UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Defining the extended substrate specificity of serine proteases

Permalink
https://escholarship.org/uc/item/89k464hj

Author
Harris, Jennifer Leslie

Publication Date
1999

Peer reviewed|Thesis/dissertation

eScholarship.org
Powered by the California Digital Library
University of California

Defining the Extended Substrate Specificity of Serine Proteases

by

Jennifer Leslie Harris

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

in the

GRADUATE DIVISIONS

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO

and

UNIVERSITY OF CALIFORNIA BERKELEY

[redacted]

....

Date University Librarian

Degree Conferred: ....

All our science, measured against reality, is primitive and
childlike—and yet it is the most precious thing we have.
Albert Einstein
(1879-1955)

To my mother, Leslie,
my father, Roger,
and my sisters,
Lara and Sarah.

UCSF LIBRARY

Preface

When I look back over my years in graduate school I am amazed that I made it! I don't simply mean obtaining the diploma, I mean the gradual transformation of becoming a scientist. I know well enough that the transformation from student to scientist does not occur within a vacuum and I am endlessly grateful to colleagues, family, and friends that have influenced my development.

I thank my thesis advisor Charles Craik for creating a lab environment conducive to scientific learning and creativity. I appreciate him for allowing me the latitude to run with my projects and for picking me up and helping me dust myself off when I stumble and fall. I have had an exceptional graduate school experience and I owe a lot of that to Charly.

The other two members of my thesis committee, Robert Fletterick and Zena Werb, have been excellent role models for me on how to run a lab and do productive research. I can't say enough of my appreciation for the support they have given me over the years. Their doors were always open when I was in need of advice.

I am humbled by the scientists with whom I have had the pleasure of interacting with and learning from at UCSF. Robert Maeda, a fellow graduate student in the lab and lab saint, has not only offered a critical and constructive sounding board for my scientific problems, but has also comforted me in times when I felt my whole world was caving in. I aspire to be the scientist and person that Rob is. He and his wife, Annick Mutero, are

UCSF LIBRARY

some of my dearest friends. Kathlynn Brown, a former postdoctoral fellow in the lab, has a love of science and learning that is contagious to those around her. She constantly encouraged me to strive to become a more rigorous and knowledgeable scientist. Nancy Douglas, the former lab supervisor, is one of those marvelous people that make you feel like a million-bucks when you are around them. She was very much responsible for building my self-esteem as a scientist and making me realize that I had what it takes to succeed. Sherin Halfon (Scrunchy), a former postdoctoral fellow, very patiently answered my questions and showed me how to run my first agarose and SDS-PAGE gel. John J. Perona (J.P.), a former postdoctoral fellow and my rotation advisor, requires my gratitude for helping me to question myself and to realize that doubt is a necessary component of science. Toshihiko Takeuchi, postdoctoral fellow in the lab, lab softball team captain, and violin teacher, constantly amazes me with his talents—which he freely shares. Cheryl Lanzo, Iain Murray, Cheng-I Wang, Sami Mahrus, Sandra Waugh, Sarah Gillmor, Qing Yang, Todd Pray were postdoctoral fellows and graduate students that, while our time of overlap was brief, had significant impact on my scientific career. All of these people are good friends of mine and I cherish their friendship.

I have also been fortunate to have many great collaborators outside of UCSF. First and foremost is Bradley Backes, a former graduate student in the Ellman lab in Berkeley. Brad has allowed me to expand my scientific horizons and has given me the confidence to wonder. Nancy Thornberry, Jon Ellman, Seenu Kothakota, Rusty Williams, and Dorothy Hudig have provided essential experimental advice, support and reagents. None of my work would have been possible without the help and support from these people.

I thank my family, Dad, Mom, Sarah and Lara, who taught me to be curious about the world around me, encouraged me in my pursuits, dried my tears of frustration and sorrow, listened to my problems without censure, provided astute advice, celebrated with me in times of joy and success, and throughout it all showed me unconditional love and support.

Finally, I would like to thank Keith, my dance partner, golf partner and best friend. He has brought balance to my life and is constantly putting a smile on my face.

UCSF LIBRARY

Abstract

Defining the Extended Substrate Specificity of Serine Proteases

Jennifer Leslie Harris

Proteases play essential roles in every biological process. The ability of a protease to discriminate among many potential substrates is a key factor in biological regulation by proteases. One mechanism that facilitates substrate discrimination is the substrate specificity determinants contained within the protease. Definition of the preferred substrate cleavage sequence, or the substrate specificity, of a protease can aid in the identification of its macromolecular substrates and thus provide insight into the biological pathway that the protease is functioning in. A complete examination of a protease's substrate specificity can also be applied to the design of sensitive and selective substrates and inhibitors.

To examine the extended substrate specificity of serine (and cysteine) proteases, generalizable tools were developed. These included the traditional use of single peptide substrates, and the combinatorial techniques of positional scanning synthetic combinatorial libraries (PS-SCL), substrate-phage display, and small pool cDNA expression and cleavage screening. Proteases involved in the blood coagulation and anticoagulation pathway, lysosomal cysteine proteases, and proteases involved in mast cell function, were profiled to validate a novel PS-SCL synthetic strategy.

To understand the role that granzyme B plays in cytotoxic lymphocyte mediated cell death, the substrate specificity from P4 to P2' was defined using a PS-SCL and

substrate-phage display. The P4-P2' sequence determined from these methods, (Ile/Val)-(Glu/Gln/Met)-(Xaa)-(Asp)-(Xaa)-(Gly), was used to identify potential macromolecular downstream substrates of granzyme B. The complementary technique of small pool cDNA expression screening was used to identify potential macromolecular substrates of granzyme B in the mouse proteome. The macromolecular substrates included the caspases, lamin B, c-abl, among others.

A model of granzyme B was constructed to elucidate the structural determinants of the extended substrate specificity. Arginine at position 192 was shown through mutagenesis studies to have a dual role in determining the P1 specificity for aspartate as well as the P3 specificity for glutamate.

The substrate specificity from P4-P1 of human and mouse granzyme A was determined from PS-SCL's and found to be (Gly)-(Tyr)-(Phe)-(Arg) for mouse granzyme A and (Ile)-(Ala)-(Asn)-(Arg) for human granzyme A. The functional non-equivalence of these enzymes raises issues of classification of enzymes based on structural similarities versus functional similarities.

UCSF LIBRARY

Table of Contents

Page

Chapter One **1**

Introduction to proteolytic substrate specificity

Chapter Two **18**

Protein engineering of substrate specificity

Chapter Three **31**

Synthesis of positional-scanning libraries of fluorogenic peptide substrates that incorporate diverse P1 substituents: extended specificity determination of plasmin and thrombin

Chapter Four **60**

Rapid and general methods for profiling protease specificity: libraries of novel 7-amino-4-acetamidecoumarin (AAC) fluorogenic substrates

Chapter Five **97**

Definition and redesign of the extended substrate specificity of granzyme B

Chapter Six **138**

Identification of macromolecular substrates of granzyme B by a small pool cDNA expression cleavage screen

Chapter Seven **155**

Mouse and human granzyme A display different extended substrate specificities

Chapter Eight **173**

Future directions

Appendix **177**

Publishers' permission to include Chapter Five

References **179**

UCSF LIBRARY

List of Tables

Page

Table 1-1. 9
Steady state kinetic constant, k_{cat}/K_m, versus Suc-A-A-P-Xaa-para nitroanalide for fiddler crab collagenase.

Table 1-2. 10
Apparent inhibition constant, K_i*, for Fiddler Crab Collagenase against ecotin with mutations of the P1-residue, position 84, and P1-P1' residues, position 84-85, and of the single disulfide bond in ecotin, between amino acids 50 and 87.

Table 1-3. 14
Steady state kinetic constants of rat anionic wild-type trypsin and trypsin [C191F, C220G].

Table 1-4. 16
Sequence identities (%) between granzymes.

Table 3-1. 59
Steady state kinetic constants, k_{cat} , K_m , and k_{cat}/K_m for amino-methylcoumarin substrates for plasmin and thrombin.

Table 4-1. 95
Fluorescence properties of 7-amino-4-methylcoumarin and 7-amino-4-acetamidecoumarin.

Table 4-2. 96
Steady state kinetic constants, k_{cat} , K_m , and k_{cat}/K_m for amino-acetamide coumarin substrates for human lung and skin tryptase.

Table 5-1 133
Purification of recombinant rat granzyme B.

Table 5-2 134
Steady-state kinetic parameters for hydrolysis of substrates by granzyme B.

Table 5-5 **135**
Steady-state kinetic parameters for the hydrolysis of substrates by granzyme B and variants.

Table 5-3 **136**
Mechanistic kinetic parameters for the hydrolysis of amide substrates by granzyme B and variants.

Table 5-4 **137**
P3, P1', and P2' substrate specificity of granzyme B determined by substrate phage.

UCSF LIBRARY

List of Figures

Page

Figure 1-1A. 11
Statistics from structure of fiddler crab collagenase bound to ecotin [M84A].

Figure 1-1B. 12
Superposition of collagenase bound to wild-type ecotin and ecotin [M84A].

Figure 1-2 13
Structure of trypsin and alignment of rat trypsin with rat granzyme B.

Figure 1-3. 15
Chemical structure of the traditionally used fluorophore, 7-amino-4-methylcoumarin, and the novel bifunctional fluorophore, 7-amino-4-acetamidecoumarin.

Figure 1-4. 17
Three steps of cytotoxic lymphocyte-mediated cell death.

Figure 2-1. 27
Schematic diagram of approaches used to engineer altered enzyme specificity.

Figure 2-2. 28
The structure of ATP and ATP-analog with the v-Src kinase amino acids.

Figure 2-3. 29
A divalent metal coordinates the two engineered histidine residues in trypsin.

Figure 2-4. 30
DNA-shuffling.

Figure 3-1. 51
Schematic diagram of the synthesis of P1-Lys PS-SCL.

Figure 3-2A. 53
Histogram of activity versus amino acid residue for plasmin in P1-Lys PS-SCL.

Figure 3-2B. 54
Histogram of activity versus amino acid residue for thrombin in P1-Lys PS-SCL.

UCSF LIBRARY

Figure 3-3A. 55
Three-dimensional model of plasmin bound to substrate.

Figure 3-3B. 56
Three-dimensional model of thrombin bound to substrate

Figure 3-4. 57
Physiological substrate cleavage sites of plasmin.

Figure 3-5. 58
Physiological substrate cleavage sites of thrombin.

Figure 4-1. 76
Schematic diagram of synthesis strategy for 7-amino-4-acetamide resin

Figure 4-2. 77
Histograms of thrombin activity with a P1-Lys-AMC PS-SCL versus a P1-Lys-AAC PS-SCL.

Figure 4-3.
Activity of enzymes in P1-Fixed (arginine, lysine or leucine) PS-SCL.

A. Human thrombin (P1-Arg) 79
B. Human uPA (P1-Arg) 80
C. Human tPA (P1-Arg) 81
D. Human Factor Xa (P1-Arg) 82
E. Cruzain (P1-Arg) 83
F. Papain (P1-Arg) 84
G. Human Lung Tryptase (P1-Arg) 85
H. Human Skin Tryptase (P1-Arg) 86
I. Human Lung Tryptase (P1-Lys) 87
J. Human Skin Tryptase (P1-Lys) 88
K. Human Plasmin (P1-Lys) 89
L. Cruzain (P1-Leu) 90

Figure 4-4.
Activity of enzymes in P1-Diverse library.

A. Rat anionic trypsin 91
B. Bovine chymotrypsin 91
C. Rat granzyme B 91
D. Fiddler crab collagenase 92
E. Human plasmin 92
F. Human thrombin 92
G. Human skin tryptase 93
H. Human granzyme A 93

UCSF LIBRARY

I. Human neutrophil elastase 93
J. Papain 94
K. Cruzain 94
L. Bovine cathepsin L 94
M. Bovine cathepsin B 94

Figure 5-1. **124**
Expression and purification of granzyme B in *Pichia pastoris.*

Figure 5-2. **126**
P4-P2 substrate specificity profile of recombinant granzyme B and the variants granzyme B [R192A] and granzyme B [R192E].

Figure 5-3. **128**
Schematic representation of substrate-phage display used to determine the P3, P1', and P2' substrate specificity profile for granzyme B.

Figure 5-4. **130**
Three dimensional model of granzyme B complexed to IEPDWG substrate.

Figure 5-5. **132**
Biological significance of granzyme B P4-P2' substrate specificity.

Figure 6-1. **148**
Schematic representing small pool cDNA expression screening.

Figure 6-2. **149**
Schematic representing cleavage fragments of lamin B and caspase 3.

Figure 6-3. **150**
Autoradiography gels of the deconvolution of cDNA pool 50.13 (Lamin B).

Figure 6-4. **151**
Protein sequence of Lamin B.

Figure 6-5. **152**
Immunoblot of *in vitro* cleavage of lamin B by granzyme B and caspases.

Figure 6-6. **153**
Autoradiogram of *in vitro* cleavage of lamin B by granzyme B.

Figure 6-7. **154**
Immunoblot of *in vitro* cleavage of caspase-3 and lamin B by granzyme B.

UCSF LIBRARY

Figure 7-1. 162
Three-dimensional dimeric model of granzyme A.

Figure 7-2A. 164
Coomassie stain of recombinant mouse granzyme A.

Figure 7-2 B. 164
Gelatin zymogram of mouse granzyme A.

Figure 7-3. 166
Coomassie stain of recombinant human and mouse granzyme A.

Figure 7-4. 168
MALDI mass spectrum of human granzyme A.

Figure 7-5. 169
Mouse and Human Granzyme A run against the P1-Diverse library.

Figure 7-6A. 171
Extended (P4-P2) substrate specificity of mouse granzyme A run against the P1-Arg PS-SCL.

Figure 7-6B. 172
Extended (P4-P2) substrate specificity of human granzyme A run against the P1-Arg PS-SCL.

UCSF LIBRARY

Chapter One:

Introduction to Proteolytic Substrate Specificity

UCSF LIBRARY

My graduate studies have focused on understanding the recognition of a substrate by an enzyme, in particular, the recognition and cleavage of a substrate by a proteolytic enzyme. Proteases play important roles in every biological process. This has become impressed on me over my years in graduate school as proteases have been shown to be essential in development (Chasan and Anderson, 1989; Kuida *et al.*, 1998; Sun *et al.*, 1998), homeostasis (Davie *et al.*, 1991; Goldberg, 1995; Werb, 1997), and cell death (Muzio *et al.*, 1996; Smyth *et al.*, 1996). Moreover, proteases are typically implicated when biological processes go awry, such as in cancer, neurodegeneration, hemophilia, atherosclerosis, asthma, etc.(Coussens *et al.*, 1999; Fath *et al.*, 1998; van der Wal and Becker, 1999). While proteases have been implicated in these processes, information on their exact function is, more often than not, lacking. One very important piece of information that can aid in dissecting proteolytic function is the identification of the protease's downstream substrates. Once this is known, other questions can be asked, such as: what is the result of the cleavage event; is it an activating or deactivating cleavage; where does it occur in the organism; can it be inhibited?

The approach that I have taken to understanding proteolytic function, is through the definition of the optimal substrate specificity displayed by a protease. Knowing if a protease has a preferred specificity in and beyond primary P1[*] specificity can first give clues as to whether the protease plays a regulatory role, manifested by a limited substrate preference, or a digestive role, manifested by a promiscuous substrate preference. Furthermore, knowing the substrate specificity of a protease can aid in the identification of physiological substrates through database searches with the optimal cleavage sequence. Alternatively, if a physiological substrate has been identified through other means, knowing the optimal substrate preference can aid in identifying the cleavage site within that substrate. Functional knowledge of substrate specificity is necessary for the identification of structural specificity determinants of a protease. Knowing the strucural determinants can aid in the redesign of substrate specificity through protein engineering. Finally, definition of the substrate specificity of a protease can guide the design of sensitive and selective substrates and inhibitors.

---

[*] Nomenclature for the substrate amino acid preference is Pn, Pn-1,...P2, P1, P1', P2',..., Pm-1', Pm'. Amide bond hydrolysis occurs between P1 and P1'. Sn, Sn-1,..., S2, S1, S1', S2',..., Sm-1', Sm' denotes the corresponding enzyme binding sites (Schechter, 1968).

My initial efforts on understanding protease substrate specificity were focused on the S1 site of the digestive proteases, collagenase and trypsin. A graduate student in the Craik laboratory, Christopher Tsu, had just finished his thesis work on the characterization of a serine collagenase from fiddler crab that had the ability to cleave after every amino acid, except acidic or beta-branched aliphatic amino acids (Table 1-1). My goal was to understand the structural basis of the broad specificity. To this end, I engineered six mutations (M84R, M84K, M84F, M84A, M84Q M85R, C50A, C87A) into the primary binding loop of ecotin, a dimeric inhibitor of serine proteases of the chymotrypsin-like fold. The inhibition constants were measured against collagenase and shown to not differ greatly between the mutants (Table 1-2). This lack of correlation between substrate specificity and inhibitor specificity highlights the basis of ecotin's ability to inhibit proteases with very different P1-specificities, the use of a secondary site interaction (Yang *et al.*, 1998). The ecotin mutants were then complexed to collgenase as a vehicle to deliver the P1-amino acid to the S1-pocket of the enzyme. The protease inhibitor complexes were crystallized under conditions of polyethylene glycol (4000 g/mol, 20-40%), 100 mM Tricine (pH 7.5-8.5), and 200 mM sodium citrate. The unit cell of the crystals were typically, a=b=90 Å, c=290 Å. Because of the long c-axis, high-resolution data was hard to obtain and completeness usually did not extend beyond 2.5 Å. I collected x-ray diffraction data and refined the strucutre of ecotin [M84A] bound to collagenase (Figure 1-1A). At 2.6 Å resolution data, the mutant structure, with the P1-position occupied by an Ala, was indistinguishable from the wild-type structure (Perona *et al.*, 1997), with the P1-position occupied by a Met (Figure 1-1B). This indicated that the S1-pocket of collagenase

UCSF LIBRARY

was rigid and did not go through major structural changes to accomodate smaller P1-amino acids. More subtle structural changes were not going to be available from this data.

The second approach I took at this time to understand P1-substrate specificity was to redesign trypsin to accept P1-acidic amino acids over basic amino acids, through engineering. The design was guided by comparitive analysis of trypsin to the only known mammalian serine protease to have P1-acidic specificity, granzyme B. Granzyme B belongs to a subclass of serine proteases of the chymotrypsin class. Other members of this class include granzymes C, D, E, F, G, H, mast cell chymases, and cathepsin G. Members of the granzyme B subclass have the following structural characteristics that distinguish them from other proteases with a chymotrypsin-like fold: absence of the disulfide bond (Cys 191-Cys 220) near the active site, a strictly conserved sequence from amino acid 24-31, a three amino acid insertion in the 30's loop and a deletion in loop2 (222-224) (Figure 1-2). I was pleased to find that upon removal of the disulfide bond and Cys191-Cys 220 there was only an ~500-fold decrease in the activity of trypsin (Table 1-3). On this disulfide minus trypsin scaffold several additional mutations were made based on the granzyme B sequence. This project was passed on to a student who was rotating with me at the time, Sandra Waugh, and I will leave it to her thesis to finish the story.

My conclusions from these various protein engineering efforts was that we do not really understand what drives substrate specificity. The lack of understanding is due in part to the lack of information. I deal with this issue in Chapter 2 of my thesis, an article that originally appeared in *Current Opinion in Chemical Biology* (Harris and Craik, 1998).

Tools to adress P1-specificity have been in place for several decades. Traditionally, to access the P1-specificity of a protease, kinetic analysis of a panel of substrates that differ in the P1-position is done (for example, see previous paragraph on collagenase and Table 1-1). If the protease has only one position of specificity, as may be the case with digestive proteases, this approach is tenable and only requires the synthesis and analysis of 20 substrates. However, if multiple sites define the substrate specificity, as may be the case with regulatory proteases, the traditional approach quickly gets out of hand. For example, to define the substrate specificity for two positions requires the synthesis and analysis of 400 substrates, three positions require 8000 substrates, four positions require 160,000 substrates, etc. To make the analysis more amenable to a graduate student time-scale and to an NIH-funding resource scale, combinatorial tools were developed to address the extended specificity requirements of regulatory proteases. I utilized two major combinatorial methods, substrate phage display and positional scanning-synthetic combinatorial libraries (PS-SCL). I will discuss my modification of substrate phage display in Chapter 4 from an article that originally appeared in the *Journal of Biological Chemistry* (Harris *et al.*, 1998).

The development and use of positional scanning-synthetic combinatorial libraries came from an invaluable collaboration with Bradley J. Backes, a chemist in Jonathan A. Ellman's group at Berkeley. Initially, we worked on a strategy that utilized an alkanesulfonamide linker (Backes and Ellman, 1999), a linker that Brad developed in graduate school, to construct PS-SCL 7-amino-4-methylcoumarin libraries that could incorporate various P1-substituents. Using this library we profiled the blood coagulation enzymes plasmin and thrombin. This work has been submitted for publication and is presented in Chapter 3 of my thesis.

The collaboration between Brad and myself further developed into the design of a strategy to prepare synthetic libraries that could introduce diversity at all positions N-terminal to the scissile bond in a straightforward manner. This was accomplished through the development of a novel bifunctional-fluorogenic leaving group, 7-amino-4-acetimidecoumarin (AAC), that has enhanced fluorescence properties over the traditionally used fluorphore, 7-amino-4-methylcoumarin (AMC) (Figure 1-3). The major benefit of this fluorophore is that it can be linked to a solid support and libraries can be readily synthesized using standard fmoc peptide-chemistry. Tetrapeptide libraries generated by this strategy were used to define the extended substrate specificity of serine proteases involved in haemostasis and cysteine proteases involved in lysosomal degradation and antigen processing. The methodology and results will be discussed in Chapter 4. This strategy presents the possibility for rapid and accurate definition of the extended substrate specificity of virtually any protease.

