DNA Polymerase." Proc. Natl. Acad. Sci. **84**: 4767-4771.

Tanaka, T., B. J. McRae, K. Cho, R. Cook, J. Fraki, D. A. Johnson and J. C. Powers (1983). "Mammalian Tissue Trypsin-like Enzymes: Comparative Reactivities of Human Skin Tryptase, Human Lung Tryptase, and Bovine Trypsin with Peptide 4-Nitroanilide and Thioester Substrates." J. Biol. Chem. **258**: 13552-13557.

Thompson, R. C. and E. R. Blount (1970). "Evidence for an Extended Active Center in Elastase." Proc. Natl. Acad. Sci. **67**: 1734-1740.

Titani, K., Sasagawa, R. G. Woodbury, L. H. Ericsson, H. Doersam, M. Kraemer, H. Neurath and R. Zwilling (1983). Biochemistry **22**: 1459.

Torriani, A. and F. Rothman (1961). "Mutants of *Escherichia coli* Constitutive for Alkaline Phosphatase." J. Bacteriol. **81**: 835-836.

Trowbridge, C. G., A. Krehbiel and J. M. Laskowski (1963). "Substrate Activation of Trypsin." Biochemistry **2**: 843-850.

Tsunematsu, H., H. Nishimura, K. Mizusaki and S. Makisumi (1985). "Kinetics of Hydrolysis of Amide and Anilide Substrates of *p*-Guanidino-L-Phenylalanine by Bovine and Porcine Trypsins." J. Biochem. **97**: 617-623.

Vajda, T. and T. Szabó (1976). "Specificity of Trypsin and α-Chymotrypsin towards Neutral Substrates." Acta Biochem. Biophys. Acad. Sci. Hung. **11**: 287-294.

Vásquez, J. R. and C. S. Craik (1990). "The Role of Electrostatics in the Trypsin Catalytic Triad." J. Am. Chem. Soc. **in preparation**:

Vásquez, J. R., L. B. Evnin, J. N. Higaki and C. S. Craik (1989). "An Expression System for Trypsin." J. Cell. Biochem. **39**: 265-276.

Wanner, B. L., M. R. Wilmes and D. C. Young (1988). "Control of Bacterial Alkaline Phosphatase Synthesis and Variation in the *Escherichia coli phoR* Mutant by Adenyl Cyclase, the cyclic AMP Receptor Protein, and the *phoM* operon." J. Bacteriol. **170**: 1092-1102.

Warner, J., B. Duncan, C. Garrett and J. Neuhard (1981). "Synthesis and Metabolism of Uracil Containing Deoxyribonucleic acid in *E. coli*." J. Bacteriol. **145**: 687-695.

Warshel, A., J. Åqvist and S. Creighton (1989). "Enzymes Work by Solvation Substitution Rather Than by Desolvation." Proc. Natl. Acad. Sci. **86**: 5820-5824.

Weinstein, M. J. and R. F. Doolittle (1972). "Differential Specificities of Thrombin, Plasmin, and Trypsin with regard to Synthetic and Natural Substrates and Inhibitors." Biochem. Biophys. Acta **258**: 577-590.

Wells, J. A., B. C. Cunningham, T. P. Graycar and D. A. Estell (1987). "Recruitment of Substrate-Specificity Properties from One Enzyme into a Related One by Protein Engineering." Proc. Natl. Acad. Sci. **84**: 5167.

Wells, J. A., D. B. Powers, R. R. Bott, T. P. Graycar and D. A. Estell (1987). "Designing Substrate Specificity by Protein Engineering of Electrostatic Interactions." Proc. Natl. Acad. Sci. **84**: 1219-1223.