In the process of my intellectual transition from digestive to regulatory proteases, the question still remained of what the minimal number of determinants required to convert one chymotrypsin-like enzyme to another were. Evolutionarily this might have already happened in the granzyme family and I thought that the question of minimal requirements could be more successfully addressed by studying the specificity of its members. The high sequence identity (31-90%) coupled with the diversity of specificity among the members of the granzyme family make it a useful system for probing how a chymotrypsin-like scaffold may be used to engineer specificity through small sequence changes (Table 1-4). In addition, the granzymes are implicated in the very important biological process of clearing the body of viral-infected and tumor cells (Figure 1-4). Knowledge of their substrate specificity could give insight into their biological function.

To begin examining the substrate specificity of the granzymes, I had the opportunity to work with Professor Dorothy Hudig from the University of Reno Nevada. She provided the cDNA clone of rat granzyme B as well as cDNA fragments of rat granzyme C and granzyme F (Ewoldt *et al.*, 1997). In addition, I subcloned from a mouse spleen cDNA library the genes for mouse granzymes A and B. With the help of Sami Mahrus, a talented young graduate student in the Craik lab, we subcloned *in silica* from the EST (Expressed Sequence Tagged) database, the genes for human granzymes A, K and M. The heterologous expression system that finally worked for the production of active granzymes was the methylotrophic yeast, *Pichia pastoris*. I expressed and purified rat granzyme B and the rat granzyme B mutants, Asn 66 Gln (removal of the glycosylation site), Arg 192 Ala (P3-specificity redesign), and Arg 192 Glu (P3-specificity redesign). I showed that granzyme B displayed extended substrate specificity from P4-P2' through the use of single substrate kinetic analysis, substrate-phage display, and positional scanning-synthetic combinatorial P1-Asp-AMC library provided by Nancy Thornberry at Merck Laboratories. The preferred substrate sequence for granzyme B is found in the activation sequences of many of the pro-caspases, endogenous cysteine proteases responsible for effecting apoptosis. Definition and redesign of the extended substrate specificity of granzyme B is described in Chapter 5, an article that appeared in the *Journal of Biological Chemistry* (Harris *et al.*, 1998). I also expressed granzymes C and F but they will not be discussed in this thesis. The x-ray crystallographic strucural analysis of granzymes B, C and F is currently being carried out by Sandra Waugh. These structures should yield insight into the strucutral determinants of substrate specificity. Mouse granzyme A and human granzymes A, K and M were also expressed and purified in *Pichia pastoris* with the help of Sami Mahrus, who will be pursuing their characterization.

Moving even farther into the physiological significance of the substrate specificity manifested by a protease, I performed a small-pool cDNA expression cleavage screen. The 14-day mouse embryo cDNA library was provided by Seenu Kothakota, a scientist in Rusty William's lab at Chiron. Using this library I screened 200,000 cDNA clones for cleavage susceptibility by granzyme B. From this library multiple macromolecular substrates were isolated, including nuclear lamin B. The results from this screen will be discussed in Chapter 6.

The most abundant protease in the granules of cytotoxic lymphocytes is granzyme A. The function of granzyme A in cytotoxic lymphocyte-mediated cell death has not been as appreciated as that of granzyme B. The mouse knock-out of granzyme A had no distinguishable phenotype in *in vitro* cytotoxicity assays. One of the existing questions associated with the granzymes is why there are so many more granzymes found in the mouse and rat than in the human (Table 1-4). Is it possible that what have been defined as homologs based on sequence may not be homologs based on function? Could it be that mouse granzymes have different substrates than human granzymes? Is mouse cytotoxic lymphocyte mediated cell death necessarily a good model for human cytotoxic lymphocyte - mediated cell death? One way to address these issues is to examine the equivalencies between the extended substrate specificities of the granzymes. For granzyme B we saw equivalent extended specificities for the rat and the human. However, when we examined the extended specificity of mouse and human granzyme A, we found them to be significantly different. What this means in terms of granzyme A's physiological function in each of the species has yet to be determined. What we can say is that they are functionally different enzymes and comparisons between them should be handled with caution. The results from this study are presented in Chapter 7.

As the era of genome sequencing comes to a close, new challenges arise for the analysis and characterization of the identified proteins. The development of assays for the rapid functional characterization of proteins will significantly aid in understanding the complex biological processes of life. My hope is that I will have contributed, if only a drop of water in the vast ocean, to the intellectual significance of examining proteolytic substrate specificity and to the tools that enable that examination.

Table 1-1.

Steady State Kinetic Constant, k_{cat}/K_m, versus Suc-A-A-P-Xaa-para-nitroanalide for Fiddler Crab Collagenase (data taken from (Tsu *et al.*, 1994)).

Substrate (Xaa)	**k_{cat}/K_m (/min/μM)**
Ala	0.10
Abu	0.53
Nva	3.50
Val	0.0044
Nle	1.8
Leu	7.3
Ile	0.0027
Met	3.0
Phe	7.4
Gln	0.74
Arg	81
Lys	3.1
Orn	0.38
Asp	0.0010
Glu	0.0031

Table 1-2.

Apparent inhibition constant, K_i*, for Fiddler Crab Collagenase against ecotin with mutations of the P1-residue, position 84, and P1-P1' residues, position 84-84, and of the single difulide bond in ecotin, positions 50 and 87.

Ecotin	**K_i* (pM)**
Wild Type (M84)	70
M84R	110
M84K	150
M84F	110
M84A	160
M84Q and M85R	90
C50A and C87A	110

UCSF LIBRARY

Figure 1-1A.

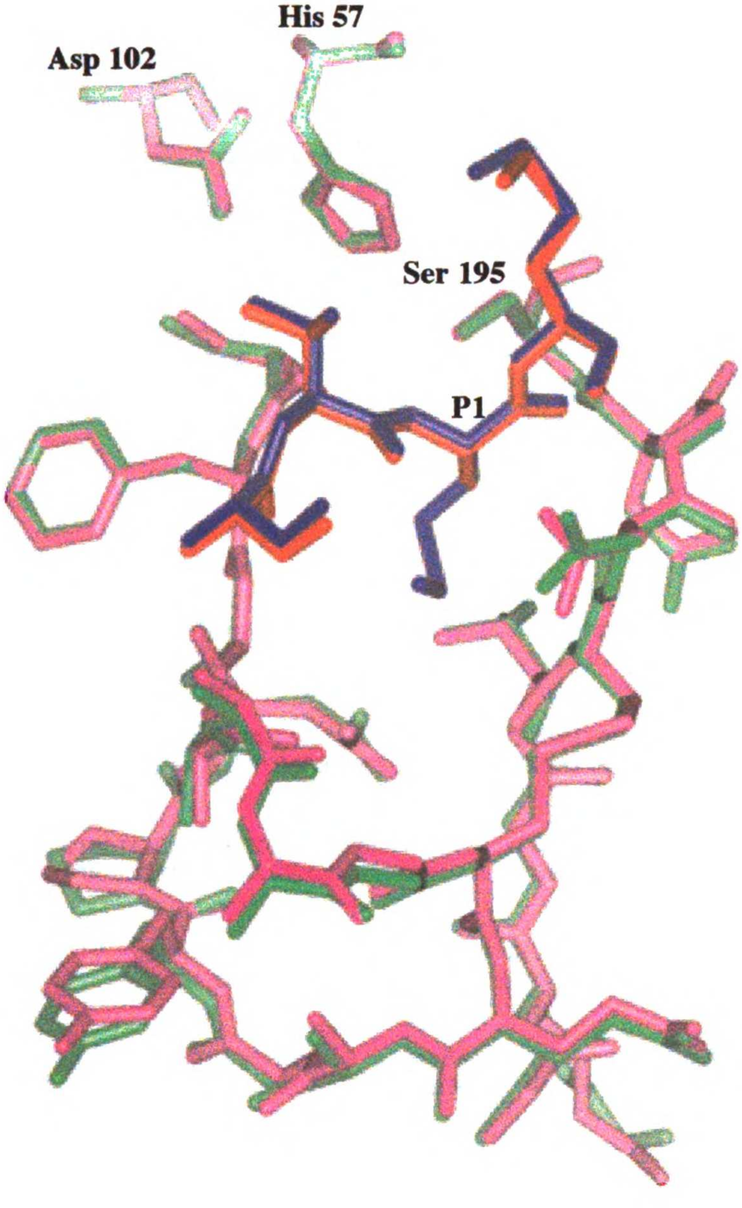
Statistics from structure of fiddler crab collagenase bound to ecotin [M84R].

Space Group	$P3_221$
Cell Dimensions (Å)	a=b=89.39, c=293.51 (wt, a=b=89.11, c=291.55)
Resolution	2.6 Å
Total Observations	42478
Unique Observations	28252
R_{merge}	0.0577
R_{free}	0.248 (6-2.5 Å)
R_{cryst}	0.201 (6-2.5 Å)
Rms_{bonds}	0.011 Å
Rms_{angles}	2.776°

UCSF LIBRARY

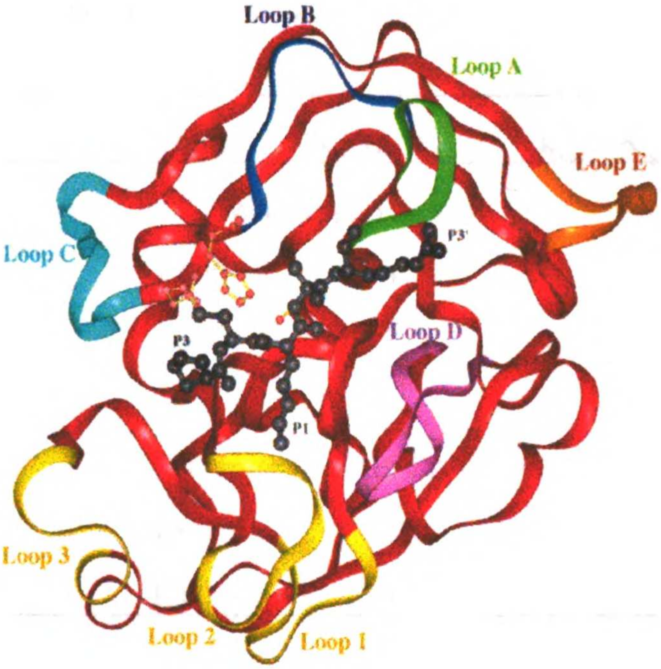
Figure 1-1B.

Superposition of wild-type ecotin (blue) complexed to Collagenase (magenta) and ecotin [M84A] (red) complexed to fiddler crab collagenase (green).


His 57
Asp 102
Ser 195
P1

UCSF LIBRARY

Figure 1-2 Structure of trypsin and alignment of trypsin with granzyme B.



```
                     16                    LOOP A                   57
Trypsin              IVGGYTCPEH SVPYQVSL.. ...NSGYHFC GGSLINDQWV VSAAHC....
Granz B              IIGGHEAKPH SRPYMAYLQI MDEYSGSKKC GGFLIREDFV LTAAHC....

                 LOOP B            LOOP E                   LOOP C
Trypsin      ......YKSR IQVRLGEHNI NVLEGD.EQF INAAKIIKHP NYS.SWTLNN
Granz B      ......SGSK INVTLGAHNI KEQEKM.QQI IPVVKIIPHP AYN.SKTISN

             102                                             LOOP D
Trypsin      DIMLIKLSSP VKLNARVAPV ALPSAC.... .APAGTQCLI SGWGNTLSN.
Granz B      DIMLLKLKSK AKRSSAVKPL NLPRRN.... .VKPGDVCYV AGWGKLGPM.

                                       LOOP 3               LOOP 1
Trypsin      ....GVNNPD LLQCVDAPVL SQADCEAAYP GE.ITSSMIC VG...FLEGG
Granz B      ....G.KYSD TLQEVELTVQ EDQKCESYLK NYFDKANEIC AG...DPKIK

                   195                           LOOP 2
Trypsin      KDSCQGDSGGP VVCNGQ.... ..LQGIVSWGY GCALPDNPGV YTKVCNFVGW IQDTIAAN~~
Granz B      RASFRGDSGGP LVCKKV.... ..AAGIVSYGQ N..DGSTPRA FTKVSTFLSW IKKTMKKS~~
```

UCSF LIBRARY

Table 1-3.

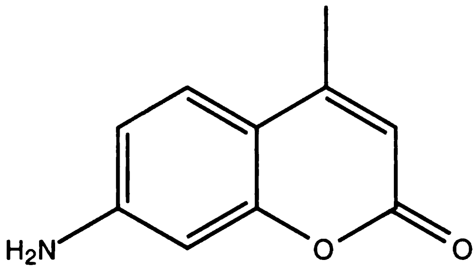
Steady state kinetic constants of rat anionic wild-type trypsin and trypsin [C191F, C220G].

Substrate	**k_{cat} (/min)**	**K_m (μM)**	**k_{cat} / K_m (/min/μM)**
	Wild-Type Trypsin		
Suc-AAPR-pNA	11,000	1.3	8400
Suc-AAPK-pNA	4200	3.2	1300
	Trypsin [C191F, C220G]		
Suc-AAPR-pNA	4000	280	14.3
Suc-AAPK-pNA	1800	430	4.2

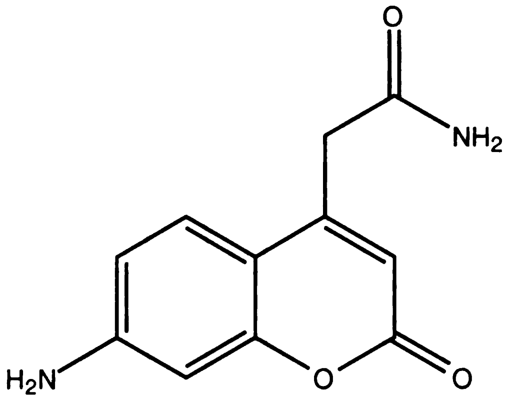
Figure 1-3.

Chemical Structure of the traditionally used fluorophore, 7-amino-4-methylcoumarin, and the novel bifunctional fluorophore, 7-amino-4-acetamidecoumarin.

7-amino-4-methylcoumarin (AMC)



7-amino-4-acetamidecouamrin (AAC)



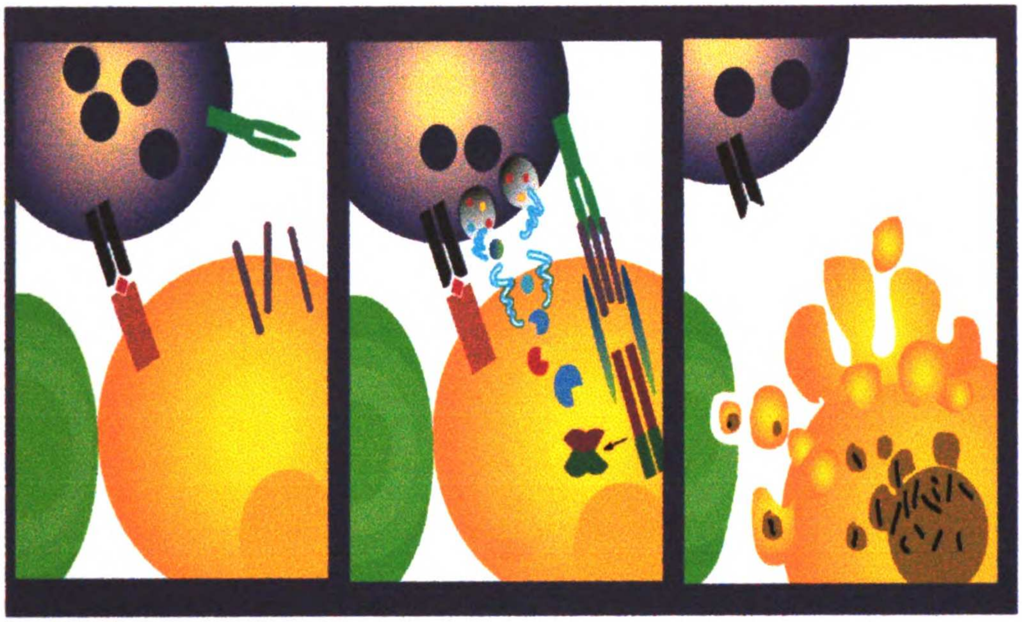
UCSF LIBRARY

Table 1-4. Sequence identities (%) between granzymes (Gr) from human (h), mouse (m), and rat (r).

	A Class of Granzymes								**B Class of Granzymes**									
% Identity	h Gr A	m Gr A	h Gr K	m Gr K	r Gr K	h Gr M	m Gr M	r Gr M	h Gr B	m Gr B	r Gr B	m Gr C	r Gr C	m Gr D	m Gr E	m Gr F	r Gr F	m Gr G
m Gr A	70																	
h Gr K	44	43																
m Gr K	45	45	74															
r Gr K	46	44	73	88														
h Gr M	36	35	36	35	37													
m Gr M	32	33	35	35	37	68												
r Gr M	34	33	34	35	37	68	86											
h Gr B	38	38	36	36	37	38	35	34										
m Gr B	41	42	36	36	38	38	36	35	69									
r Gr B	40	41	38	36	38	37	38	36	69	81								
m Gr C	37	38	36	35	36	37	36	34	60	67	66							
r Gr C	38	39	36	35	36	37	36	34	58	67	66	85						
m Gr D	37	36	33	34	33	32	31	30	57	57	55	56	54					
m Gr E	36	37	31	32	33	34	32	31	53	56	53	55	56	90				
m Gr F	40	38	34	34	35	36	34	34	52	57	53	56	57	71	74			
r Gr F	36	37	33	32	35	35	36	35	55	58	56	57	56	66	69	71		
m Gr G	38	37	34	35	36	35	33	32	55	58	58	58	60	77	81	78	71	
h Gr H	39	38	36	34	36	37	36	35	70	62	62	59	58	57	57	57	59	57

Figure 1-4.

Schematic representation of the stages of cytotoxic lymphocyte (CL)-mediated cell death.



1. Recognition of the virus-infected or tumor cell by the CL.

2. Lethal hit by two contact dependent mechanisms: granule exocytosis or ligation of fas/fas-receptor.

3. Apoptosis of the target cell. CL disassociates to kill again.

Chapter Two:

Protein Engineering of Substrate Specificity

LIBRARY

Abstract

Protein engineering is the application of knowledge to design and alter protein function and structure. While powerful methods, from specific to random, have been developed for the redesign of protein architecture, their successful application is dependent on the information known about the protein. The ways in which mechanistic and structural information is used to direct the redesign of enzyme specificity is discussed. This database of information is providing a foundation for establishing rules that govern enzyme-substrate interactions. (This chapter was published in Harris, J.L. and Craik, C.S. (1998) Engineering enzyme specificity. *Current Opinion in Chemical Biology* **2**, 127-132.)

The ultimate goal in protein engineering is the creation of novel enzymatic functions derived from first principles. While the *de novo* design of protein structure has shown some recent advances (Betz *et al.*, 1997; Dalal *et al.*, 1997; Severin *et al.*, 1997), the knowledge to consistently design enzymatic function is still lacking. As a result, much effort has been put into understanding and redesigning properties of preexisting proteins (Chang *et al.*, 1994; Dufour *et al.*, 1995; el Hawrani *et al.*, 1996; Huang *et al.*, 1996; Kast *et al.*, 1996; Lawson *et al.*, 1997; Pinto *et al.*, 1997; Quaemaeneur *et al.*, 1998). Protein engineering has been particularly insightful in the elucidation of structure-activity relationships by modification of substrate specificity (Arnold and Moore, 1997; Cantu *et al.*, 1996; Cubitt *et al.*, 1995; Heim and Tsien, 1996; Park *et al.*, 1997). Substrate specificity is the preference that an enzyme manifests for one substrate over competing substrates. Implicit in this definition is the understanding that specificity is dependent both on substrate binding and on the utilization of the binding energy for catalytic turnover. Binding of a substrate is necessary, but not sufficient, for catalysis. Indeed, substrates that bind in modes that are not conducive to catalysis are defined as inhibitors. This review addresses some of the current uses of protein design and subsequent engineering to better understand the factors involved in enzyme substrate specificity.

Rational design defines one end of the spectrum of approaches used to engineer enzyme substrate specificity. In this case, a detailed understanding of the catalytic mechanism and determinants of substrate specificity of a particular enzyme are used as a basis to alter its specificity in a predictable fashion. At the other end of this design spectrum is the random, or “irrational”, approach. Region-specific random mutagenesis has proven to be useful in modulating substrate specificity in cases where a limited amount of information is known about the target enzyme (el Hawrani *et al.*, 1996). Completely random mutagenesis does not require prior information of the catalytic mechanism and substrate specificity of the target enzyme. However, success of the random approach depends on the ability to search large libraries of variant proteins to identify the desired function (Zhao and Arnold, 1997a). The size of the library that must be sampled is often inversely proportional to the database of information known about the enzyme. Whatever the method chosen, analysis of the resulting protein is important for generating additional information about the enzyme, refining experimental design, and contributing to knowledge of general design principles (Figure 2-1).

An elegant example of how a structure based method was used to design a novel substrate specificity was presented by Shokat and coworkers (Shah *et al.*, 1997). They chose to tackle the problem of identifying the direct substrates of a particular kinase, which can be difficult because of the redundancy and overlapping substrate consensus sequences among the plethora of kinases. Their proposed solution to this problem was to create a unique nucleotide binding site in a kinase that can bind and efficiently catalyze the phosphotransfer of an ATP analog not utilized by other kinases, while maintaining the original peptide substrate recognition properties. Analysis of the three-dimensional structures of the cAMP-dependent kinase and the cyclin-dependent kinase-2 identified two key amino acids in the purine nucleotide binding site, valine 323 and isoleucine 338. Substitution of these bulky side chains with alanines created a pocket near the N^6-amino position of bound ATP. Creation of this cavity in the variant enzyme allowed the kinase to bind and utilize ATP analogs derivatized at the N^6-position (Figure 2-2). Introduction of these mutations into the catalytic domain of the prototypical tyrosine kinase, v-Src, permitted the direct substrates of v-Src to be labeled by $[\gamma\text{-}^{32}P]N^6$-(cyclopentyl) ATP. The $[\gamma\text{-}^{32}P]N^6$-(cyclopentyl) ATP analog was only recognized by the variant v-Src and not other cellular kinases; this allowed the authors to show that the main phosphorylation target of v-Src was itself. This approach appears to be generally applicable and is currently being extended to other members of the protein tyrosine kinases (Liu *et al.*, 1998).

The wealth of information on metal binding sites in proteins can extend the possibilities of protein design to include metals as structural components and cofactors for substrate specificity and catalysis (Bonagura *et al.*, 1996; Braha *et al.*, 1997; Klemba and Regan, 1995). Based on this extensive protein database of preferred geometry, distance, and ligands for divalent transition metal binding sites, a tridentate metal binding site was rationally designed into trypsin to confer metal-assisted substrate specificity (Willett *et al.*, 1996). The metal binding site was designed to bind metal in a tetrahedral conformation and consisted of two histidine residues engineered into the S2' subsite of trypsin and a third histidine residue provided by the P2' position of the substrate (Figure 2-3). A substrate containing a P2'-histidine in the presence of metal allowed the modified trypsin to hydrolyze the peptide bond C-terminal to the α-carbon of tyrosine, a usually poor substrate for trypsin. As designed, a metal capable of binding in a tetrahedral conformation was required for cleavage. Zinc (II), which prefers tetrahedral geometry, showed increased catalysis over nickel (II), which prefers square planar geometry but can adopt a tetrahedral coordination. In the presence of copper (II), which strongly prefers square planar geometry, cleavage of the substrate was not observed. X-ray crystallographic structure analysis of the metal-trypsin complexes validates these preferred geometries (Brinen *et al.*, 1996). Engineering a metal binding site, to both increase the occupancy and tune the orientation of a non-preferred substrate, underscores the important role that binding substrate in the correct position for catalysis has in determining substrate specificity.

Another approach along the spectrum of design methods is to use examples found in nature as a guide to alter substrate specificity. Comparative analysis of homologous proteins that differ in substrate specificity can identify key functional determinants and direct the redesign in a family of enzymes. A good example of this method came from analysis of the subtilisin class of serine proteases. The subtilisin family includes the bacterial subtilisin BPN', a very efficient but promiscuous enzyme, the yeast processing enzyme kex2, which has dibasic P2-P1-specificity, and the mammalian prohormone endopeptidase furin, which has tribasic P4-P2-P1-specificity. Using the primary sequences of this class of enzymes and the three-dimensional structure of subtilisin, Ballinger *et al.* created a two amino acid variant of subtilisin BPN' that displayed kex2-like substrate specificity (Ballinger *et al.*, 1995). Subsequently, the specificity of this variant was extended to that of furin by engineering P4-basic specificity through introduction of a third mutation (Ballinger *et al.*, 1996). An unexpected result of this series of variants is the synergy that exists between the engineered sites. In fact, since the P1-arginine preference of the triple mutant was significantly more stringent than the double mutant, the autocatalytic cleavage site had to be changed to furin's preferred cleavage sequence for efficient processing and activation of the protease.

Another system where comparative analysis of homologous enzymes has been used to direct substrate specificity redesign is in the case of the decarboxylating dehydrogenases: isocitrate dehydrogenase (IDH) and isopropylmalate dehydrogenase (IMDH). These enzymes catalyze the Mg^{2+} and nicotinamide dinucleotide cofactor-dependent oxidation and decarboxylation of 2-hydroxy acids. A prominent difference between IDH and IMDH is their preference for nicotinamide dinucleotide. IDH shows a 7000-fold preference for NADP over NAD while IMDH shows a 200-fold preference for NAD over NADP. Using comparative analysis, Dean and coworkers (Chen *et al.*, 1996) identified key amino acid residues that directly contacted the cofactor. Additional "second shell" substitutions were designed based on the consideration of packing constraints that could correctly position the "first shell" determinants of cofactor binding. The modified IDH with seven amino acid substitutions displayed an 850-fold preference for NAD. The modified IMDH with four amino acids substitutions and replacement of a 7-amino acid loop for one with 13-amino acids resulted in a 1000-fold preference for NADP. The structure of the engineered NAD-dependent IDH was recently solved and verified the importance of distal interactions in determining substrate specificity (Hurley *et al.*, 1996). The identification by mutagenesis of substrate specificity determinants remote from the substrate binding site has been seen in hydrolytic enzymes as well (Mace *et al.*, 1995; Perona *et al.*, 1995).

Applying design and engineering principles to macromolecular human therapeutics parallels the approach a medicinal chemist uses to improve the properties of a small molecule that exhibits promising efficacy in a drug screen. For example, the anticoagulation property of the serine protease thrombin was explored and improved by region specific mutagenesis (Tsiang *et al.*, 1996). Thrombin plays a pivotal role in controlling both clot formation as well as clot dissolution at sites of vascular injury. This dual function for thrombin is dependent on the regulation of its substrate specificity. Specifically, cleavage of fibrinogen promotes coagulation, while cleavage of protein C in the presence of thrombomodulin activates an anticoagulant pathway. Alanine scanning mutagenesis of thrombin was used to identify amino acid residues that decrease fibrinogen specificity while maintaining protein C activation (Gibbs *et al.*, 1995). Subsequently, these positions were mutated to all 19 other amino acids and screened for preferential protein C activation over fibrinogen cleavage. A single point mutation, E229K, was found to shift the substrate specificity of thrombin 130-fold in favor of protein C activation over fibrinogen cleavage. These *in vitro* results were conserved when tested *in vivo* in Cynomolgus monkeys. The enhanced anticoagulant activity of the thrombin variant could prove useful in treating the consequences of thrombosis, such as stroke or myocardial infarction.

A similar therapeutic objective in redesigning preexisting enzymatic substrate specificity is the creation of "designer enzymes" to enhance the activity of small molecule drugs (Black *et al.*, 1996; Wentworth *et al.*, 1996). A combination of random mutagenesis, genetic selection and drug-sensitivity screens was used to generate herpes simplex virus type 1 thymidine kinase (HSV-1 TK) variants that demonstrate substrate specificity towards the chain-terminating nucleoside analog pro-drugs ganciclovir and acyclovir (Christians *et al.*, 1997). Preferential phosphorylation of these pro-drugs would increase their chemotherapeutic value in treating particular cancers by enhancing incorporation of the drug into DNA. Based on the sequence conservation among the thymidine kinases and previous mutagenesis studies, six amino acids were targeted for saturation mutagenesis. These sites were changed to all other possible amino acids and the resulting library of randomized HSV-1 TKs was sequentially selected for thymidine kinase activity and screened for sensitivity to ganciclovir and acyclovir. A variant, with four amino acid substitutions in the kinase active site, resulted from the screen that was 43-fold more sensitive to ganciclovir and 20-fold more sensitive to acyclovir. This approach can provide a potentially useful tool for gene therapy applications.

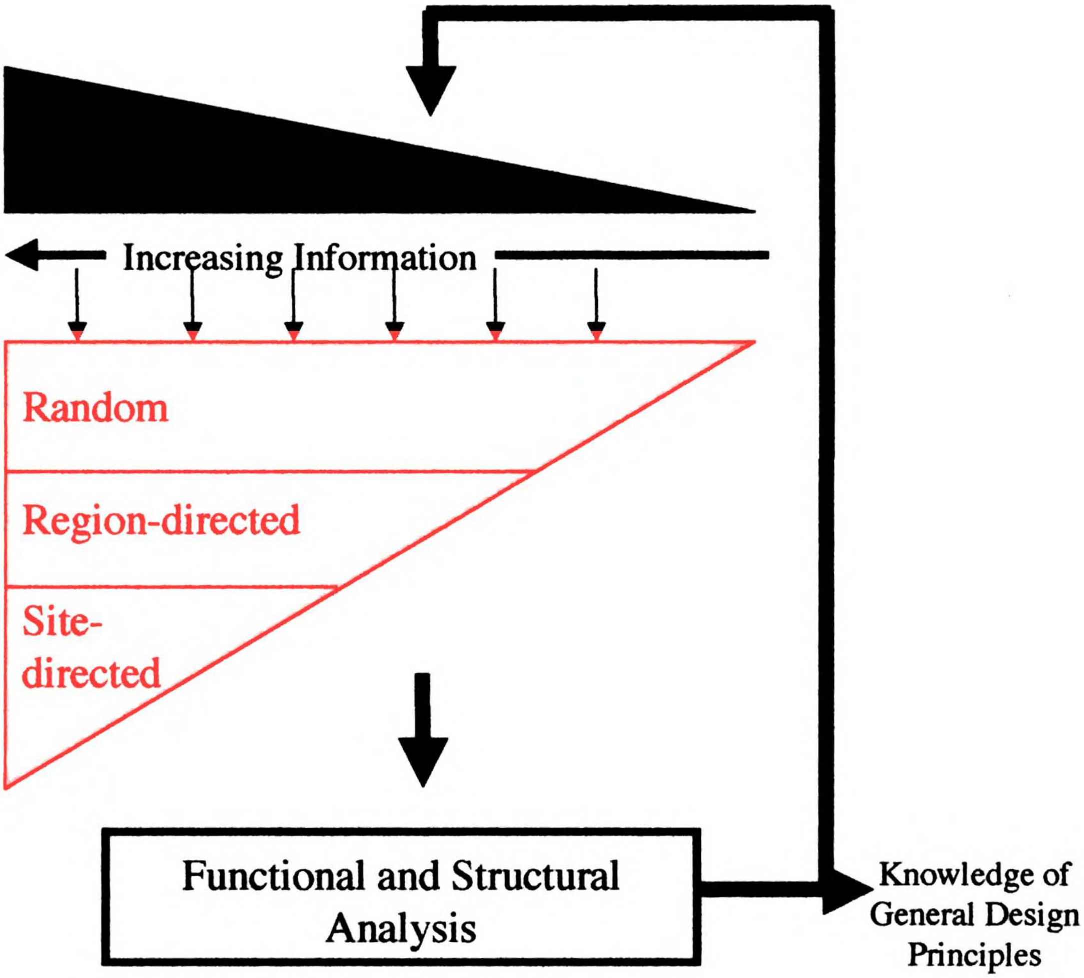
The rational design of altered substrate specificity through directed site-specific mutagenesis or region-specific random mutagenesis requires some initial knowledge of amino acid residues or regions in a protein critical for substrate specificity. Because this information is not available for most enzymes, coupled with the increasing appreciation of distal specificity determinants and synergy between binding sites, considerable effort has been directed toward generating altered substrate specificity through the process of directed evolution, or the accumulation of effectual mutations over generations of randomization and selection (Crameri *et al.*, 1996; Moore and Arnold, 1996; You and Arnold, 1996; Zhao and Arnold, 1997b; Zhao *et al.*, 1996). One promising method for generating infrequent, random, point mutations is DNA shuffling, which is also referred to as *in vitro* evolution. The process of DNA shuffling involves reiterative mutation and recombination by non-specific DNA fragmentation followed by reassembly and extension by primerless PCR (Smith, 1994; Stemmer, 1994b) (Figure 2-4). One particularly powerful application of this method is to determine the resistance outcome of a drug candidate *in vitro* to guide secondary drug generation before resistance arises *in vivo*. This technique was used with TEM-1 β lactamase and the antibiotic cefotaxime (Stemmer, 1994a). *In vitro* evolution of β-lactamase resulted in an enzyme that could hydrolyze cefotaxime 32,000-fold more efficiently than the parent enzyme. Presumably this type of approach can be used to look at the resistance profile of cefotaxime derivatives to obtain antibiotics that are less prone to resistance.

Recently, Stemmer and his colleagues used DNA shuffling and screening (rather than selection) to direct β-D-fucosyl hydrolysis activity onto β-galactosidase (Zhang *et al.*, 1997). After seven rounds of DNA shuffling, an evolved β-galactosidase was isolated that showed a 10-fold improvement in fucosidase activity accompanied by a 40-fold decrease in galactosidase activity. The DNA sequence of this molecularly evolved fucosidase showed thirteen mutations resulting in six amino acid changes in the translated β-galactosidase gene. Interestingly, only three mutations appear near the active site, the other mutations are distal to the binding site and may either play a role as long-range specificity determinants or be an important factor in the *in vivo* selection, such as increased stability or preferred codon usage, that is not assayed *in vitro.* The use of molecular evolution methods to select for several complex traits *in vivo* is exemplified in a recent report of engineering multiple genes in an arsenate detoxification pathway (Crameri *et al.*, 1997).

In this limited review, we have attempted to introduce some of the uses of protein engineering in elucidating substrate specificity determinants. Methods discussed in this review include the use of structure-directed point mutations, metal binding sites, comparative analysis with subsequent directed mutagenesis, region specific mutagenesis, random mutagenesis, and *in vitro* evolution. In combination, these methods, among others (Hanes and Pleuckthun, 1997; O'Neil and Hoess, 1995; Wade and Scanlan, 1997), provide an impressive arsenal of tools for creating enzymes with complex and novel functions. The choice of the design and engineering method will depend on the type of information known about a protein and its function. Whatever the approach, the challenge now becomes the thorough analysis of the altered function to increase our knowledge of protein structure and function.

Figure 2-1.

Schematic diagram of approaches used to design and engineer altered function into enzymes. Greater knowledge about an enzyme will provide more options for redesign. Thorough analysis of the resulting variants allows both positive feedback for future design and potentially generalizable design principles.

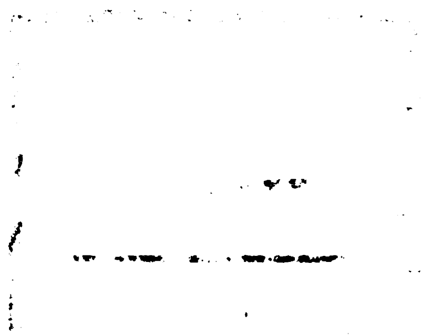

Increasing Information
Random
Region-directed
Site-directed
Functional and Structural Analysis
Knowledge of General Design Principles

Figur

Panel

mutag

show



A.

Figure 2-2.

Panel A shows the structure of ATP with the v-Src kinase amino acids targeted for mutagenesis, V323 and I338, shown in bold line. The N^6-(cyclopentyl) ATP analog is shown in panel B in the context of the variant v-Src, [V323A, I338A].

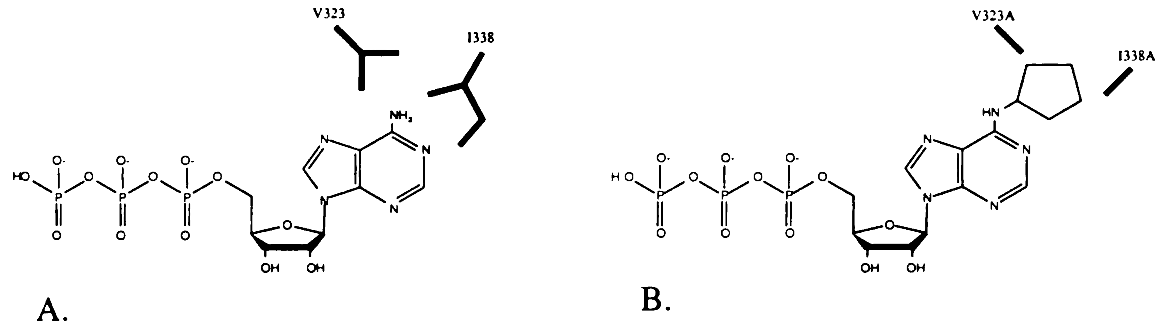

V323
I338
NH₂
N
HO
O-
P
O
OH
A.
V323A
I338A
HN
B.

Figure 2-3.

A divalent metal coordinates the two engineered histidine residues in trypsin (His 143 and His 151) and the P2'-histidine residue of the substrate. This metal dependent substrate specificity allows trypsin to correctly position the substrate relative to the catalytic residues (Asp102, His57, Ser195) for cleavage after the normally refractive P1-tyrosine.

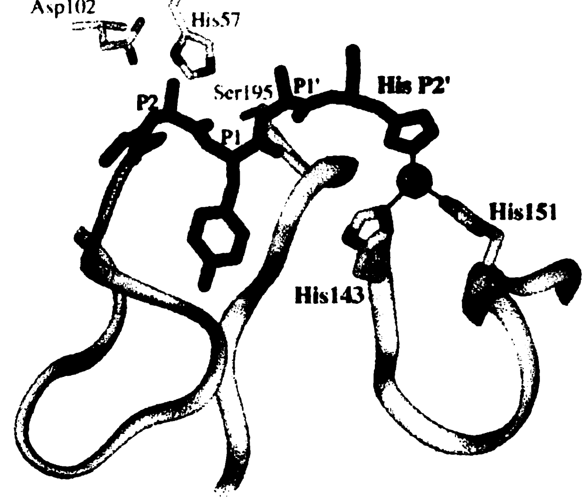
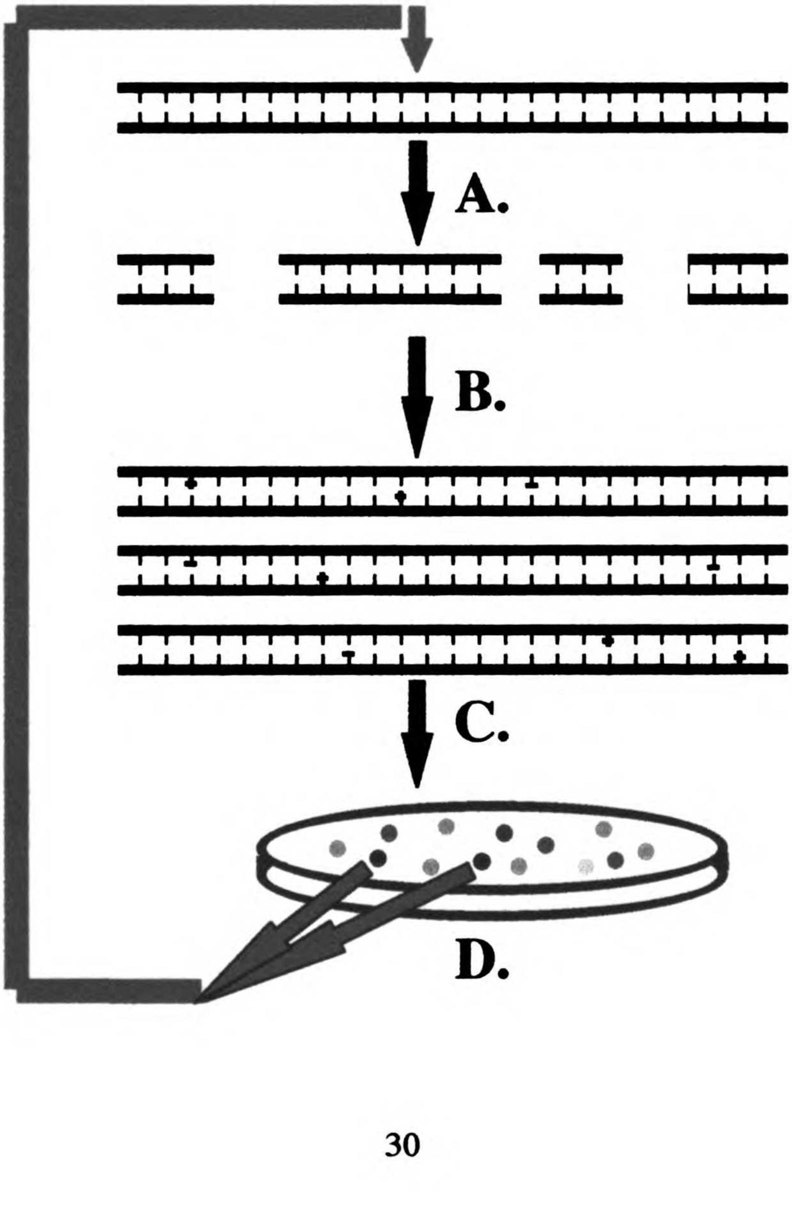

Asp102
His57
Ser195
P2
P1
P1'
His P2'
His151
His143

Figure 2-4.

DNA Shuffling. (a) Cleavage of the β-galactosidase gene by DNase 1 into 50-200 bp fragments. (b) Creation of the β-galactosidase library through two PCR steps. "Primerless" mutagenic PCR was used to introduce (+) and negative (-) point mutations and subsequent amplification of the reassembled gene. (c) Visual colorimetric screening of the resultant library for fucosidase activity on 5-bromo-4-chloro-3indol β-fucopyranoside plates. (d) Clones displaying positive activity are used as a templet for the next round of DNA shuffling.