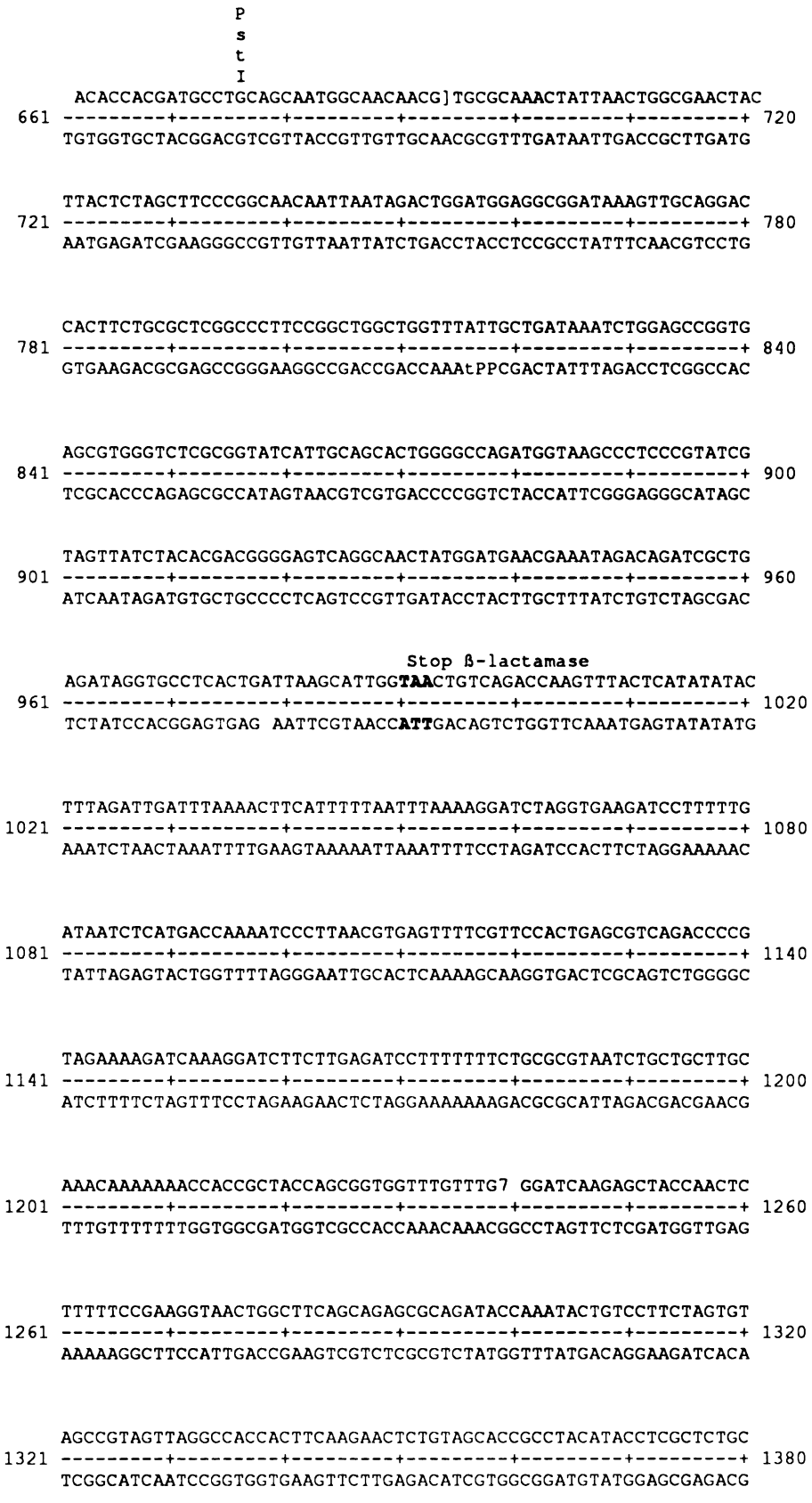
Wilke, M. and R. J. Fletterick (1990). "1.6Å X-ray Crystal Structure of Trypsin Gly226Ala Complexed to Benzamidine." J. Mol. Biol. **in press.**

Wilks, H. M., K. W. Hart, R. Feeney, C. R. Dunn, H. Muirhead, W. N. Chia, D. A. Barstow, T. Atkinson, A. R. Clarke and J. J. Holbrook (1988). "A Specific, Highly Active Malate Dehydrogenase by Redesign of a Lactate Dehydrogenase Framework." Science **242**: 1541-1544.