Chapter Three:

Synthesis of Positional-Scanning Libraries of Fluorogenic Peptide Substrates that Incorporate Diverse P1 Substituents: Extended Specificity Determination of Plasmin and Thrombin

Abstract

A positional-scanning synthetic combinatorial library (PS-SCL) synthesis strategy is presented in which diversity at P1 has been achieved. The strategy is demonstrated through the synthesis of a tetrapeptide positional library in which the P1-amino acid is held constant as a lysine and the P4-P3-P2 positions are positionally randomized. The 6,859 members of the library are synthesized on solid support with an alkanesulfonamide linker. The members of the peptide library are displaced from the solid support by condensation with a fluorogenic 7-amino-4-methylcoumarin-derivatized lysine. This library was used to determine the extended substrate specificities of two trypsin-like enzymes involved in the blood coagulation pathway, plasmin and thrombin. The optimal P4 to P2 substrate specificity for plasmin was P4-Lys/Nle/Val/Ile/Phe, P3-Xaa, and P2-Tyr/Phe/Trp. This cleavage sequence has recently been identified in some of plasmin's physiological substrates. The optimal P4 to P2 extended substrate sequence determined for thrombin was P4-Nle/Leu/Ile/Phe/Val, P3-Xaa, and P2-Pro, a sequence found in many of the physiological substrates of thrombin. Single substrate kinetic analysis of plasmin and thrombin was used to validate the substrate preferences resulting from the PS-SCL. Three-dimensional structural modeling of the substrates into the active sites of plasmin and thrombin identified potential determinants of the defined substrate specificity. (This work has been submitted for publication by Backes, B.J. (co-first author), Harris, J.L. (co-first author), Leonetti, F., Craik, C.S. and Ellman, J.A. Synthesis of Positional-Scanning Libraries of Fluorogenic Peptide Substrates that Incorporate Diverse P1 Substituents: Extended Specificity Determination of Plasmin and Thrombin.)

Abbreviations Used

AMC, 7-amino-4-methyl coumarin; Nle and Z, norleucine; Xaa, a non-specific amino acid; PS-SCL, positional scanning-synthetic combinatorial library; MUGB, 4-methylumbelliferyl *p*-guanidinobenzoate; Tris, *tris*-(hydroxymethyl)-amino-methane; BSA, bovine serum albumin; DICI, diisopropylcarbodiimide; HOBt, 1-hydroxybenzotriazole; TFA, trifluoroacetic acid; Fmoc, 9-fluorenylmethoxycarbonyl; pbf, 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl; Trt, trityl; Boc, *tert*-butoxycarbonyl; *t*-Bu, *tert*-butyl DMF, dimethylformamide; NMP, N-methylpyrrolidine.

Introduction

Proteases play essential roles in numerous biological processes. Substrate specificity, or the ability to discriminate among many potential substrates, is central to the function of proteases. Knowledge of a protease's substrate specificity may not only give valuable insights into its biological function and provide the basis for potent substrate and inhibitor design. Synthetic substrates are typically employed to define substrate specificity. However, the synthesis and assay of single substrates is tedious and impractical for proteases with specificity beyond P1 and often results in a limited substrate specificity profile.

Combinatorial approaches have recently been used to address the identification of multiple substrate recognition sites in proteases. All of these combinatorial methods involve the generation of libraries of substrates, proteolysis of the substrates, and identification of the optimal substrate sequence. Substrate libraries can be broken down into two categories: those that are biologically generated and those that are synthetically generated. Biological library methods include the display of peptide libraries on filamentous phage (Ding *et al.*, 1995; Matthews and Wells, 1993), the randomization of amino acids at physiological cleavage sites (Bevan *et al.*, 1998), and the identification of macromolecular cleavage sites in *in vitro* transcription/translation cDNA libraries (Kothakota *et al.*, 1997; Lustig *et al.*, 1997). The diversity of these libraries is often constrained to the transformation efficiency of the host organism and can only contain naturally occurring amino acids. Synthetic substrate libraries circumvent the dependence on transformation efficiency and naturally occurring amino acids. While combinatorial synthesis allows for the creation of millions of compounds, these methods are only useful when coupled with powerful analytical assays that allow for the identification of the preferred substrate. Discontinuous analysis of the cleavage products through Edman degradation (Birkett *et al.*, 1991; Petithory *et al.*, 1991), mass spectroscopy [Berman, 1992 #37; (Berman *et al.*, 1992; McGeehan *et al.*, 1994), and chromatography

(Schellenberger *et al.*, 1993) has proven useful for qualitative assessment of optimal substrates from soluble or support-bound peptides.

Substrate consensus sequences have also been obtained using support-bound fluorescence quenched substrate libraries prepared by the process of split-synthesis, which results in single substrate sequences on each of the resin beads (Lam and Lebl, 1998). Partial proteolysis of the support-bound libraries and subsequent sequence determination of the substrates on the most fluorescent beads provides the consensus sequences (Meldal, 1998; Meldal *et al.*, 1994). Unfortunately, this method suffers from two major limitations: first, the kinetics of support-bound substrates can differ greatly from soluble substrates (Del Nery *et al.*, 1997; St. Hilaire *et al.*, 1999) and second, like the other methods previously mentioned, identification of the substrate occurs after the cleavage event, making the kinetic analysis more cumbersome. A method that avoids these limitations, and gives a quantitative assessment of protease substrate preference, is the use of positional scanning-synthetic combinatorial libraries (PS-SCL) (Dooley and Houghten, 1998).

Positional-scanning (Pinilla *et al.*, 1992) synthetic combinatorial libraries (PS-SCL) of fluorogenic peptide substrates have the potential to be a very powerful tool for determining protease specificity. In contrast to other combinatorial libraries, this library format provides rapid and continuous information on each of the varied substituents in the substrate. A positional-scanning library with the general structure Ac-X-X-X-Asp-AMC was prepared previously by Rano et al. to rapidly and accurately assess the P4-P2 specificity for caspases that require Asp in the P1 position (Rano *et al.*, 1997; Thornberry *et al.*, 1997). Specific cleavage of the amide bond after the Asp residue liberates a fluorescent 7-amino-4-methylcoumarin (AMC) leaving group, thus allowing for the simple determination of cleavage rates for a library of substrates. The P1 Asp-coumarin substrate was convieniently linked to an insoluble polymer through the Asp carboxylic acid side chain, which allowed for library synthesis by standard peptide synthesis. However, the method used to synthesize the library was specific for an aspartic acid at P1. The employment of strategies to link P1 amino acid-coumarin derivatives through side chain functionality may prove viable for some residues. However, linkage through hydrophobic side chain functionalities (Leu, Phe, Val, etc.) will prove difficult, therefore, the utility of this strategy is limited. By developing a general strategy to incorporate all 20 proteinogenic amino acids at the P1 position of a PS-SCL, the extended specificity of virtually any protease could be rapidly determined.

The present study describes the design and development of a general method for the preparation and screening of positional scanning-synthetic combinatorial substrate libraries. This design is free from the limitations of previous approaches and allows for complete randomization at the P1-position. A library generated by this method has been applied to map the P4, P3, P2 extended substrate specificity of plasmin and thrombin, serine proteases involved in, among other processes, blood coagulation and fibrinolysis. The resulting enzymatic profiles from the library resemble the known physiological cleavage sites of these enzymes and were verified by single substrate kinetic analysis. Potential substrate recognition determinants on the enzymes were identified through the 3-D modeling of the substrates in the active sites of the enzymes.

Results

Library Design and Synthesis. To incorporate diversity at the P1-position, we condensed fluorogenic AMC P1-amino acid derivatives with a support-bound PS-SCL to provide library compounds in solution (Figure 3-1). Three support-bound sub-libraries were prepared (P2, P3, P4) employing an alkanesulfonamide linker (Backes and Ellman, 1999) and solid-phase peptide synthesis. The properties of the linker allow for the incorporation of a fluorogenic-leaving group through the nucleophilic addition of a AMC-derivatized amino acid. Each sub-library consisted of 19 resins (one unnatural amino acid, norleucine, was included; cysteine and methionine were excluded) for which a single position was spatially addressed by the coupling of a single amino acid. The two remaining positions of each resin were supplied by the coupling of isokinetic mixtures (Ostresh *et al.*, 1994) to give a resin-bound mixture of 361 amino acids. The 57 resins comprising the entire PS-SCL were put in individual wells and cleaved from the resin with a P1-amino acid-coumarin derivative. Filtration, side-chain-deprotection, and concentration provided a PS-SCL of 57 wells containing 361 tetrapeptide-coumarin derivatives (total-6859 peptides). Thus analysis of the three libraries identifies the enzyme's preferences for amino acids at P4, P3, and P2. A P1-Lys library was prepared and used to elucidate the specificity of plasmin and thrombin.

Profiling of Plasmin with the Positional Scanning P1-Lys Library: The preferred tetrapeptide substrate recognition sequence for plasmin was determined to be P4-Lys, P3-Xaa, P2-Tyr/Phe/Trp, and P1-Lys (Figure 2A). To validate the results from the PS-SCL and to quantitate dependence and utilization of extended interactions, kinetic parameters were determined for several single AMC substrates (Table 3-1). As indicated from the library, the majority of plasmin's extended substrate specificity resides in P2 and P4. Hydrolysis of the sub-optimal P2 substrate, Ac-Lys-Thr-**Ser**-Lys-AMC, is up to 34-fold disfavored when compared to substrates that possess an aromatic amino acid at P2, Ac-Lys-Thr-**Tyr**-Lys-AMC, Ac-Lys-Thr-**Phe**-Lys-AMC, and Ac-Lys-Thr-**Trp**-Lys-AMC (Table I), 0.020 $\mu M^{-1}\ s^{-1}$ in k_{cat}/K_m versus 0.544 $\mu M^{-1}\ s^{-1}$, 0.677 $\mu M^{-1}\ s^{-1}$, and 0.601 $\mu M^{-1}\ s^{-1}$ respectively. The subtle preference for lysine over phenylalanine at P4 is also verified by single substrates, Ac-**Phe**-Thr-Tyr-Lys-AMC retains 63% of the activity of Ac-**Lys**-Thr-Tyr-Lys-AMC, 0.342 $\mu M^{-1}\ s^{-1}$ in k_{cat}/K_m and 0.544 $\mu M^{-1}\ s^{-1}$ in k_{cat}/K_m respectively.

Structural Determinants of P4-Lysine and P2-Aromatic Substrate Specificity in Plasmin: The structure of plasmin was solved in the absence of a substrate or inhibitor in the active site (Wang *et al.*, 1998). Because of this, analysis of enzyme-substrate interactions required the molecular modeling of the optimal substrate, Lys-Thr-Phe-Lys, into the active site. The resulting model reveals potential structural determinants for substrate recognition. As is appreciated for trypsin, the major determinant for P1-basic specificity lies in Asp-189, a residue at the base of the S1-pocket (Figure 3A). The (δ+) ring hydrogens from P2-Phe can interact with the carboxylate group of Glu-60a that protrudes from above the active site and toward to P2-Phe ring edge. The positively charged (δ+) amino group of Gln-192 could then make contact with the (δ-) π-electrons of the P2-Phe face (Figure 3-3A). Significant interactions between plasmin and the P3-amino acid side chain are not readily apparent from the structural model. This is a result of the P3-amino acid side chain being directed away from the enzyme and into bulk solvent. A substrate with a P4-Lys could make contact with Glu-180, with the aliphatic portion of P4-lys packing against Trp-215 (Figure 3-3A). Position 180 is normally occupied by a hydrophobic amino acid in other chymotrypsin-like serine proteases. The additional, although lesser, preference for P4-Nle/Val/Ile/Phe could in part be due to favorable interaction with Trp-215.

Profiling of Thrombin with the Positional Scanning P1-Lys Library: Profiling of thrombin with the PS-SCL revealed that the preferred P4-P2 extended substrate specificity is for large aliphatic amino acids at P4, such as norleucine, leucine and isoleucine, negligible discrimination at P3, and narrow specificity for proline at P2 (Figure 3-2B). These preferences were validated through single substrate kinetic analysis (Table 3-1). Replacement of the optimal P2 amino acid proline with the sub-optimal amino acid leucine results in a 45-fold decrease in activity, 3.83 μM^{-1} s^{-1} in k_{cat}/K_m for Ac-Nle-Thr-**Pro**-Arg-AMC versus 0.085 μM^{-1} s^{-1} in k_{cat}/K_m for Ac-Nle-Thr-**Leu**-Arg-AMC. The requirement for aliphatic amino acids at P4 has a less pronounced effect than the requirement for proline at P2, as reflected by the relative activities in the PS-SCL. However, upon replacement of the preferred leucine at P4 for the sub-optimal glycine, there is a 23-fold decrease in specific activity, 0.154 μM^{-1} s^{-1} in k_{cat}/K_m for Ac-**Leu**-Gly-Val-Arg-AMC versus 0.007 μM^{-1} s^{-1} in k_{cat}/K_m for Ac-**Gly**-Gly-Val-Arg-AMC.

Structural Determinants of P4-Aliphatic and P2-Proline Substrate Specificity in Thrombin: The coordinates used for enzyme-substrate analysis of thrombin were that of thrombin complexed with the tripeptide inhibitor, D-Phe-Pro-Arg-chloromethylketone (Bode *et al.*, 1989). The P3 side chain was converted to a Thr of the L-enantiomer and a P4-Val was added to the amino-terminus of the inhibitor. As suggested from the original structural analysis by Bode et al., the preference for P1-basic amino acids is determined by Asp-189 and the preference for proline arises from the insertion, relative to the digestive serine proteases, of seven amino acids in the 60's loop (Bode *et al.*, 1989). This loop is above the active site Ser-195 and creates a rigid pocket for P2-Pro interaction (Figure 3-3B). The specificity for P3 amino acids is less well understood from the original structure. The use of P3-D-Phe allows the side chain to point into the enzyme, occupying, in part, the S4 pocket. When the P3-D-Phe is replaced with P3-L-Thr, the side chain is pointed away from the enzyme with significantly fewer interactions (Figure 3-3B). The S4 pocket on thrombin is very clearly hydrophobic, with Ile-174 making significant interactions with P4-Val modeled into the pocket. Additional hydrophobic determinants for this pocket include Trp-215 at the floor of the pocket, Met-180 at the end of pocket and Leu-99 near the top of the pocket (Figure 3-3B).

Discussion

The rapid discovery of new proteases presents the need for generalized assays to aid in the elucidation of their biological functions. The use of synthetic positional scanning combinatorial libraries offers the ability to rapidly test and evaluate the extended substrate specificity of a protease. The major limitation of previous synthetic methods is the inability to permit complete diversity at the P1-position. The development of the synthetic strategy presented in this manuscript surmounts this limitation and allows for complete diversity of any naturally occurring amino acid at the P1 position. In fact, a variety of nucleophiles, including unnatural amino acid derivatives, could be incorporated in the cleavage step.

The utility of a positional scanning library with P1-lysine was demonstrated through the determination of the extended specificity of plasmin and thrombin, proteases involved in the regulation of hemostasis. Plasmin has been traditionally characterized as a protease with broad substrate specificity. Results from the current study show this not to be the case. Plasmin demonstrates a distinct preference for aromatic amino acids at P2 and a moderate preference for lysine and hydrophobic amino acids at P4. Molecular modeling of a substrate bound into the active site of plasmin can aid in the identification of potential structural interactions between enzyme and substrate. Plasmin has an insertion in the 60's loop, relative the digestive protease trypsin, that could allow for the creation of a S2 pocket. Plasmin's S2 pocket is not simply a hydrophobic pocket, but is specific for aromatic amino acids. Experimental crystallographic evidence exists for structural determinants used to achieve discrimination between aromatic and aliphatic amino acids, including aromatic ring interactions with oxygen (Thomas *et al.*, 1982) and amide nitrogens (Burley and Petsko, 1986). Likewise, modeling of a P2-Phe into the putative S2 pocket shows that Glu-60a may contribute to ring hydrogen bonding, while the polar amino group of Gln-192 may interact with the π-electrons of P2-Phe (Figure 3-3A). The P3-amino acid side chain is pointed out into solvent and makes few interactions with the enzyme. The moderate P4 preference for lysine may be driven by an electrostatic interaction with Glu-180 (Figure 3-3A). The aliphatic portion of the Lys side chain as well as other aliphatic amino acids (Nle/Val/Ile) could form favorable interactions with Trp-215.

Plasmin is an important enzyme in attenuating blood coagulation and restoring blood flow through its degradation of fibrin and pro-coagulant factors (McKee *et al.*, 1975; Omar and Mann, 1987). Several non-fibrin substrates of plasmin have recently been demonstrated (Campbell and Andress, 1997; Gundersen *et al.*, 1997; Tsirka *et al.*, 1997). While the cleavage sites of the physiological substrates have not all been determined, many of those that have been identified resemble the optimal site determined in the current study, especially in the conservation of a P2-aromatic amino acid (Figure 3-4). It has been proposed that plasmin may function in accelerating fibrinolysis and arresting coagulation through the cleavage of the normally pro-coagulant factor X into an anti-coagulant cofactor (Pryzdial *et al.*, 1999). It was determined that plasmin cleaves factor X after the Arg in the site Ile-Thr-Phe-Arg. In a similar vein, plasmin may act in the desensitization of the protease activated receptor PAR1. PAR1 is a transmembrane receptor that when cleaved by thrombin results in the activation of platelets; however plasmin cleaves PAR1 at sites that not only do not result in activation but also remove the thrombin activation site (Kuliopulos *et al.*, 1999). One of the plasmin cleavage sites in PAR1 was determined as Thr-Glu-Tyr-Arg. Plasmin may attenulate its own production through the cleavage of vitronectin, a protein that binds both plasminogen activator inhibitor-1 (PAI-1) and extracellular matrix (ECM) (Chain *et al.*, 1991; Kost *et al.*, 1996). The cleavage of vitronectin by plasmin may result in the release of PAI-1 from the ECM to inhibit plasminogen activators and subsequent inhibition of plasmin formation. The cleavage site of vitronectin is Lys-Gly-Tyr-Arg, a site resembling the optimal site determined in this study. Plasmin may also play a role in regulating bone resorption through the cleavage of osteocalcin at the site Glu-Ala-Tyr-Arg (Novak *et al.*, 1997).

Thrombin has been shown to be more restrictive in its substrate cleavage profile than trypsin, a protease whose main function is digestive rather than regulatory (Bode *et al.*, 1992; Pozsgay *et al.*, 1981). This is due, in part, to preferences exhibited in the S3-S2' subsites (Le Bonniec *et al.*, 1996; Vindigni *et al.*, 1997). Here, for the first time, the sequence space for thrombin's P4, P3 and P2 substrate preference has been completely sampled. Like previous results (Kawabata *et al.*, 1988; Lottenberg *et al.*, 1983; Pozsgay *et al.*, 1981; Vindigni *et al.*, 1997) the PS-SCL shows that thrombin has a pronounced preference for proline at the P2 position of the substrate. The structure of thrombin bound to the tripeptide inhibitor, D-Phe-Pro-Arg-chloromethylketone, reveals that the constraints for proline in the P2 position most likely arise from the insertion in the 60's loop of thrombin (Bode *et al.*, 1989). When a P3-L-amino acid was modeled into the active site, rather than the P3-D-Phe solved in the structure, the side chain had minimal interaction with the enzyme and was pointed out into solvent (Figure 3-3B). This structural analysis of the P3-S3 interaction supports the lack of defined P3 specificity seen in the substrate library. However, a P4-aliphatic amino acid modeled into the active site can make significant interactions with a hydrophobic pocket of thrombin. The walls of this pocket are formed predominantly by Ile-174, other amino acids such as Met-180, Leu-99, and Trp-215 also contribute to this hydrophobic environment (Figure 3-3B). The interaction between Ile-174 and a P4-aliphatic amino acid was also observed in the structure of thrombin bound to the peptide fragment Leu-Asp-Pro-Arg (Mathews *et al.*, 1994). The structural analysis was supported by the substrate library analysis that showed thrombin to have a definite preference for P4-aliphatic amino acids.

The results from the library cleavage analysis, structural analysis and kinetic analysis show a unique specificity fingerprint of thrombin for aliphatic amino acids at P4 and for proline at P2. The importance of thrombin's specificity profile for the recognition of proper substrates in coagulation and vascular integrity is appreciated in the context of its physiological substrate cleavage sites (Figure 3-5). Thrombin plays an important role in blood coagulation by activating platelets through the cleavage of PAR1 after arginine in the sequence, Leu-Asp-Pro-Arg (Vu *et al.*, 1991). Thrombin further enhances coagulation through the feedback cleavage activation of factor V and factor VIII at the sequences Leu-Ser-Pro-Arg (1018) and Trp-Tyr-Leu-Arg (1545) in factor V and Ile-Gln-Ile-Arg(372) and Gln-Ser-Pro-Arg(1689) in factor VIII (Keller *et al.*, 1995; Pittman *et al.*, 1994). Ultimately, thrombin assists in the formation of the fibrin clot through the cleavage of fibrinogen. The sequence that thrombin cleaves in fibrinogen Aα-chain, Gly-Gly-Val-Arg, is suboptimal by the current analysis. However, it has been demonstrated that the fibrinogen Aα-chain supplements its specific binding energy through the use of thrombin's distal anion-binding exosite 1 (Stubbs and Bode, 1993). Fibrinogen Bβ-chain has a more optimal thrombin cleavage site, Phe-Ser-Ala-Arg, and has not been showed to utilize the anion-binding exosite 1. As thrombin moves away from the wound site, it aids in the attenuation of blood clotting through the activation of Protein C at the site, Leu-Asp-Pro-Arg (Ehrlich *et al.*, 1990). While thrombin plays multiple roles in blood coagulation, the common theme is the requirement for proper recognition elements in the physiological substrates. Information from the P1-Lys PS-SCL facilitates the identification of recognition elements for productive thrombin-substrate interaction.

The method presented offers a rapid and accurate means of identifying the optimal substrate specificity of a protease. Utility of this method was demonstrated for the serine proteases, plasmin and thrombin, two enzymes that require P1-basic amino acids. This library can have broad application to other enzymes since proteases that cleave P1-basic amino acids are well represented in the serine and cysteine protease families. The defined extended substrate specificity for plasmin and thrombin was in agreement with the cleavage sites in known physiological substrates. Substrate specificity information can aid in the discovery of new physiological substrates and the cleavage sites within substrates. A direct outcome from the library analysis is the creation of sensitive substrates to monitor activity. This information can also be used as a starting point in the design and synthesis of potent and selective inhibitors.

Experimental Procedures

Materials: Unless otherwise noted, chemicals were obtained from commercial suppliers and used without further purification. Aminomethyl Merrifield resin was purchased from Novabiochem, and the substitution level of the resin was determined (0.84 meq/gram) by employing a spectrophotometric Fmoc-quantitation assay (Bunin, 1998). Alkanesulfonamide resin (0.75 mmol/g) was prepared by the method of Backes and Ellman

(Backes and Ellman, 1999) or purchased from Novabiochem. Fmoc-amino acids were purchased from Novabiochem. Iodoacetonitrile and anhydrous NMP were purchased from Aldrich. Iodoacetonitrile was filtered through a small plug of basic alumina immediately prior to use. Anhydrous, low amine content DMF was purchased from EM Science. High-loading tosyl chloride resin (PS-TsCl, 1.55 mmol/g) was purchased from Argonaut Technologies. An Argonaut Quest 210 Organic Synthesizer was employed for library synthesis. Chromatography was carried out using Merck 60 230-240 mesh silica gel according to the procedure of Still (Still *et al.*, 1978). Thin-layer chromatography was carried out on Merck 60 F_{254} 250-µm silica gel plates. IR spectra were recorded neat (for oils) and as films from CH_2Cl_2 or $CHCl_3$ (for crystalline compounds) and only partial data is reported. NMR chemical shifts are reported in ppm down field from an internal solvent peak, or trimethylsilane, and *J* values are in hertz. Elemental analyses were performed by M-H-W labs, Phoenix, AZ. A Savant Speed Vac Plus was employed for concentrating single substrate solutions in vials and library member solutions configured in microtiter plates. The human enzymes, thrombin and plasmin were purchased from Haematologic Technologies Inc. (Essex Jct., VT) and used as received.

Library Synthesis (Work done by Bradley J. Backes and Francesco Leonetti): Fmoc-Amino Acids (Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(O-*t*-Bu)-OH, Fmoc-Glu(O-*t*-Bu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Gly-OH, Fmoc-His(Boc)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Nle, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(O-*t*-Bu)-OH, Fmoc-Thr(O-*t*-Bu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(O-*t*-Bu)-OH, Fmoc-Val-OH) were coupled to the alkanesulfonamide resin (Backes and Ellman, 1999) and the amino acid loading levels were determined employing a spectrophotometric Fmoc-quantitation assay (Bunin, 1998). For the preparation of the P2 sub-library, 0.10 mmol (ca. 200 mg) of each of the 19 Fmoc-amino acid resins were added to 19 reaction vessels (one Fmoc-amino acid resin/vessel) of the Argonaut Quest 210

Organic Synthesizer and solvated with DMF (2 mL/vessel). After agitating 20 min, the DMF was drained and a solution of 20% piperidine in DMF (2 mL/vessel) was added. The resin was agitated for 25 min, filtered, and washed with DMF (3 x 2 mL/vessel). In order to install the randomized P3 position, 10 equiv (ca 1.0 mmol/well, 19 mmol) of an isokinetic mixture (Ostresh *et al.*, 1994) of Fmoc-amino acids (Fmoc-amino acid, mole%: Fmoc-Ala-OH, 3.4; Fmoc-Arg(Pbf)-OH, 6.5; Fmoc-Asn(Trt)-OH, 5.3; Fmoc-Asp(O-*t*-Bu)-OH, 3.5; Fmoc-Glu(O-*t*-Bu)-OH, 3.6; Fmoc-Gln(Trt)-OH, 5.3; Fmoc-Gly-OH, 2.9; Fmoc-His(Boc)-OH, 3.5; Fmoc-Ile-OH, 17.4; Fmoc-Leu-OH, 4.9; Fmoc-Lys(Boc)-OH, 6.2; Fmoc-Nle, 3.8; Fmoc-Phe-OH, 2.5; Fmoc-Pro-OH, 4.3; Fmoc-Ser(O-*t*-Bu)-OH, 2.8; Fmoc-Thr(O-*t*-Bu)-OH, 4.8; Fmoc-Trp(Boc)-OH, 3.8; Fmoc-Tyr(O-*t*-Bu)-OH, 4.1; Fmoc-Val-OH, 11.3) was pre-activated with DICI (3.0 mL, 19 mmol) and HOBt (2.6 mg, 19 mmol) in DMF (57 mL) in a 100 mL round bottom flask. After the 2 min preactivation period, 3 mL of the solution was added to each of the 19 reaction vessels. The resin was agitated for 3 h, filtered, and washed with DMF (3 x 2 mL/vessel). After Fmoc-removal (treatment with 20% piperidine in DMF (2 mL/vessel), agitation for 25 min, filtration, and washing with DMF (3 x 2 mL/vessel)) the randomized P4 position was incorporated in an identical manner. The Fmoc-removal step was followed by filtration and washing with DMF (3 x 2 mL/vessel). A capping solution of AcOH (7.6 mmol), DICI (1.2 mL, 7.6 mmol), HOBt (2.3 g, 7.6 mmol), and DMF (38 mL) was premixed in a 100 mL round bottom flask, and 2 mL was added to each of the 19 reaction vessels. After agitating for 3 h, each resin was filtered, washed (DMF: 3 x 2 mL/vessel, THF: 3 x 2 mL/vessel, MeOH: 3 x 2 mL/vessel), and dried overnight under high vacuum with P_2O_5. In order to install the P1 residue, the 19 resins with a fixed P2 residue (0.020 mmol, ca. 40 mg) were added to the reaction vessels (one fixed P2 residue/vessel), swollen with NMP(1 mL/vessel), agitated for 20 min, and filtered. A solution of ICH_2CN (1.4 mL, 19 mmol), *i*-Pr_2EtN (0.65 mL, 3.8 mmol), and NMP (19 mL) was prepared and added to each of the 19 reaction vessels (1 mL/vessel). The vessels were shielded from light with aluminum foil.

After agitating 24 h, each resin was filtered, washed with NMP (5 x 2 mL/vessel), and DMF (5 x 2 mL), and filtered again. Resin washes were agitated for 5 min/wash. A solution of coumarin-Lys(Boc)-NH_2 (760 mg, 1.9 mmol) in DMF (9.5 mL) was prepared and 0.5 mL (5 equiv) of the solution was added to each vessel. The resin was agitated at 80 °C for 12 h to liberate the coumarin-tetrapeptide derivatives. The 19 reaction mixtures were brought to room temperature (rt), and filtered into 19 individual scintillation vials, each containing DMF, high-loading tosyl chloride resin (125 mg, 0.300 mmol), Et_3N (41 μL, 0.35 mmol), and DMF (1 mL). Each of the 19 resins were washed with DMF (3 x 0.5 mL), and again, the supernatents were filtered into the 19 tosyl chloride resin-containing vials. The vials were agitated with orbital stirring for 3-4 h. The triethylamine salts produced were free-based by adding K_2CO_3 (200 mg, 1.5 mmol) to each vial followed by agitation over 2 h. The 19 reaction mixtures were filtered into 19 scintillation vials and concentrated. Side-chain deprotection was accomplished by adding 1 mL of a TFA:H_2O:triisopropylsilane mixture (95:2.5:2.5) to each vial. After aging for 1 h, the reaction mixtures were concentrated, and EtOH (1 mL) was added to each vial followed by concentration. Ethanol (1 mL) was again added to each vial followed by concentration. The contents of each of the 19 vials were lyophilized after the addition of 1:5 acetonitrile:H_2O (1 mL/well). The synthesis of individual substrates prepared by these methods provided products in 50%-60% yield based upon the loading level of the P2 support-bound Fmoc-amino acid. The yield of coumarin-peptide compounds in each vial was therefore estimated to be ca. 0.01 mmol. The P3 sublibrary and P4 sublibrary were prepared in a similar fashion with the exception that the randomized P2 position was incorporated by hand-mixing the preloaded and quantified Fmoc-amino acid resins. The resin was then transferred to the 19 vessels (0.10 mmol/vessel). In order to supply the fixed positions, each of the 19 Fmoc–amino acids (0.5 mmol, 5 equiv) were individually premixed with DICI (78 mL, 0.5 mmol) and HOBt (68 mg, 0.5 mmol) in DMF (2 mL) in a vial and added to the designated resin-containing vessel. The resin was agitated for 3 h, filtered, and washed with DMF (3 x 2 mL/vessel).

Synthesis of Lys(Boc)-7-amino-4-methylcoumarin (Work done by Bradley J. Backes and Francesco Leonetti: To a 100 mL round bottom flask were added 7-amino-4-methylcoumarin (2.00 g, 11.4 mmol), Fmoc-Lys-OH (5.62 g, 12.0 mmol), and DMF (40 mL). After stirring for 5 min, HATU (4.56 g, 12.0 mmol) and collidine (3.2 mL, 24 mmol) were added. The reaction mixture was stirred overnight at rt, diluted with EtOAc (500 mL), and extracted with 2N HCl (3 x 300 mL). The organic layer was washed with brine (3 x 300 mL), dried (Na_2SO_4), and concentrated. Purification over silica gel (5 x 20 cm eluted with 96:4 $CHCl_3$/MeOH) provided 5.0 g (70%) of Fmoc-Lys(Boc)-7-amino-4-methylcoumarin as a colorless solid: mp 189-191 °C; IR 3305, 1734, 1686, 1663, 1615; ^{1}H NMR (300 MHz) δ 1.31 (s, 9), 1.35-1.38 (m, 6), 1.62-1.66 (m, 2), 2.36 (s, 3), 2.88-2.90 (m, 2), 6.23 (s, 1), 6.70 (bt, 1), 7.29-7.36 (m, 2), 7.38-7.42 (m, 2), 7.44 (d, 1, J = 8.7) 7.70-7.80 (m, 5), 7.85 (d, 2, J = 7.5), 10.50 (s, 1); ^{13}C (101 MHz) δ 17.9, 23.0, 28.2, 29.2, 31.3, 46.6, 55.6, 65.7, 77.3, 105.6, 112.3, 115.0, 115.2, 120.1, 125.3, 125.9, 127.1, 127.6, 140.7, 142.2, 143.7, 143.8, 153.1, 153.6, 155.6, 156.2, 160.0, 172.0. Anal. Calcd for $C_{36}H_{39}N_3O_7$: C, 69.10; H, 6.23; N, 6.71. Found: C, 69.11; H, 6.21; N, 6.71. To a 100 mL round bottom were added Fmoc-Lys(Boc)-7-amino-4-methylcoumarin (5.0 g, 8.0 mmol), DMF (40 mL), and Et_2NH (1.7 mL, 16 mmol). After stirring for 1 h, the reaction mixture was concentrated and purified over silica gel (5 x 20 cm eluted with 95:5 CH_2Cl_2/MeOH) to provide 3.0 g (93%) of NH_2-Lys(boc)-AMC isolated as a solid: mp 113-116 °C; IR 1698, 1690, 1617; ^{1}H NMR (300 MHz) δ 1.30-1.40 (m, 15), 1.51-1.58 (m, 2), 2.36 (d, 3, J = 1.10), 2.85 (d, 2, J = 5.9), 3.28-3.30 (m, 1), 6.21 (d, 1, J = 1.10), 6.75 (t, 3, J = 5.5), 7.50 (dd, 1, J = 2.0, J = 9.5), 7.66 (d, 1, J = 9.5), 7.80 (d, 1, J = 2.0), 10.50 (s, 1); ^{13}C (101 MHz) δ 18.4, 23.2, 28.7, 29.9, 35.0, 56.2, 72.2, 77.8, 106.1, 112.6, 115.4, 115.7, 126.3, 142.8, 153.6, 154.1, 156.0, 160.5, 175.8. Anal. Calcd for $C_{21}H_{29}N_3O_5$: C, 62.50; H, 7.19; N, 10.41. Found: C, 62.46; H, 7.28; N, 10.31.

Synthesis of Arg(Pbf)-7-amino-4-methylcoumarin (Work done by Bradley J. Backes and Francesco Leonetti: To a 100 mL round bottom flask were added 7-amino-4-methylcoumarin (780 mg, 4.5 mmol), Fmoc-Arg(Pbf)-OH (4.42 g, 6.7 mmol), and DMF (10 mL). After stirring for 5 min, HATU (2.0 g, 6.7 mmol) and collidine (1.8 mL, 13 mmol) were added. The reaction mixture was stirred overnight at rt, and then concentrated. The viscous oil was dissolved in hot EtOAc (25 mL) and allowed to cool to rt. The precipitate that had formed upon standing was filtered and washed with EtOAc (3 x 5 mL) to provide 2.2 g (55%) of Fmoc-Arg(Pbf)-7-amino-4-methylcoumarin as a grey solid: mp 224-225 °C; IR 3305, 1719, 1692, 1619; ^{1}H NMR (300 MHz) δ 1.10 (t, 2, *J* = 7.0), 1.35 (s, 6), 170-175 (m, 2), 1.80-1.85 (m, 2), 1.98 (s, 3), 2.34 (s, 6), 2.40 (s, 3), 2.88 (s, 2), 3.01-3.05 (m, 2), 3.99-4.02 (m, 1), 4.17-4.25 (m, 2), 4.30-4.35 (m, 2), 6.3 (s, 1), 6.40 (s, 1), 6.66-6.68 (m, 1), 7.28-7.30 (m, 2), 7.38 (d, 2, *J* = 7.4), 7.47 (d, 1, *J* = 8.8), 7.50-7.53 (m, 1), 7.70-7.73 (m, 3), 7.75-7.78 (m, 2), 7.86 (d, 2, *J* = 7.4), 10.50 (s, 1); ^{13}C (101 MHz) δ 12.2, 14.1, 17.6, 18.0, 18.9, 20.8, 28.2, 30.0, 42.4, 46.7, 55.3, 59.8, 65.7, 86.2, 105.7, 112.3, 115.1, 116.2, 120.1, 124.3, 125.3, 125.9, 127.1, 127.6, 131.4, 137.3, 140.7, 142.2, 143.7, 143.8, 153.1, 153.6, 156.1, 157.4, 160.0, 169.0, 171.8. Anal. Calcd for $C_{45}H_{48}N_5SO_8$: C, 65.56; H, 5.83; N, 8.68. Found: C, 65.28; H, 5.57; N, 8.42. To a 100 mL round bottom were added Fmoc-Arg(Pbf)-7-amino-4-methylcoumarin (2.0 g, 2.4 mmol), DMF (12 mL) and Et_2NH (500 μL, 4.8 mmol). After stirring for 1 h, the reaction mixture was concentrated and purified over silica gel (5 x 20 cm eluted with 90:10 $CHCl_3$/MeOH) to provide 1.3 g (90%) of NH_2-Arg(pbf)-AMC isolated as a colorless solid: mp 134-137 °C; IR 3305, 1719, 1692, 1619; ^{1}H NMR (300 MHz) δ 1.05 (t, 2, *J* = 7.0), 1.35 (s, 6), 1.41-1.52 (m, 3), 1.58-1.60 (m, 1), 1.94 (s, 3), 2.36 (s, 3), 2.38 (s, 3), 2.43 (s, 3), 2.89 (s, 2), 3.03-3.06 (m, 2), 3.30-3.38 (m, 1), 3.40-3.43 (m, 2), 4.33 (s, 1), 6.23 (s, 1), 6.36 (s, 1), 6.71 (s, 1), 7.50-7.60 (m,1), 7.68 (d, 1, $J = 8.7$), 7.81 (d, 1, $J = 1.8$), 10.08 (s, 1); ^{13}C (101 MHz) δ 12.2, 17.6, 17.9, 18.6, 18.9, 26.0, 28.3, 32.1, 42.4, 55.4, 56.0, 86.3, 105.6, 112.2, 114.9, 115.3, 116.2, 124.3, 125.9, 131.4, 137.3, 142.3, 153.1, 153.6, 156.1, 157.4, 160.1, 175.1.

Synthesis of Single Substrates: Single substrates for kinetic analysis were prepared by the previously described methods, except that single amino acids were employed in place of mixtures. The P1 residue was introduced employing Lys(Boc)-7-amino-4-methylcoumarin or Arg(Pbf)-7-amino-4-methylcoumarin. After side-chain deprotection, the unpurified products were subjected to C18 reverse phase HPLC with a 10-40% gradient of 0.1% trifluoroacetic acid and 0.08% trifluoroacetic acid/95% acetonitrile. The purified products were subsequently lyophilized. All coumarin tetrapeptides displayed appropriate molecular masses as determined by mass spectrometry (data not shown).

Enzymatic Assay of Library: The protein concentration of the enzymes was determined by absorbance measured at 280 nm, plasmin's extinction coefficient is 1.70 mL mg^{-1} cm^{-1} (Robbins *et al.*, 1981) and thrombin's extinction coefficient is 1.83 mL mg^{-1} cm^{-1} (Fenton *et al.*, 1977). The proportion of catalytically active protein was quantitated by active-site titration with MUGB (Jameson, 1973a). Briefly, fluorescence was monitored, with excitation at 360nm and emission at 450nm, upon addition of enzyme to MUGB. The concentration of enzyme was determined from the increase in fluorescence based on a standard concentration curve.

Substrates from the PS-SCL were dissolved in DMSO. Approximately 2.5x10^{-9} mol of each sub-library (361 compounds) were added to 57 wells of a 96-well Microfluor White "U" bottom plate (Dynex Technologies, Chantilly, VA). Final substrate concentration was approximately 0.25 μM, making the hydrolysis of the AMC group directly proportional to the specificity constant, k_{cat}/K_m. Hydrolysis reactions were initiated by the addition of enzyme (0.5 nM-10 nM) and monitored fluorometrically with a Perkin Elmer LS50B Luminescence Spectrometer 96-well plate reader, with excitation at 380nm and emission at 460nm (Zimmerman *et al.*, 1977). Assays were performed in a buffer containing 50 mM Tris, pH 8.0, 100 mM NaCl, 5mM $CaCl_2$, 1% DMSO (from substrates) and either 1mg/mL BSA or 0.01% Tween-20.

Single substrate kinetic assays: Enzyme activity was monitored at 25°C in assay buffer containing 50 mM Tris pH 8.0 and 100 mM NaCl, 5 mM $CaCl_2$ and 0.01% Tween-20. Substrate stock solutions were prepared in DMSO. The final concentration of substrate ranged from 0.005-2 mM, the concentration of DMSO in the assay was less than 5%. Enzyme concentrations ranged from 5-50 nM. Hydrolysis of AMC-substrates was monitored fluorometrically with an excitation wavelength of 380 nm and emission wavelength of 460 nm on a Fluoromax-2 spectrofluorimeter.

Molecular modeling of Thrombin-substrate and Plasmin-substrate complex. The coordinates for thrombin bound to D-Phe-Pro-Arg-chloromethylketone (1PPB) (Bode *et al.*, 1989) and plasmin complexed with streptokinase (1BML) (Wang *et al.*, 1998) were obtained from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). The Biopolymer module of the Insight II (Molecular Simulations Inc. San Diego, CA) molecular modeling package was used to build and model the substrates into the active sites of the proteases.

Figure 3-1. Figure 3-1. Three sub-libraries (P2, P3, P4) each made up of 19 wells containing 361 compounds are prepared by a segment condensation reaction with a P1 fluorogenic amino acid substrate. Individual proteinogenic amino acids are employed to incorporate spatially addressed positions "O", while an isokinetic mixture of proteinogenic amino acids is employed to incorporate varied positions "X".

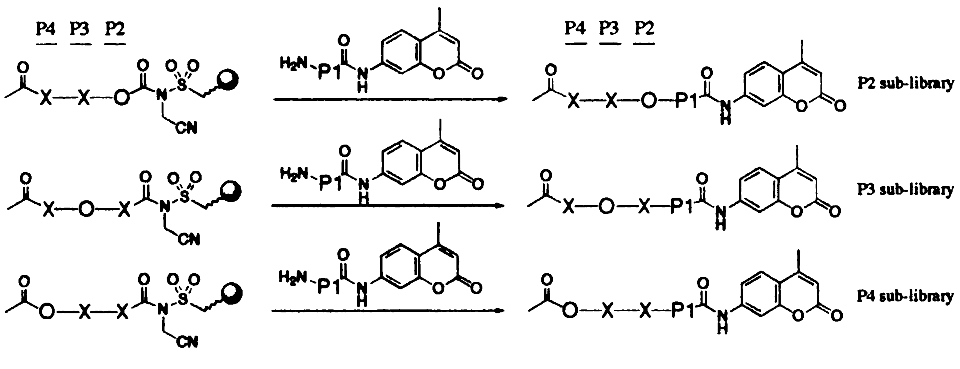

P4 P3 P2
P4 P3 P2
P2 sub-library
P3 sub-library
P4 sub-library

Figure 3-2.

A. Activity of Plasmin in a P1-Lys Positional Scanning-Synthetic Combinatorial Library. Y-axis is pM of fluorophore released per second. X-axis indicates the amino acid held constant at each position, designated by the one-letter code (norleucine is represented by "Z").

B. Activity of Thrombin in a P1-Lys Positional Scanning-Synthetic Combinatorial Library. Y-axis is pM of fluorophore released per second. X-axis indicates the amino acid held constant at each position, designated by the one-letter code (norleucine is represented by "Z".

Figure 3-2A.