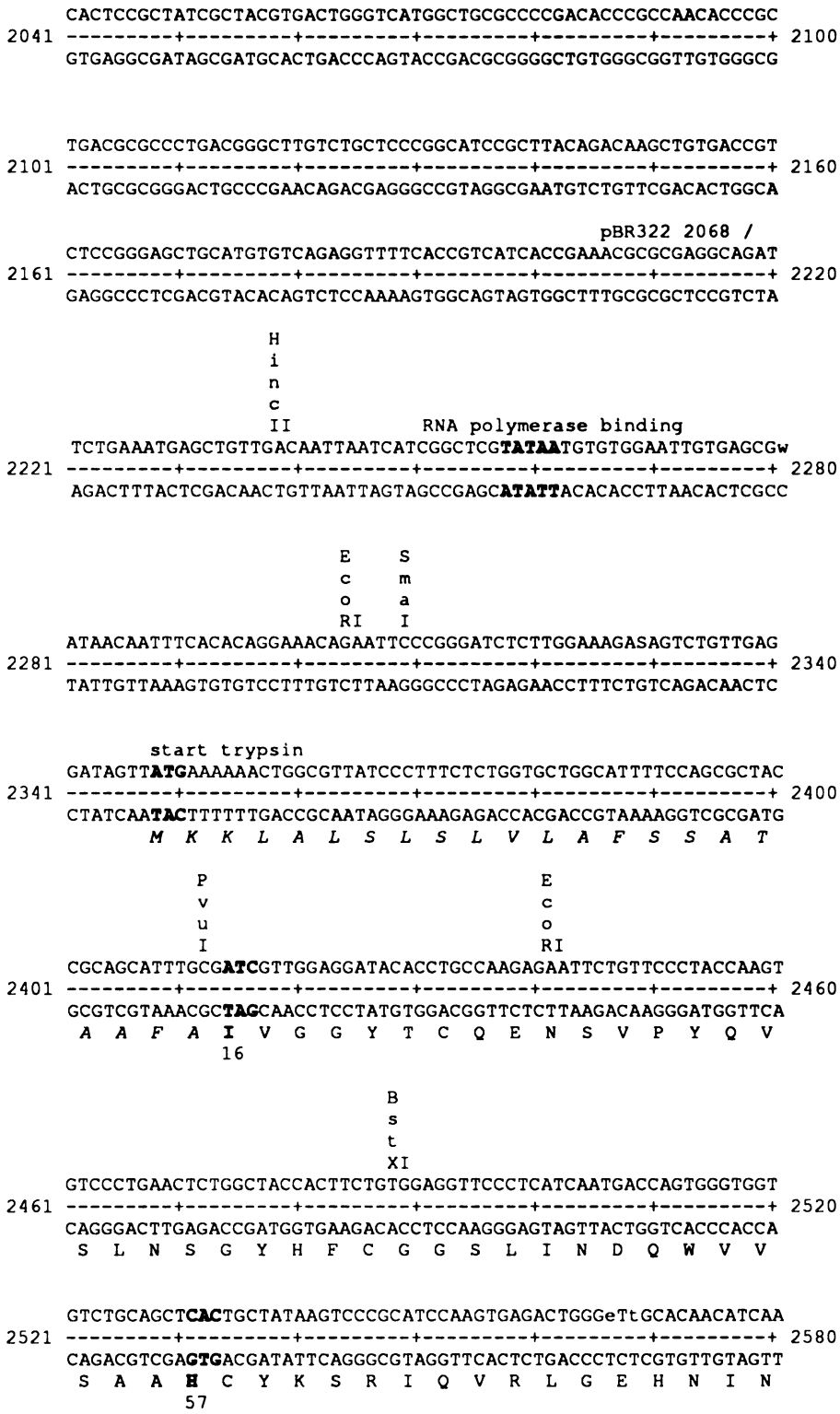
Young, J. D.-E., C.-C. Liu, L. G. Leong, A. Damiano, P. M. Persechini and Z. A. Cohn (1987). "An Automated Colorimetric Microassay System for Rapid and Quantitative Determination of Soluble Enzyme Activities." BioTechniques **5**: 572-576.