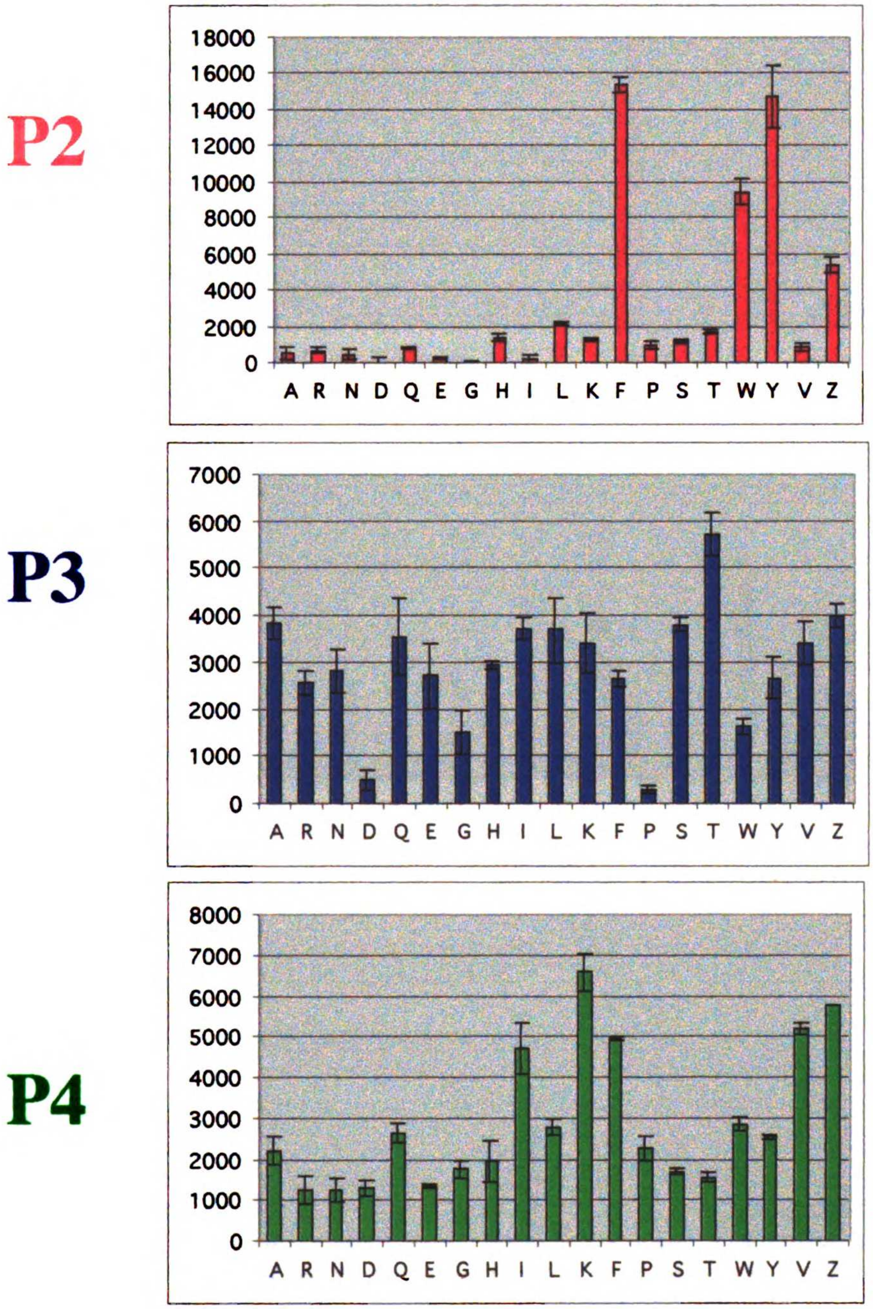

P2
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
A R N D Q E G H I L K F P S T W Y V Z
P3
7000
6000
5000
4000
3000
2000
1000
0
A R N D Q E G H I L K F P S T W Y V Z
P4
8000
7000
6000
5000
4000
3000
2000
1000
0
A R N D Q E G H I L K F P S T W Y V Z

Figure 3-2B.

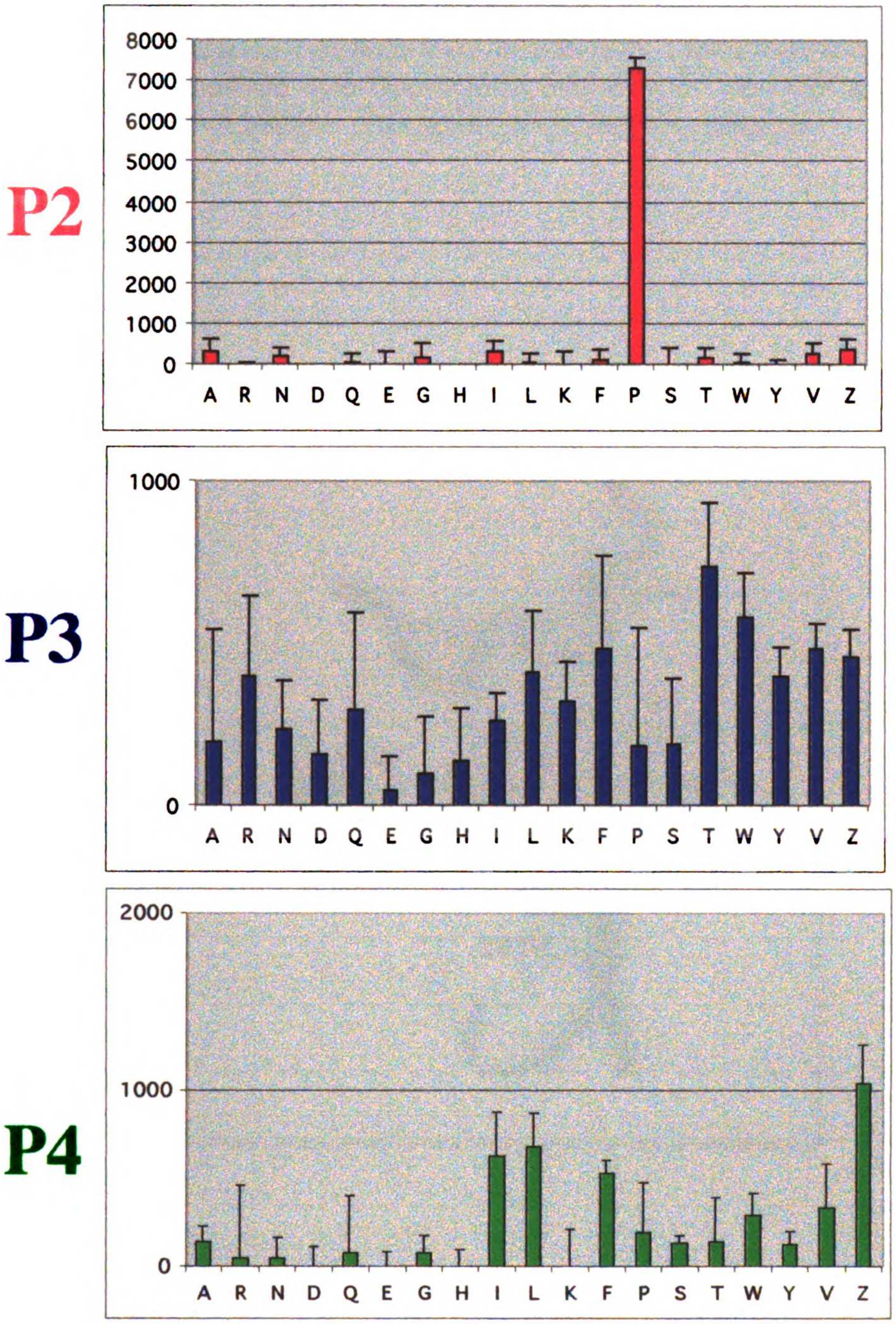

P2
8000
7000
6000
5000
4000
3000
2000
1000
0
A R N D Q E G H I L K F P S T W Y V Z
P3
1000
0
A R N D Q E G H I L K F P S T W Y V Z
P4
2000
1000
0
A R N D Q E G H I L K F P S T W Y V Z

Figure 3-3A. Three-dimensional model of plasmin bound to the tetrapeptide substrate Lys-Thr-Phe-Lys. The enzyme is shown in green, with the catalytic triad (His 57, Asp 102, and Ser 195) in yellow. The substrate is shown in magenta.

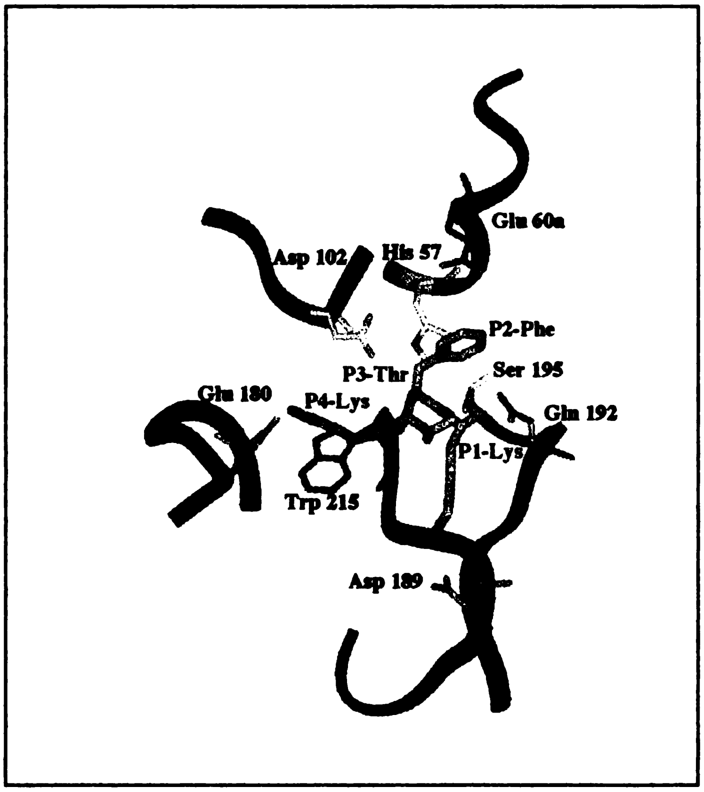
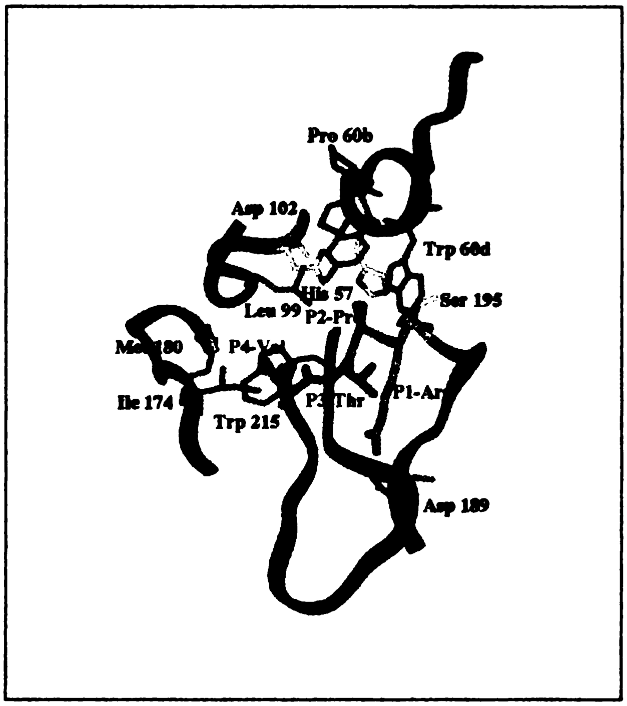

Glu 60a
Asp 102
His 57
P2-Phe
P3-Thr
Ser 195
Glu 180
P4-Lys
Gln 192
P1-Lys
Trp 215
Asp 189

Figure 3-3B. Three-dimensional model of thrombin bound to the tetrapeptide substrate Val-Thr-Pro-Arg. The enzyme is shown in green, with the catalytic triad (His 57, Asp 102, and Ser 195) in yellow. The substrate is shown in magenta.



Pro 60b
Asp 102
Trp 60d
His 57
Ser 195
Leu 99
P2-Pr
Me 180
P4-V
Ile 174
P3 Thr
P1-Ar
Trp 215
Asp 189

Figure 3-4. Physiological Substrate Cleavage Sites of Plasmin.

	P4	**P3**	**P2**	**P1**
Optimal Plasmin Substrate Specificity from PS-SCL	K/V/I/F	X	F/Y/W	R/K
Vitronectin	**K**	**G**	**Y**	**R**
Osteocalcin	**E**	**A**	**Y**	**R**
Factor X	**I**	**T**	**F**	**R**
PAR1 (70)	**T**	**E**	**Y**	**R**

Figure 3-5. Physiological Substrate Cleavage Sites of Thrombin.

	P4	**P3**	**P2**	**P1**
Optimal Thrombin Substrate Specificity from PS-SCL	*L, I, V, F*	*X*	*P*	*R*
PAR1	**L**	**D**	**P**	**R**
Factor V (709)	**L**	**G**	**L**	**R**
Factor V (1018)	**L**	**S**	**P**	**R**
Factor V (1545)	**W**	**Y**	**L**	**R**
Factor VIII (372)	**I**	**Q**	**I**	**R**
Factor VIII (740)	**I**	**E**	**P**	**R**
Factor VIII (1689)	**Q**	**S**	**P**	**R**
Factor XI	**I**	**K**	**P**	**R**
Factor XIII	**V**	**V**	**P**	**R**
Fibrinogen A α	**G**	**G**	**V**	**R**
Fibrinogen B β	**F**	**S**	**A**	**R**
Protein C	**L**	**D**	**P**	**R**
uPA	**L**	**R**	**P**	**R**

Table 3-1.

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} s^{-1})
Plasmin			
Ac-K-T-Y-K-AMC	11.3 ± 0.4	20.8 ± 3.3	0.544 ± 0.071
Ac-K-T-F-K-AMC	20.1 ± 0.6	29.7 ± 3.6	0.677 ± 0.069
Ac-K-T-W-K-AMC	11.9 ± 0.3	19.9 ± 2.2	0.601 ± 0.054
Ac-K-T-S-K-AMC	8.8 ± 1.3	440 ± 100	0.020 ± 0.002
Ac-F-T-Y-K-AMC	17.5 ± 0.9	51.0 ± 10.2	0.342 ± 0.054
Ac-L-T-F-K-AMC	33.2 ± 3.9	296 ± 70	0.112 ± 0.015
Ac-L-E-F-K-AMC	5.5 ± 0.3	74.6 ± 9.9	0.073 ± 0.006
Thrombin			
Ac-Z-T-P-R-AMC	45.0 ± 1.1	11.3 ± 1.3	3.83 ± 0.35
Ac-V-T-P-R-AMC	30.8 ± 1.3	29.6 ± 4.5	1.04 ± 0.13
Ac-Z-T-L-R-AMC	5.8 ± 0.3	67.4 ± 10.7	0.085 ± 0.009
Ac-L-G-V-R-AMC	15.6 ± 1.8	101.5 ± 29.5	0.154 ± 0.029
Ac-G-G-V-R-AMC	1.2 ± 0.1	180.7 ± 55.3	0.007 ± 0.001

Chapter Four:

Rapid and general methods for profiling protease specificity: libraries of novel 7-amino-4-acetamidecoumarin (AAC) fluorogenic substrates

ABSTRACT

The function of a protease is largely dependent on its ability to recognize and cleave specific substrates. A method is presented here for the utilization of fluorogenic peptide substrates to rapidly identify the primary and N-terminal-extended specificity of a protease. The substrates contain the fluorogenic-leaving group, 7-amino-4-acetamidecoumarin (AAC), that has an increased fluorescence yield over the traditionally used 7-amino-4-methylcoumarin (AMC). The bifunctionality of AAC allows for the production of libraries using standard solid-phase synthesis techniques with the incorporation of any amino acid at the P1-position. Using this strategy, soluble positional protease substrate libraries of 6,859 and 137,180 members, possessing amino acid diversity at three and four positions, respectively, were constructed. A library incorporating the AAC leaving group gave a comparable kinetic profile to one that incorporated the AMC leaving group. These libraries allowed for the rapid determination of the substrate specificities of a diverse array of proteases, including the human serine proteases thrombin, plasmin, factor Xa, uPA, tPA, skin tryptase, lung tryptase, granzyme A, the rat serine proteases granzyme B and trypsin, the bovine serine protease chymotrypsin, the porcine elastase serine protease, the serine collagenase from fiddler crab, and the cysteine proteases cathepsin B, cathepsin L, papain and cruzain.

ABBREVIATIONS USED

AMC, 7-amino-4-methyl coumarin; AAC, 7-amino-4-acetamide coumarin; RFU, relative fluorescence units; Z and Nle, norleucine; PS-SCL, positional scanning-synthetic combinatorial library; MUGB, 4-methylumbelliferyl *p*-guanidinobenzoate; MUTMAC, 4-methylumbelliferyl *p*-trimethylammonium cinnamate chloride; Tris, *tris*-(hydroxymethyl)-amino-methane; BSA, bovine serum albumin; Ac, acyl; DICI, diisopropylcarbodiimide; HOBt, 1-hydroxybenzotriazole; TIS, triisopropyl silane; HATU, 2-(1H-9-Azabenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; TFA, trifluoroacetic acid; Fmoc, 9-fluorenylmethoxycarbonyl; pbf, 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl; trt, trityl; boc, tert butoxycarbonyl; DMF, dimethylformamide; NMP, N-methylpyrrolidine; uPA, urokinase-type plasminogen activator; tPA, tissue-type plasminogen activator.

INTRODUCTION

Substrate specificity, or the ability of an enzyme to discriminate among many potential substrates, is an important factor in maintaining the fidelity of most biological functions. Substrate specificity can be regulated on many levels, including spatial and temporal localization of enzyme and substrate, concentration of enzyme and substrate, cofactor requirement, and primary catalytic determinants of the enzyme. With the rapid discovery of new enzyme sequences, the development of high throughput assays is necessary to ascribe functional properties to these enzymes. Here we describe a rapid, generalized, and high-throughput assay for determination of the extended substrate specificity of proteases.

MATERIALS AND METHODS

Reagents and General Methods. Rink Amide AM resin and Fmoc-amino acids were purchased from Novabiochem. The amine substitution level of the Rink resin was determined (0.80 meq/gram) by a spectrophotometric Fmoc-quantitation assay (Bunin, 1998). Anhydrous, low-amine content DMF was purchased from EM Science. HATU was purchased from Perseptive Biosystems. DICI, HOBt, AcOH, Fmoc-Cl, TFA, and TIS were purchased from Aldrich. An Argonaut Quest 210 Organic Synthesizer was employed for the preparation of Fmoc-amino acid-substituted AAC resins. All library synthesis was performed in 96-well plates employing the Multi-Chem synthesis apparatus of Robbins Scientific. A Savant Speed Vac Plus was employed for concentrating library compound wells and solutions of single substrates. The human enzymes, thrombin, plasmin, and factor Xa were purchased from Haematologic Technologies Inc. (Essex Jct., VT) and used as received. Human light chain uPA, human neutrophil elastase, bovine cathepsin B and bovine cathepsin L were purchased from Calbiochem and used as received. Papain and bovine chymotrypsin were purchased from Sigma and used as received. Rat granzyme B was heterologously expressed in *Pichia pastoris* and purified as previously described (Harris *et al.*, 1998). Human granzyme A was heterologously expressed in *Pichia pastoris* and purified as described in Chapter Seven. Human skin and lung tryptase was heterologously expressed in *Pichia pastoris* as previously described (Niles *et al.*, 1998). Fiddler crab collagenase was purified fiddler crab pancreatic extract as previously described (Tsu *et al.*, 1994). Cruzain was heterologously expressed in *E. coli* and purified as previously described (Eakin *et al.*, 1992). Rat trypsin was heterologously expressed and purifed from the yeast *Pichia pastoris* (Halfon, 1996).

AAC-Resin Synthesis (Sythesized by Brad Backes) (Figure 4-1): 7-Fmoc-aminocoumarin-4-acetic acid was prepared by treating 7-aminocoumarin-4-acetic acid with

Fmoc-Cl. To a 1-L flask was added 7-aminocoumarin-4-acetic acid (10.0 g, 45.6 mmol) and H_2O (228 mL). $NaHCO_3$ (3.92 g, 45.6 mmol) was added in small portions followed by the addition of acetone (228 mL). The solution was cooled with an ice bath, and Fmoc-Cl (10.7 g, 41.5 mmol) was added in several portions with vigorous stirring over the course of an hour. The ice bath was then removed and the solution was stirred overnight. The acetone was removed with rotary evaporation and the resulting gummy solid was collected via filtration, and washed with several portions of hexane. The material was dried over P_2O_5 to give 14.6 g (80%) of a cream-colored solid.

AAC-Resin was prepared by the condensation of Rink Amide AM resin with 7-Fmoc-aminocoumarin-4-acetic acid. To a 500 mL round bottom flask were added Rink Amide AM resin (21 g, 17 mmol) and DMF (200 mL). The mixture was agitated employing a shaker arm for 30 min and filtered with a filter cannula (Pharmacia, Uppsala, Sweden) wherupon 20% piperidine in DMF (200 mL) was added. After agitating for 25 min, the resin was filtered and washed with DMF (3 x 200 mL). 7-Fmoc-aminocoumarin-4-acetic acid (15 g, 34 mmol), HOBt (4.6 g, 34 mmol), and DMF (150 mL) were added to the flask followed by the addition of DICI (5.3 mL, 34 mmol). The mixture was agitated overnight, filtered, washed (DMF: 3 x 200 mL, THF: 3 x 200 mL, MeOH: 3 x 200 mL), and dried over P_2O_5. The substitution level of the resin was 0.58 mmol/g (>95%) by Fmoc-analysis.