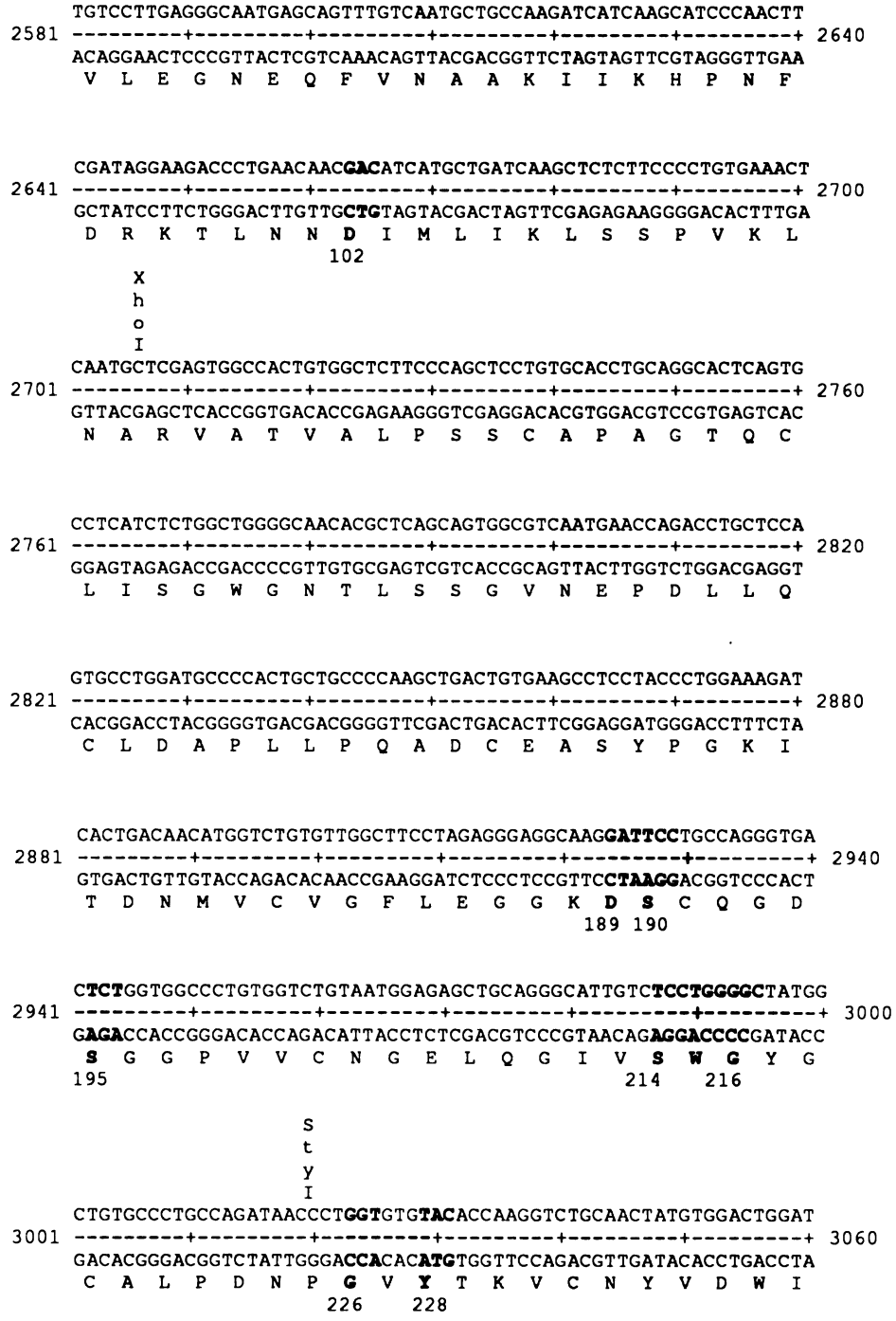
Zhang, R. G., A. Joahimiak, C. L. Lawson, R. W. Schevitz, Z. Otwinowski and P. B. Sigler (1987). "The Crystal Sturcture of *trp* Aporepressor at 1.8Å Shows How Binding Tryptophan Enhances DNA Affinity." Nature **327**: 591-597.

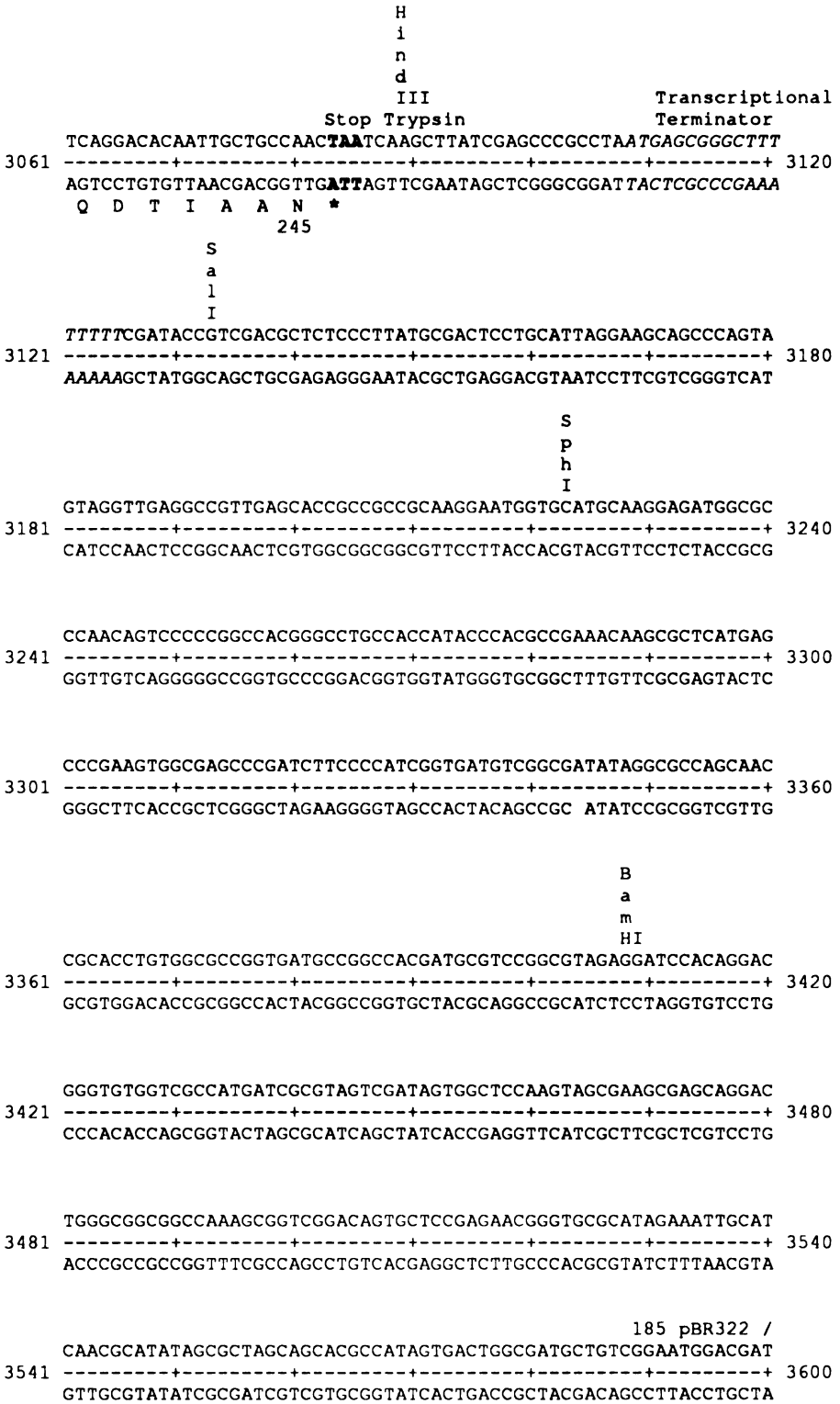
DNA Sequence of the Expression Vector $p^{m}T3$ (from 1-4132)