P1-Substituted AAC-Resin Synthesis (Synthesized by Brad Backes). Fmoc-AAC-Resin (100 mg, 0.058 mmol) was added to each of the 20 reaction vessels of an Argonaut Quest 210 Organic Synthesizer and solvated with DMF (2 mL). The resin was filtered and 20% piperidine in DMF (2 mL) was added to each vessel. After agitating for 25 min, the resin was filtered and washed with DMF (3 x 2 mL). An Fmoc-amino acid (0.29 mmol), DMF (0.7 mmol), collidine (76 μL, 0.58 mmol) and HATU (110 mg, 0.29 mmol) were added to the designated reaction vessel followed by agitiation for 20 h. The resins were then filtered, washed with DMF (3 x 2 mL), and subjected a second time to the coupling conditions.[i] A solution of AcOH (40 µL, 0.7 mmol), DICI (109 µL, 0.7 mmol), nitrotriazole (80 mg, 0.7 mmol) in DMF(0.7 mL) was added to each of the reaction vessels folled by agitation over a 24 h period. The resins were filtered, washed (DMF: 3 x 2 mL; THF: 3 x 2 mL; MeOH: 3 x 2 mL), and dried over P_2O_5. The substitution level of each resin was determined by Fmoc-analysis.

P1-Fixed Library Synthesis. Multi-gram quantities of P1-substituted AAC-resin could be synthesized using the methods described. P1-Lys, P1-Arg, and P1- Leu P1-fixed libraries were prepared. Fmoc-Amino acid-substituted AAC resin (ca. 25 mg, 0.0125 mmol, of Lys, Arg, or Leu) was placed in a wells A1-A10, B1-B9, D1-D10, E1-E9, G1-G10, H1-H9 of a Multi-Chem reaction block. The resin containing wells were solvated with DMF (0.5 mL), agitated for 30 min, and filtered. A 20% piperidine in DMF solution (0.5 mL) was added to the wells followed by agitation for 30 min. The wells were filtered and washed with DMF (3 x 0.5 mL). To each well A1-A10 and B1-B9 was added a single Fmoc-amino acid (0.15 mmol), HOBt (20 mg, 0.15 mmol), DICI (23 µL, 0.15 mmol) and DMF (0.5 mL). To wells D1-D10, E1-E9, G1-G10, H1-H9 was added an isokinetic mixture of Fmoc-amino acids as described in the previous section. The reaction block was agitated for 3h, filtered, and washed with DMF (3 x 0.5 mL). A 20% piperidine in DMF solution (0.5 mL) was added to the wells followed by agitation for 30 min. The wells were filtered and washed with DMF (3 x 0.5 mL). To wells C1-C10 and D1-D9 were added a single Fmoc-amino acid (0.15 mmol), HOBt (20mg, 0.15 mmol), DICI (23 µL, 0.15 mmol) and DMF (0.5 mL). To wells A1-A10, B1-B9, G1-G10, H1-H9 was added an isokinetic mixture of Fmoc-amino acids. The reaction block was agitated for 3h, filtered, and washed with DMF (3 x 0.5 mL). A 20% piperidine in DMF solution (0.5

mL) was added to the wells followed by agitation for 30 min. The wells were filtered and washed with DMF (3 x 0.5 mL). To wells G1-G10 and H1-H9 were added a single Fmoc-amino acid (0.15 mmol), HOBt (20mg, 0.15 mmol), DICI (23 μL, 0.15 mmol) and DMF (0.5 mL). To wells A1-A10, B1-B9, C1-C10, D1-D9 was added an isokinetic mixture of Fmoc-amino acids. The reaction block was agitated for 3h, filtered, and washed with DMF (3 x 0.5 mL). A 20% piperidine in DMF solution (0.5 mL) was added to the wells followed by agitation for 30 min. The wells were filtered and washed with DMF (3 x 0.5 mL). The resin-containing wells were treated with 0.5 mL of a capping stock solution of AcOH (150 μL, 2.5 mmol), HOBt (338 mg, 2.5 mmol) and DICI (390 μL, 2.5 mmol) in DMF (10 mL). After the reaction block was agitated for 4h, the resin–containing wells were washed with DMF (3 x 0.5 mL), CH_2Cl_2 (3 x 0.5 mL), and treated with a solution of 95:2.5:2.5 TFA/TIS/$H_2$0. Cleavage work-up was identical to that described in the previous section.

P1-Diverse Library Synthesis. To wells A-1 through A-10 and B-1 through B-10 of a Multi-Chem 96-well reaction apparatus was added an individual P1-substituted Fmoc-amino acid AAC-resin (ca. 25 mg, 0.0125 mmol). To each of the resin-containing wells was added DMF (0.5 mL). The reaction block was agitated for 30 min and filtered. A 20% piperidine in DMF solution (0.5 mL) was added to each of the 20 wells followed by agitation for 30 min. The wells of the reaction block were filtered and washed with DMF (3 x 0.5 mL). In order to introduce the randomized P2 position, an isokinetic mixture (Ostresh *et al.*, 1994) of Fmoc-amino acids (2.5 mmol, 10 equiv/well; Fmoc-amino acid, mol%: Fmoc-Ala-OH, 3.4; Fmoc-Arg(Pbf)-OH, 6.5; Fmoc-Asn(Trt)-OH, 5.3; Fmoc-Asp(O-*t*-Bu)-OH, 3.5; Fmoc-Glu(O-*t*-Bu)-OH, 3.6; Fmoc-Gln(Trt)-OH, 5.3; Fmoc-Gly-OH, 2.9; Fmoc-His(Trt)-OH, 3.5; Fmoc-Ile-OH, 57%, 17.4; Fmoc-Leu-OH, 4.9; Fmoc-Lys(Boc)-OH, 6.2; Fmoc-Nle-OH, 3.8; Fmoc-Phe-OH, 2.5; Fmoc-Pro-OH, 4.3; Fmoc-

Ser(O-*t*-Bu)-OH, 2.8; Fmoc-Thr(O-*t*-Bu)-OH, 4.8; Fmoc-Trp(Boc)-OH, 3.8; Fmoc-Tyr(O-*t*-Bu)-OH, 4.1; Fmoc-Val-OH, 11.3.) was pre-activated with DICI (390 μL, 2.5 mmol), and HOBt (338 mg, 2.5 mmol) in DMF (10 mL) in a 50 mL round bottom flask. After the 2 min pre-activation period, 0.5 mL of the solution was added to each of the 20 resin-containing wells. The reaction block was agitated for 3h, filtered, and each well was washed with DMF (3 x 0.5 mL). After Fmoc-removal as before, the randomized P3 position was incorporated in an identical manner. The randomized P4 position was installed after P3 Fmoc-removal. The Fmoc of the P4 amino acid was removed and the resin was washed with DMF (3 x 0.5 mL), and treated with 0.5 mL of a capping solution of AcOH (150 μL, 2.5 mmol), HOBt (338 mg, 2.5 mmol) and DICI (390 μL, 2.5 mmol) in DMF (10 mL). After 4 h of agitation, the resin was washed with DMF (3 x 0.5 mL), CH_2Cl_2 (3 x 0.5 mL), and treated with a solution of 95:2.5:2.5 TFA/TIS/H_2O. After aging for 1h, the reaction block was placed in dry ice to freeze the well contents. The reaction block was opened and placed on a 96-well deep-well titer plate and allowed to come to room temperature whereupon the wells were washed with additional cleavage solution (2 x 0.5 mL). The deep-well collection plate was concentrated, and the substrate-containing wells were diluted with EtOH (0.5 mL), concentrated, diluted with EtOH (0.5 mL) and concentrated again. The contents of the individual wells were lyophilized from CH_3CN:H_2O mixtures. The total mmol of substrate in each well was conservatively estimated to be 0.00625 mmol (50%) based upon yields of single substrates.

Synthesis of Single Substrates. Single substrates for kinetic analysis were prepared employing the methods described previously. P1-substituted Fmoc-amino acid AAC-resin was employed to prepare the single substrates for kinetic analysis. The P2-P4 positions were coupled employing Fmoc-amino acid (5 equiv), HOBt (5 equiv), and DICI (5 equiv), in DMF (0.3 M). The resin was capped with AcOH and cleaved with 95:2.5:2.5

TFA/TIS/H20 as described. The unpurified products were subjected to reversed-phase HPLC preparatory chromatography followed by lyophilization.

Fluorescence Properties of 7-Amino-4-Acetamide-coumarin. The fluorescence of free AAC and peptidyl-derivitatized AAC were detected on a Spex fluorimeter thermostated to 25°C. Excitation wavelengths of 300-410nm, at 5nm intervals, were used with emission wavelengths of 410-500nm, at 5nm intervals, to determine optimal excitation and emission parameters.

Enzymatic Assay of Library. The concentration of the protein enzymes was determined by absorbance measured at 280 nm using the following extinction coefficients: plasmin, 1.70 mL mg^{-1} cm^{-1} (Robbins *et al.*, 1981); thrombin, 1.83 mL mg^{-1} cm^{-1} (Fenton *et al.*, 1977); granzyme B, 0.52 mL mg^{-1} cm^{-1} (Harris *et al.*, 1998); trypsin, 1.45 mL mg^{-1} cm^{-1}; chymotrypsin, 2.16 mL mg^{-1} cm^{-1}; porcine elastase, 2.64 mL mg^{-1} cm^{-1}; uPA, 1.59 mL mg^{-1} cm^{-1}; tPA, 1.42 mL mg^{-1} cm^{-1}; factor Xa, 1.28 mL mg^{-1} cm^{-1}; papain, 2.29 mL mg^{-1} cm^{-1}; and cruzain, 2.69 mg^{-1} cm^{-1} (Gill and von Hippel, 1989). The proportion of catalytically active thrombin, plasmin, trypsin, uPA, tPA was quantitated by active-site titration with MUGB (Jameson, 1973a). The proportion of catalytically active chymotrypsin was quantitated by active-site titration with MUTMAC. Briefly, fluorescence was monitored, with excitation at 360nm and emission at 450nm, upon addition of enzyme to MUGB or MUTMAC. The concentration of enzyme was determined from the increase in fluorescence based on a standard concentration curve.

Substrates from the PS-SCLs were dissolved in DMSO. Approximately 1.0 x 10^{-10} mol of each P1-diverse sub-library (6859 compounds) was added to 20 wells of a 96-well microfluor white "U" bottom plate (Dynex Technologies, Chantilly, VA) for a final concentration of 0.01 µM in each compound. Approximately $1.0\text{x}10^{-9}$ mol of each P1-Lys, P1-Arg, or P1-Leu sub-library (361 compounds) was added to 57 wells of a 96-well microfluor plate for a final concentration of 0.1 μM. Hydrolysis reactions were initiated by the addition of enzyme (0.02 nM-100nM) and monitored fluorometrically with a Perkin Elmer LS50B Luminescence Spectrometer, with excitation at 380nm and emission at 450nm or 460nm. Assays for the serine proteases were performed in a buffer containing 50 mM Tris, pH 8.0, 100 mM NaCl, 0-5mM $CaCl_2$, 0.01% Tween-20, and 1% DMSO (from substrates). Assay of the cysteine proteases, papain and cruzain, was performed in a buffer containing 100 mM sodium acetate, pH 5.5, 100 mM NaCl, 5 mM DTT, 1 mM EDTA, 0.01% Brij-35, and 1% DMSO (from substrates).

RESULTS

Construction AAC-Based Substrate Libraries. While several traceless linkage strategies to tether the coumarin to the support were considered, direct attachment of a bifunctional Fmoc-coumarin-acid **2** to acid labile Rink linker **1** employed for Fmoc-peptide synthesis was employed (Fig. 4-1). Coupling amino acids to the poor coumarin nucleophile in high yields is a formidible task. Modified coupling conditions of Carpino perfomed well (Carpino *et al.*, 1996), providing P1-substituted AAC Resin in good yields (57->98%) for the majority of residues. Traditionally dificult to couple derivatives such as Ile can be coupled a second time to increase the substitution level. The remaining free aniline fluorophore is capped with AcOH using a nitrotirazole active ester perpared *in situ*. The substitution level of the P1-substituted AAC resins can be acurately assessed with an Fmoc-analysis assay.

Fluorescence properties of 7-amino-4-acetamide-coumarin. The excitation and emission maxima of the amino-conjugated 7-amino-4-acetamide-coumarin (AAC) substrate is 325 nm and 390 nm, respectively (Table 4-1). Cleavage of the substrate to release the free 7-amino-4-acetamide coumarin results in a shift of the excitation and emission maxima to 350 nm and 450 nm, respectively (Table 4-1). The AAC fluorophore has a significantly higher fluorescence yield than the AMC coumarin at the excitation and emission wavelengths of 380 nm and 460 nm (Table 4-1). The enhanced fluorescence of the AAC group allows for the more sensitive detection of proteolytic activity.

Proteolytic comparison of AAC versus AMC. In order to compare the performance of the AAC to AMC as a proteolytic-substrate leaving group, an AAC-P1-fixed Lys PS-SCL was prepared for comparison to an AMC-P1-fixed Lys PS-SCL. For the synthesis of the AAC library, Lys was positioned at the P1 site. Three sub-libraries denoting the second fixed position (P4,P3,P2) and consisting of 19 wells addressing a fixed amino acid (Cys was omitted and Nle was substituted for Met) were prepared (361 compounds/well and 6,859 compounds/library). The extended substrate specificity of thrombin was profiled with the AAC-P1-fixed Lys PS-SCL and shown to have a P4-preference for the aliphatic amino acids, Leu, Ile, and Nle, a broad P2-preference, and a strict P2-preference for proline (Figure 4-2). This specificity profile is comparable to that observed in the AMC-P1-fixed Lys library.

Profiling of serine proteases with P1-fixed positional libraries. The extended P4-P2 substrate specificity of several serine proteases of the chymotrysin-family were profiled with tetrapeptide libraries in which the P1-position was held constant as either lysine, arginine or leucine, the P2, P3, P4 positions were individually positioned, and the remaining two positions were randomized. Because the enhanced fluorescence properties of the AAC fluorophore (Table 4-1), the concentration of each substrate was minimized to 0.1 μM.

Plasmin, a protease involved in fibrinolysis, has a P1-preference for lysine. Recently, we have shown plasmin to have a distinct preference for aromatic amino acids at the P2 position and lysine at P4 (Harris *et al.*, 1999). As is consistent with that data, the substrate specificity profile of plasmin in the AAC P1-fixed lysine library is for P4-lysine, broad P3-specificity, and P2-aromatic amino acids (Figure 4-3K).

Thrombin prefers cleavage after P1-arginine over cleavage after P1-lysine. However, the specificity preference of thrombin, when profiled with both the P1-Arg (Figure 4-3A) and P1-Lys (Figure 4-2) libraries, shows little difference in the extended subsites. Thrombin has a preference for aliphatic amino acids at the P4 position, little preference at P3, and strict preference for proline at the P2-position (Harris *et al.*, 1999).

Correlation of thrombin's optimal substrate sequence with that found in its physiological substrates has been observed (Harris *et al.*, 1999).

Two enzymes that have been extensively characterized for their substrate specificity are tissue-type plasminogen activator (t-PA) (Coombs *et al.*, 1996; Ding *et al.*, 1995) and urokinase plasminogen activator (u-PA) (Ke *et al.*, 1997a; Ke *et al.*, 1997c). Both t-PA and u-PA are responsible for converting the zymogen, plasminogen, to the active form of the enzyme, plasmin. However, their P4-P2 substrate specificities are very different. Through the profiling of the extended substrate specificity of these enzymes (Figure 4-3 B & C), we observe them both to have preference for small amino acids at P2 (Gly/Ala/Ser) and no significant preference at P4, except the decreased activity for acidic amino acids. However, their P3 preferences are quite disparate with t-PA showing a preference for aromatic amino acids (Phe and Tyr) and u-PA for small polar amino acids (Thr and Ser). This difference in P3-specificity was also noted by Ke et al. to be a major distinction between the two plasminogen activators (Ke *et al.*, 1997c).

Factor Xa is an enzyme that plays the critical function of activating prothrombin in blood coagulation (Davie *et al.*, 1991). The extended substrate specificity of factor Xa has not been previously characterized. Through profiling with the P1-Arg library, we find it to have a slight preference for P4-aliphatic amino acids, a slight P3 preference for arginine, with an absence of P3-proline activity, and an unusual P2-preference for glycine (Figure 4-3 D).

Tryptase is a protease released from the granules of mast cells (Caughey, 1994; Welle, 1997). There are two isoforms found in humans, skin and lung tryptase, which differ only in a glycosylation site. Tryptase has an unusual structure in that it is a tetramer that occludes the active site to large macromolecules. We profiled both isoforms with the P1-Arg and P1-Lys libraries. Both isoforms of the enzyme were shown to have similar preferences for P4-proline, P3-basic amino acids (Arg/Lys), and P2-asparagine (Figure 4-3 G-J). We made tetrapeptide substrates to test the dependence on the determined optimal substrate sequence through determination of steady state constants (Table 4-2). The steady state kinetic results show good agreement with the library analysis.

Profiling of cysteine proteases with P1-fixed positional libraries. The positional substrate libraries with the AAC fluorogenic leaving group are also conducive for defining cysteine protease specificity. The P4-P2 extended substrate specificity for papain and cruzain were defined using the AAC P1-fixed Arg library. Cysteine proteases of the papain family have been shown to have a major substrate specificity at the P2-position. This position usually shows a preference for hydrophobic amino acids. Indeed, from the library screening (Figure 4-3 E & F), we observe papain to have a preference for P2-Val>Phe>Tyr>Nle and cruzain to have a P2-preference for Leu>Tyr>Phe>Val. While the P3 specificity is rather broad, papain does show a preference for Pro, whereas cruzain has a preference for the basic amino acid, arginine and lysine. The P4 position is again, very broad for both enzymes, but interesting observations arise from testing all possible substrates. Papain shows very little specificity at P4. While cruzain also shows broad activity at the P4-position, there is a distinct lack of specificity for large aliphatic and aromatic amino acids. This absence is also seen in a P4 library in which the P1-position is held constant as leucine (Figure 4-3 L).

Profiling proteases with a P1-diverse library of 137,180 substrates. To test the possibility of attaching all amino acids to the P1-site of the substrate and the possibility of adding another position of diversity, a P1-diverse library was created. The P1-diverse library consisted of 20 sub-libraries in which only the P1-position is systematically held constant as all amino acids, excluding cysteine and including norleucine. The P2, P3, and P4 positions consist of an equimolar mixture of all amino acids. To test the ability of this library to identify the optimal P1 amino acid for a protease, a panel of well-characterized serine and cysteine proteases were run againsts the substrate library. Chymotrypsin showed the well-known preference for hydrophobic amino acids (Figure 4-4B). Trypsin and thrombin showed preference for P1-basic amino acids (Arg>Lys) (Figure 4-4 A & F). Plasmin also showed a preference for basic amino acids, however it prefered Lys over Arg (Figure 4-4 E). Tryptase also showed a preference for basic amino acids, lysine slightly more preferred than arginine (Figure 4-4G). Granzyme B, the only know mammalian serine protease to have P1-Asp specificity, showed a distinct preference for aspartic acid over all other amino acids, even the other acidic amino acid, Glu (Figure 4-4C). Fiddler crab collagenase has been characterized as having broad P1 substrate specificity (Table 1-1) and indeed that preference is manifested in the P1-diverse library (Figure 4-4D). Human neutrophil elastase has the canonical preference for alanine as well as valine (Figure 4-4 I). The cysteine proteases, papain, cruzain, cathepsin L and cathepsin B show the broad P1-substrate specificity known for these enzymes, however, there is a preference for arginine and an absence of isoleucine and proline activity (Figure 4-4 J-M).

Figure 4-1. Preparation of 7-amino-4-acetamide coumarin resin.