P
s
t
I
ACACCACGATGCCTGCAGCAATGGCAACAACG]TGCGCAAACTATTAACTGGCGAACTAC
661
TGTGGTGCTACGGACGTCGTTACCGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATG
TTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGAC
721
780
AATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTG
CACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTG
781
840
GTGAAGACGCGAGCCGGGAAGGCCGACCGACCAAAtPPCGACTATTTAGACCTCGGCCAC
AGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCG
841
900
TCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGCATAGC
TAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTG
901
960
ATCAATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGAC
Stop ß-lactamase
AGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATAC
961
1020
TCTATCCACGGAGTGAG AATTCGTAACCATTGACAGTCTGGTTCAAATGAGTATATATG
TTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTG
1021
1080
AAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAAC
ATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCG
1081
1140
TATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGC
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGC
1141
1200
ATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACG
AAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG7 GGATCAAGAGCTACCAACTC
1201
1260
TTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAG
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGT
1261
1320
AAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACA
AGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC
1321
1380
TCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACG

```
     TAATCCTGTTYCCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACT
1381 ---------+---------+---------+---------+---------+---------+ 1440
     ATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGA


     CAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACAC
1441 ---------+---------+---------+---------+---------+---------+ 1500
     GTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTG


     AGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAG
1501 ---------+---------+---------+---------+---------+---------+ 1560
     TCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGTAACTC


     AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCG
1561 ---------+---------+---------+---------+---------+---------+ 1620
     TTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGC


     GAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTG
1621 ---------+---------+---------+---------+---------+---------+ 1680
     CTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGAC


     TCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
1681 ---------+---------+---------+---------+---------+---------+ 1740
     AGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCT


     GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
1741 ---------+---------+---------+---------+---------+---------+ 1800
     CGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAA


     TTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCT
1801 ---------+---------+---------+---------+---------+---------+ 1860
     AACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGA


     TTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCG
1861 ---------+---------+---------+---------+---------+---------+ 1920
     AACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGCGTCGCTCAGTCACTCGC


     AGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCAC
1921 ---------+---------+---------+---------+---------+---------+ 1980
     TCCTTCGCCTTCTCGCGGACTACGCCATAAAAGAGGAATGCGTAGACACGCCATAAAGTG


     ACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATA
1981 ---------+---------+---------+---------+---------+---------+ 2040
     TGGCGTATACCACGTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAT
```


CACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGC
2041 + 2100
GTGAGGCGATAGCGATGCACTGACCCAGTACCGACGCGGGGCTGTGGGCGGTTGTGGGCG
TGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGT
2101 + 2160
ACTGCGCGGGACTGCCCGAACAGACGAGGGCCGTAGGCGAATGTCTGTTCGACACTGGCA
pBR322 2068 /
CTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGAT
2161 + 2220
GAGGCCCTCGACGTACACAGTCTCCAAAAGTGGCAGTAGTGGCTTTGCGCGCTCCGTCTA
H
i
n
c
II
RNA polymerase binding
TCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGw
2221 + 2280
AGACTTTACTCGACAACTGTTAATTAGTAGCCGAGCATATTACACACCTTAACACTCGCC
E
c
o
RI
S
m
a
I
ATAACAATTTCACACAGGAAACAGAATTCCCGGGATCTCTTGGAAAGASAGTCTGTTGAG
2281 + 2340
TATTGTTAAAGTGTGTCCTTTGTCTTAAGGGCCCTAGAGAACCTTTCTGTCAGACAACTC
start trypsin
GATAGTTATGAAAAAACTGGCGTTATCCCTTTCTCTGGTGCTGGCATTTTCCAGCGCTAC
2341 + 2400
CTATCAATACTTTTTTGACCGCAATAGGGAAAGAGACCACGACCGTAAAAGGTCGCGATG
M K K L A L S L S L V L A F S S A T
P
v
u
I
E
c
o
RI
CGCAGCATTTGCGATCGTTGGAGGATACACCTGCCAAGAGAATTCTGTTCCCTACCAAGT
2401 + 2460
GCGTCGTAAACGCTAGCAACCTCCTATGTGGACGGTTCTCTTAAGACAAGGGATGGTTCA
A A F A I V G G Y T C Q E N S V P Y Q V
16
B
s
t
XI
GTCCCTGAACTCTGGCTACCACTTCTGTGGAGGTTCCCTCATCAATGACCAGTGGGTGGT
2461 + 2520
CAGGGACTTGAGACCGATGGTGAAGACACCTCCAAGGGAGTAGTTACTGGTCACCCACCA
S L N S G Y H F C G G S L I N D Q W V V
GTCTGCAGCTCACTGCTATAAGTCCCGCATCCAAGTGAGACTGGGeTtGCACAACATCAA
2521 + 2580
CAGACGTCGAGTGACGATATTCAGGGCGTAGGTTCACTCTGACCCTCTCGTGTTGTAGTT
S A A H C Y K S R I Q V R L G E H N I N
57


TGTCCTTGAGGGCAATGAGCAGTTTGTCAATGCTGCCAAGATCATCAAGCATCCCAACTT
2581 ---------+---------+---------+---------+---------+---------+ 2640
ACAGGAACTCCCGTTACTCGTCAAACAGTTACGACGGTTCTAGTAGTTCGTAGGGTTGAA
V L E G N E Q F V N A A K I I K H P N F
CGATAGGAAGACCCTGAACAACGACATCATGCTGATCAAGCTCTCTTCCCCTGTGAAACT
2641 ---------+---------+---------+---------+---------+---------+ 2700
GCTATCCTTCTGGGACTTGTTGCTGTAGTACGACTAGTTCGAGAGAAGGGGACACTTTGA
D R K T L N N D I M L I K L S S P V K L
102
X
h
o
I
CAATGCTCGAGTGGCCACTGTGGCTCTTCCCAGCTCCTGTGCACCTGCAGGCACTCAGTG
2701 ---------+---------+---------+---------+---------+---------+ 2760
GTTACGAGCTCACCGGTGACACCGAGAAGGGTCGAGGACACGTGGACGTCCGTGAGTCAC
N A R V A T V A L P S S C A P A G T Q C
CCTCATCTCTGGCTGGGGCAACACGCTCAGCAGTGGCGTCAATGAACCAGACCTGCTCCA
2761 ---------+---------+---------+---------+---------+---------+ 2820
GGAGTAGAGACCGACCCCGTTGTGCGAGTCGTCACCGCAGTTACTTGGTCTGGACGAGGT
L I S G W G N T L S S G V N E P D L L Q
GTGCCTGGATGCCCCACTGCTGCCCCAAGCTGACTGTGAAGCCTCCTACCCTGGAAAGAT
2821 ---------+---------+---------+---------+---------+---------+ 2880
CACGGACCTACGGGGTGACGACGGGGTTCGACTGACACTTCGGAGGATGGGACCTTTCTA
C L D A P L L P Q A D C E A S Y P G K I
CACTGACAACATGGTCTGTGTTGGCTTCCTAGAGGGAGGCAAGGATTCCTGCCAGGGTGA
2881 ---------+---------+---------+---------+---------+---------+ 2940
GTGACTGTTGTACCAGACACAACCGAAGGATCTCCCTCCGTTCCTAAGGACGGTCCCACT
T D N M V C V G F L E G G K D S C Q G D
189 190
CTCTGGTGGCCCTGTGGTCTGTAATGGAGAGCTGCAGGGCATTGTCTCCTGGGGCTATGG
2941 ---------+---------+---------+---------+---------+---------+ 3000
GAGACCACCGGGACACCAGACATTACCTCTCGACGTCCCGTAACAGAGGACCCCGATACC
S G G P V V C N G E L Q G I V S W G Y G
195 214 216
S
t
y
I
CTGTGCCCTGCCAGATAACCCTGGTGTGTACACCAAGGTCTGCAACTATGTGGACTGGAT
3001 ---------+---------+---------+---------+---------+---------+ 3060
GACACGGGACGGTCTATTGGGACCACACATGTGGTTCCAGACGTTGATACACCTGACCTA
C A L P D N P G V Y T K V C N Y V D W I
226 228


Hind III
Stop Trypsin
Transcriptional Terminator
TCAGGACACAATTGCTGCCAACTAATCAAGCTTATCGAGCCCGCCTAATGAGCGGGCTTT
3061
AGTCCTGTGTTAACGACGGTTGATTAGTTCGAATAGCTCGGGCGGATTACTCGCCCGAAA
Q D T I A A N *
245
SalI
TTTTTCGATACCGTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTA
3121
AAAAAGCTATGGCAGCTGCGAGAGGGAATACGCTGAGGACGTAATCCTTCGTCGGGTCAT
SphI
GTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGC
3181
CATCCAACTCCGGCAACTCGTGGCGGCGGCGTTCCTTACCACGTACGTTCCTCTACCGCG
CCAACAGTCCCCCGGCCACGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAG
3241
GGTTGTCAGGGGGCCGGTGCCCGGACGGTGGTATGGGTGCGGCTTTGTTCGCGAGTACTC
CCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAAC
3301
GGGCTTCACCGCTCGGGCTAGAAGGGGTAGCCACTACAGCCGC ATATCCGCGGTCGTTG
BamHI
CGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCCACAGGAC
3361
GCGTGGACACCGCGGCCACTACGGCCGGTGCTACGCAGGCCGCATCTCCTAGGTGTCCTG
GGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGAC
3421
CCCACACCAGCGGTACTAGCGCATCAGCTATCACCGAGGTTCATCGCTTCGCTCGTCCTG
TGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCAT
3481
ACCCGCCGCCGGTTTCGCCAGCCTGTCACGAGGCTCTTGCCCACGCGTATCTTTAACGTA
185 pBR322 /
CAACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGACGAT
3541
GTTGCGTATATCGCGATCGTCGTGCGGTATCACTGACCGCTACGACAGCCTTACCTGCTA

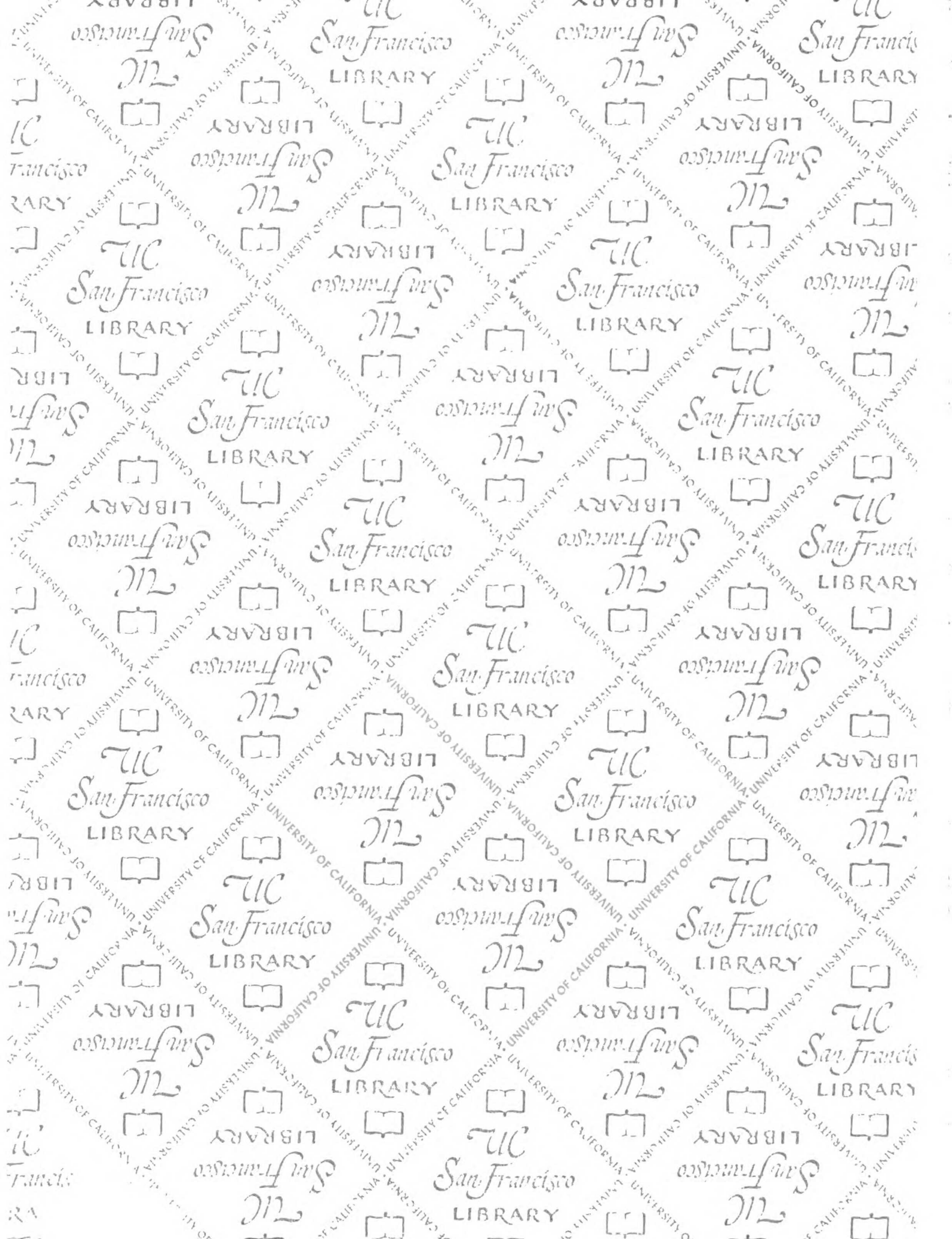
```
                              P
                              v
                              u
     / 527 Bluescript (-)     I
     CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAA
3601 ---------+---------+---------+---------+---------+---------+ 3660
     GACCGCATTATCGCTTCTCCGGGCGTGGCTAGCGGGAAGGGTTGTCAACGCGTCGGACTT


     TGGCGAATGGCGCGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTG
3661 ---------+---------+---------+---------+---------+---------+ 3720
     ACCGCTTACCGCGCTTTAACATTTGCAATTATAAAACAATTTTAAGCGCAATTTAAAAAC


     TTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAA
3721 ---------+---------+---------+---------+---------+---------+ 3780
     AATTTAGTKGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTT


     AGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAA
3781 ---------+---------+---------+---------+---------+---------+ 3840
     TCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTT


     GAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACG
3841 ---------+---------+---------+---------+---------+---------+ 3900
     CTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATGC


     TGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAA
3901 ---------+---------+---------+---------+---------+---------+ 3960
     ACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTT


     CCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAA
3961 ---------+---------+---------+---------+---------+---------+ 4020
     GGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTT


     GGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCT
4021 ---------+---------+---------+---------+---------+---------+ 4080
     CCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGA


     GCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCC
4081 ---------+---------+---------+---------+---------+-- 4132
     CGCGCATTGGTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGG
```

Enzymes that do cut:

AccI	AflIII	AhaII	AluI	AlwI	AlwNI	ApaLI	Asp700I
AvaI	AvaII	BalI	BamHI	BanI	BanII	BbeI	BbvI
BbvII	BceFIX	BclI	BglI	Bsp1286I	BspHI	BspMI	BsrI
BstNI	BstXI	CfrI	Cfr10I	CviJI	CviQI	DdeI	DpnI
DraI	DraIII	DsaI	Eco31I	Eco47III	Eco57I	Eco78I	EcoNI
EcoRI	EcoRII	EspI	FinI	Fnu4HI	FokI	FspI	GdiII
GsuI	HaeI	HaeII	HaeIII	HgaI	HgiAI	HgiEII	HhaI
HincII	HindIII	HinfI	HinP1I	HpaII	HphI	Ksp632I	MaeI
MaeII	MaeIII	MboII	MfeI	MmeI	MnlI	MseI	NaeI
NarI	NciI	NdeI	NheI	NlaIII	NlaIV	NspBII	NspHI
PflMI	PleI	PstI	PvuI	RsaI	SalI	Sau3AI	Sau96I
ScaI	SciI	ScrFI	SecI	SfaNI	SmaI	SphI	SsoII
SspI	StyI	TaqI	TaqII	ThaI	Tsp45I	TspEI	Tth111I
Tth111II	Uba26I	VspI	XcaI	XhoI	XhoII	XmaI	

Enzymes that do **not** cut:

AatII	AflII	AocI	ApaI	Asp718I	AsuII	AvrII	BglII
BsmI	BspMII	BssHII	BstEII	ClaI	DraII	EcoRV	HpaI
KpnI	MluI	NcoI	NotI	NruI	NsiI	PmaCI	PpuMI
PssI	PvuII	RsrII	SacI	SacII	SfiI	SnaBI	SpeI
SplI	StuI	XbaI	XmaIII				


UC
San Francisco
LIBRARY
UNIVERSITY OF CALIFORNIA · UNIVERSITY OF CALIFORNIA