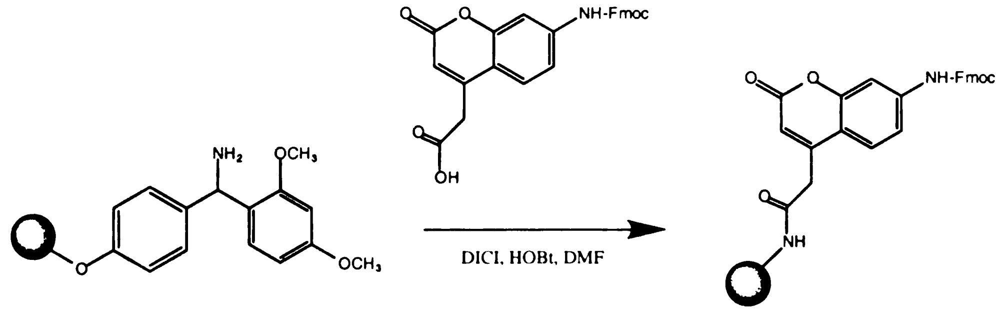
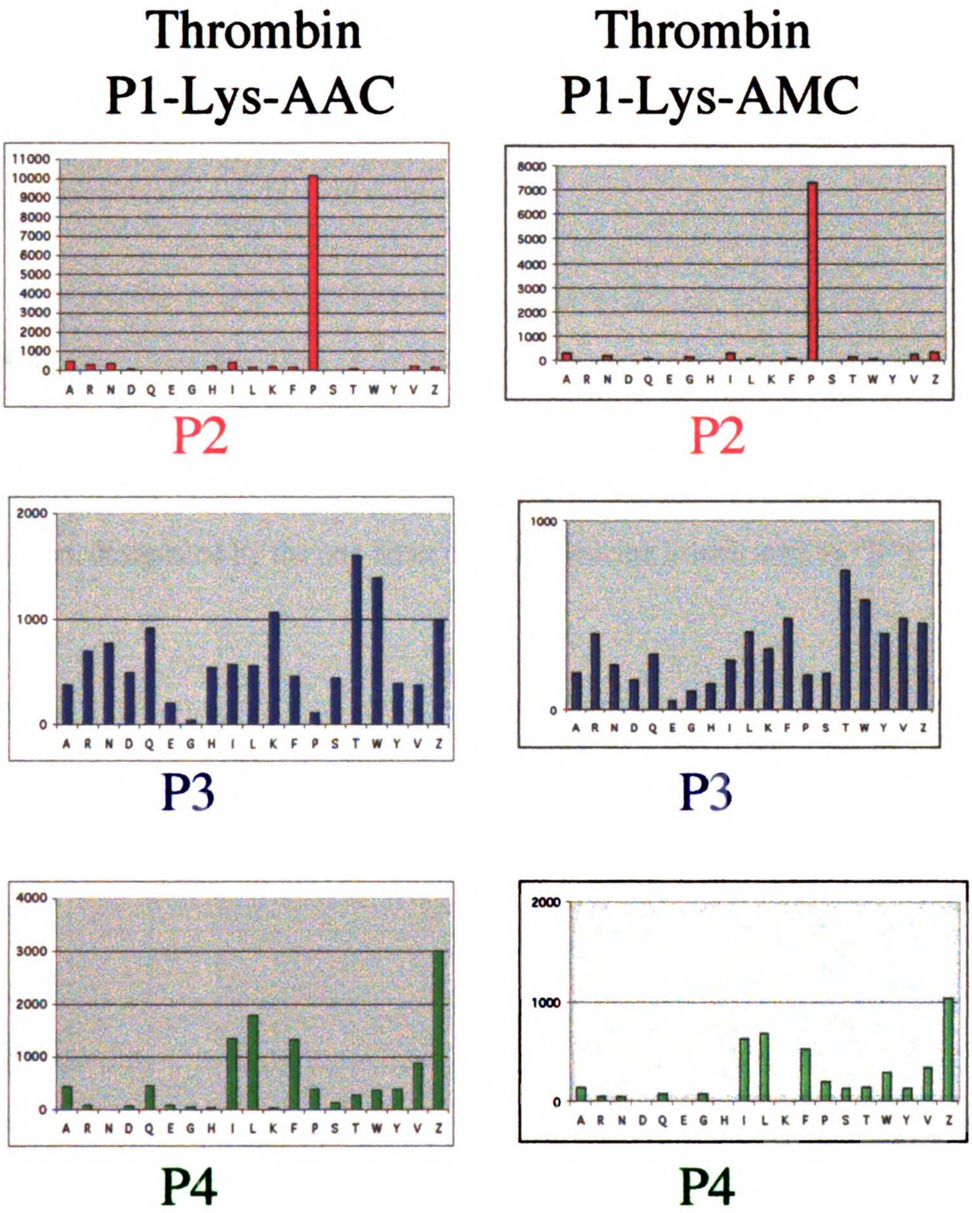

NH-Fmoc
OH
NH₂
OCH₃
OCH₃
O
DICI, HOBt, DMF
NH-Fmoc
NH

Figure 4-2.



Thrombin
P1-Lys-AAC

Thrombin
P1-Lys-AMC

Figure 4-3.

(A-L) Activity of proteases in a P1-(Arg, Lys or Leu) Positional Scanning-Synthetic Combinatorial Library. Y-axis is pM of fluorophore released per second. X-axis indicates the amino acid held constant at each position, designated by the one-letter code, (norleucine is indicated by "Z").

Figure 4-4.

(A-M) Activity of proteases in a P1-Diverse Positional Scanning-Synthetic Combinatorial Library. Y-axis is pM of fluorophore released per second. X-axis indicates the amino acid held constant at each position, designated by the one-letter code (norleucine is indicated by "Z").

Figure 4-3 A.

Human Thrombin (P1-Arg)

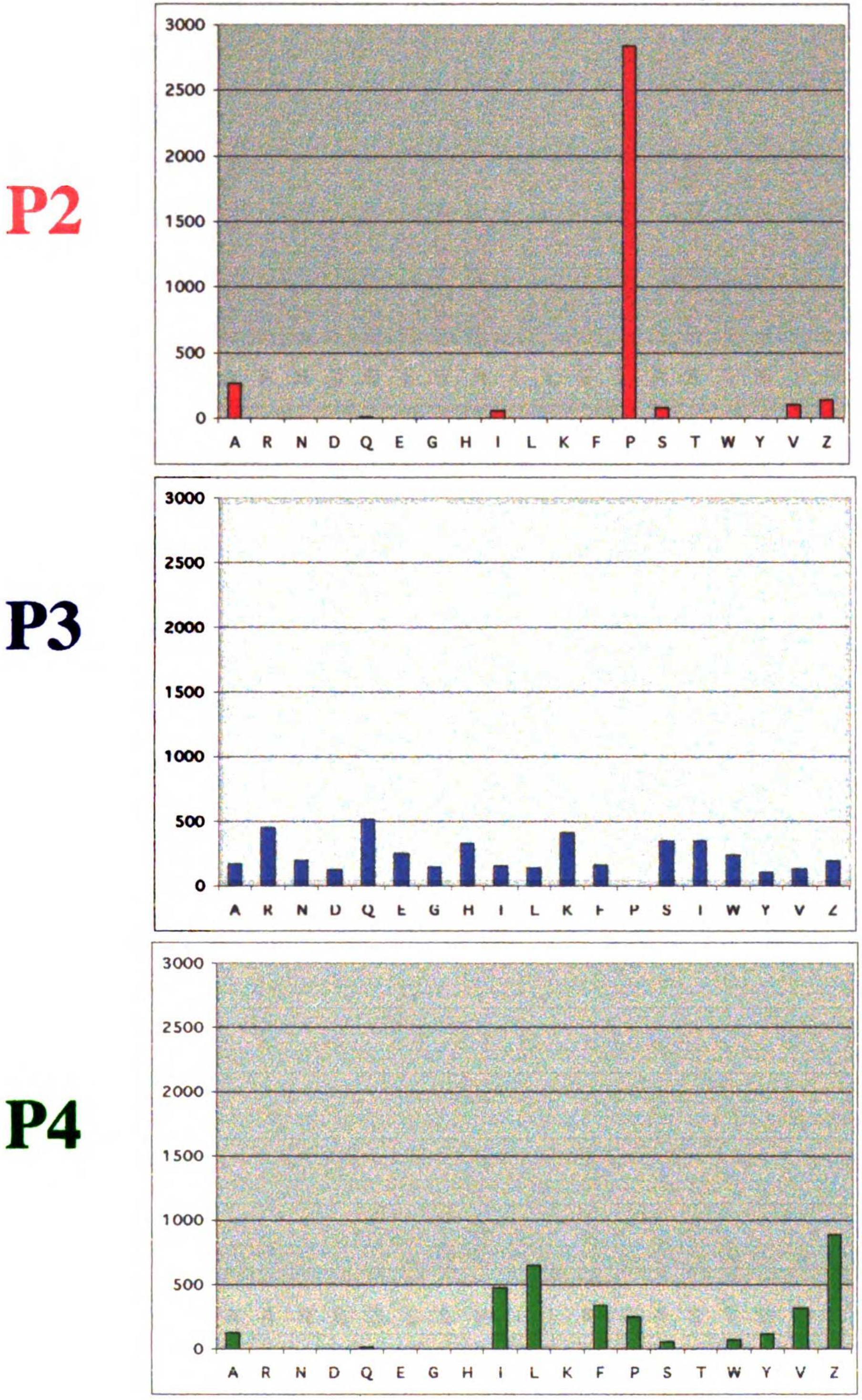
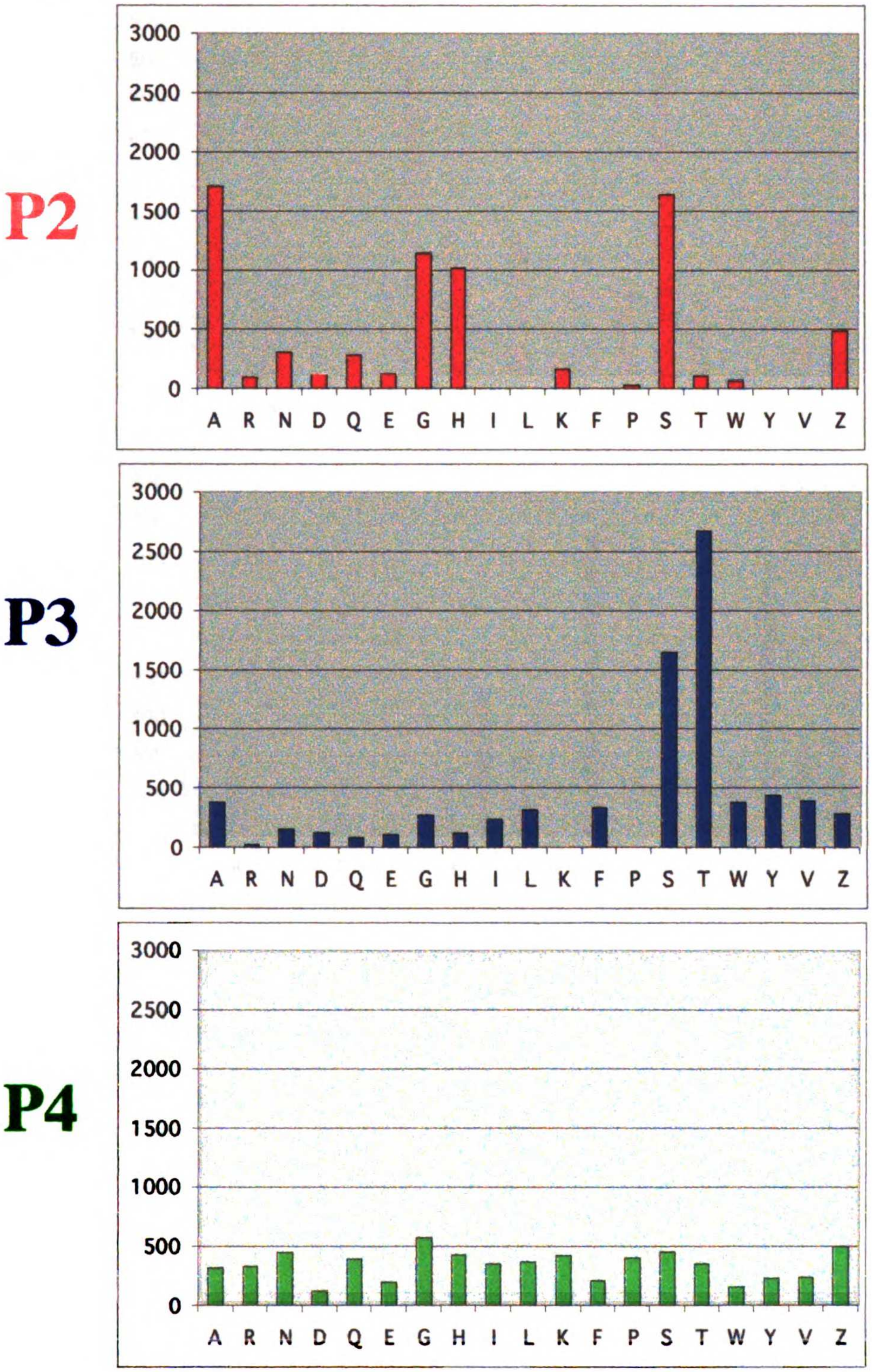


Figure 4-3 B.

Human u-PA (P1-Arg)



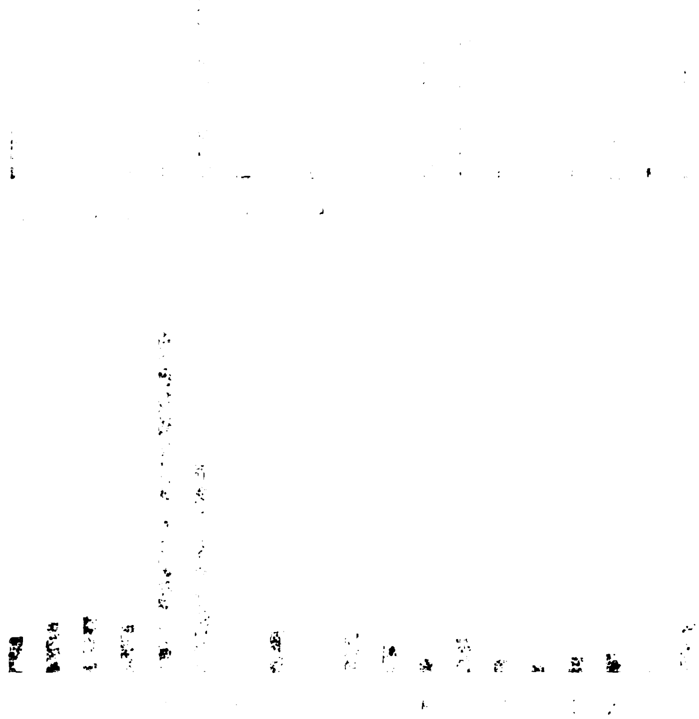
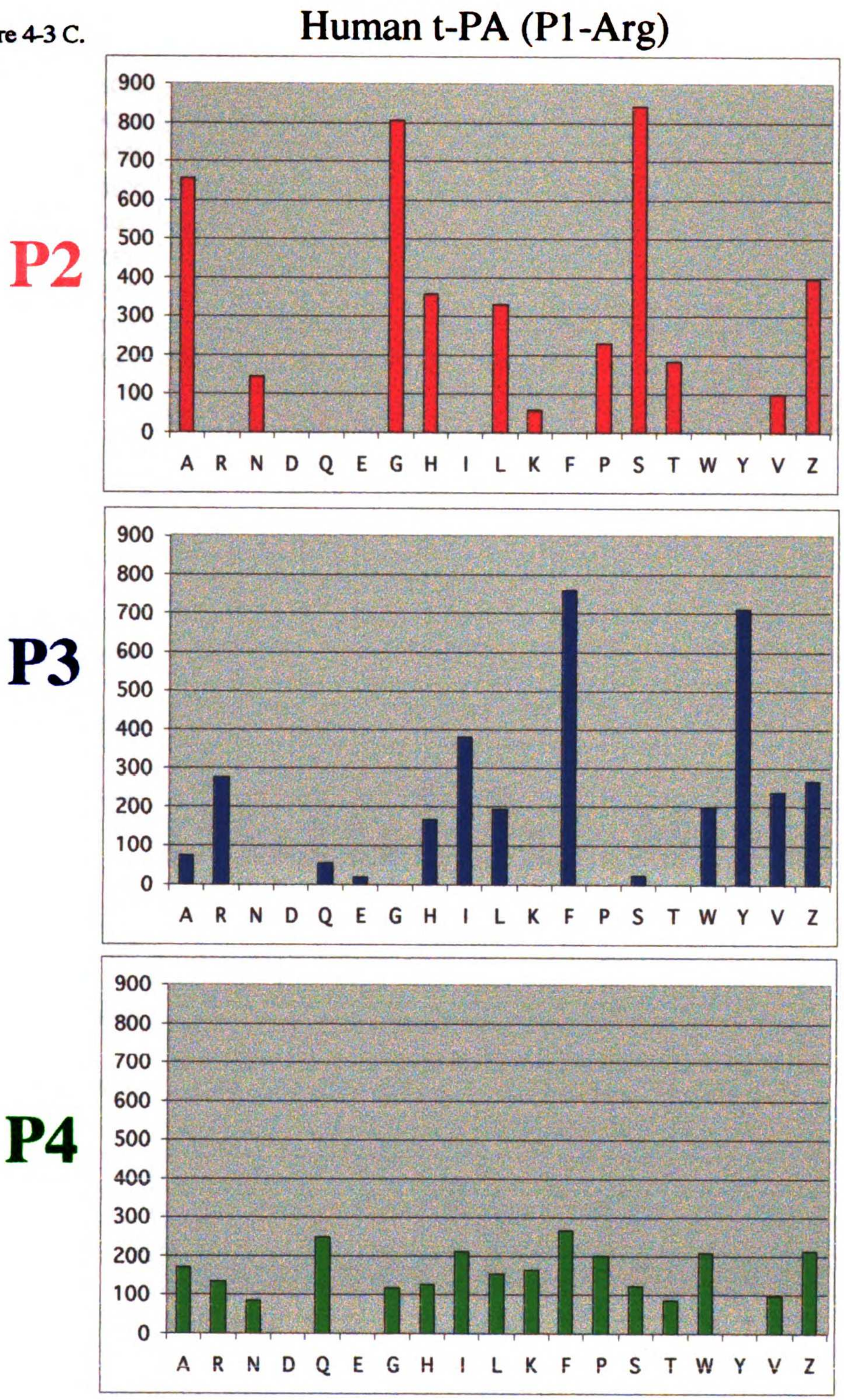


Figure 4-3 C.


Human t-PA (P1-Arg)
P2
900
800
700
600
500
400
300
200
100
0
A R N D Q E G H I L K F P S T W Y V Z
P3
900
800
700
600
500
400
300
200
100
0
A R N D Q E G H I L K F P S T W Y V Z
P4
900
800
700
600
500
400
300
200
100
0
A R N D Q E G H I L K F P S T W Y V Z

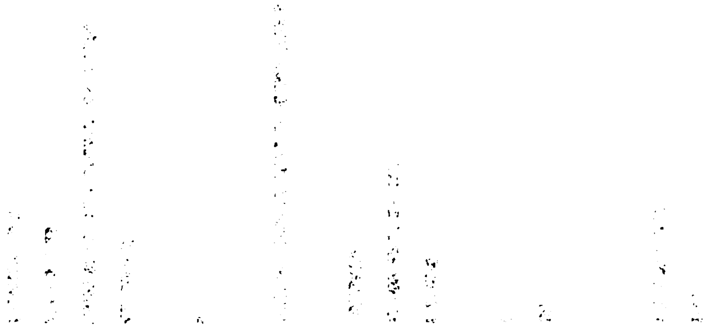


Figure 4-3 D.

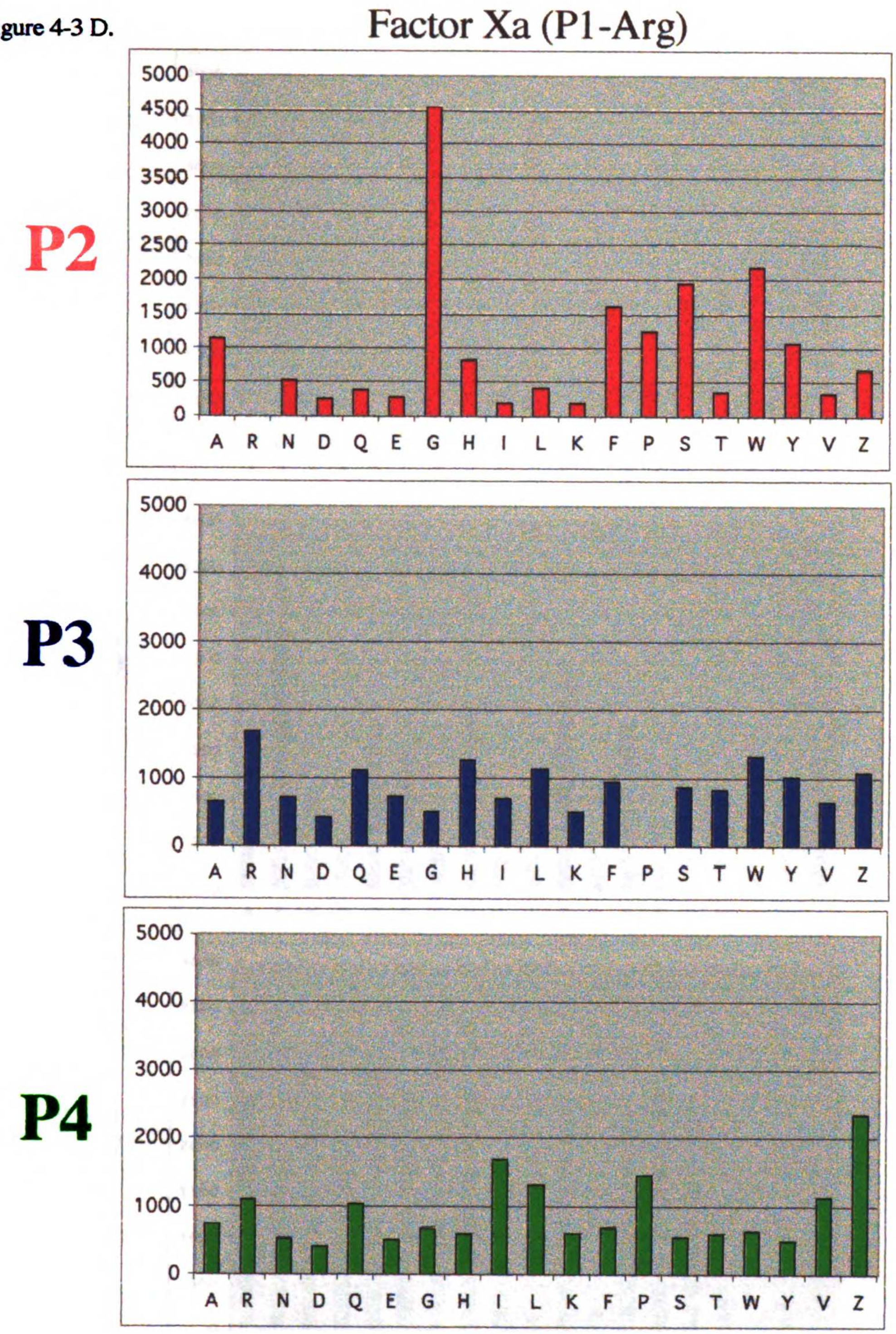
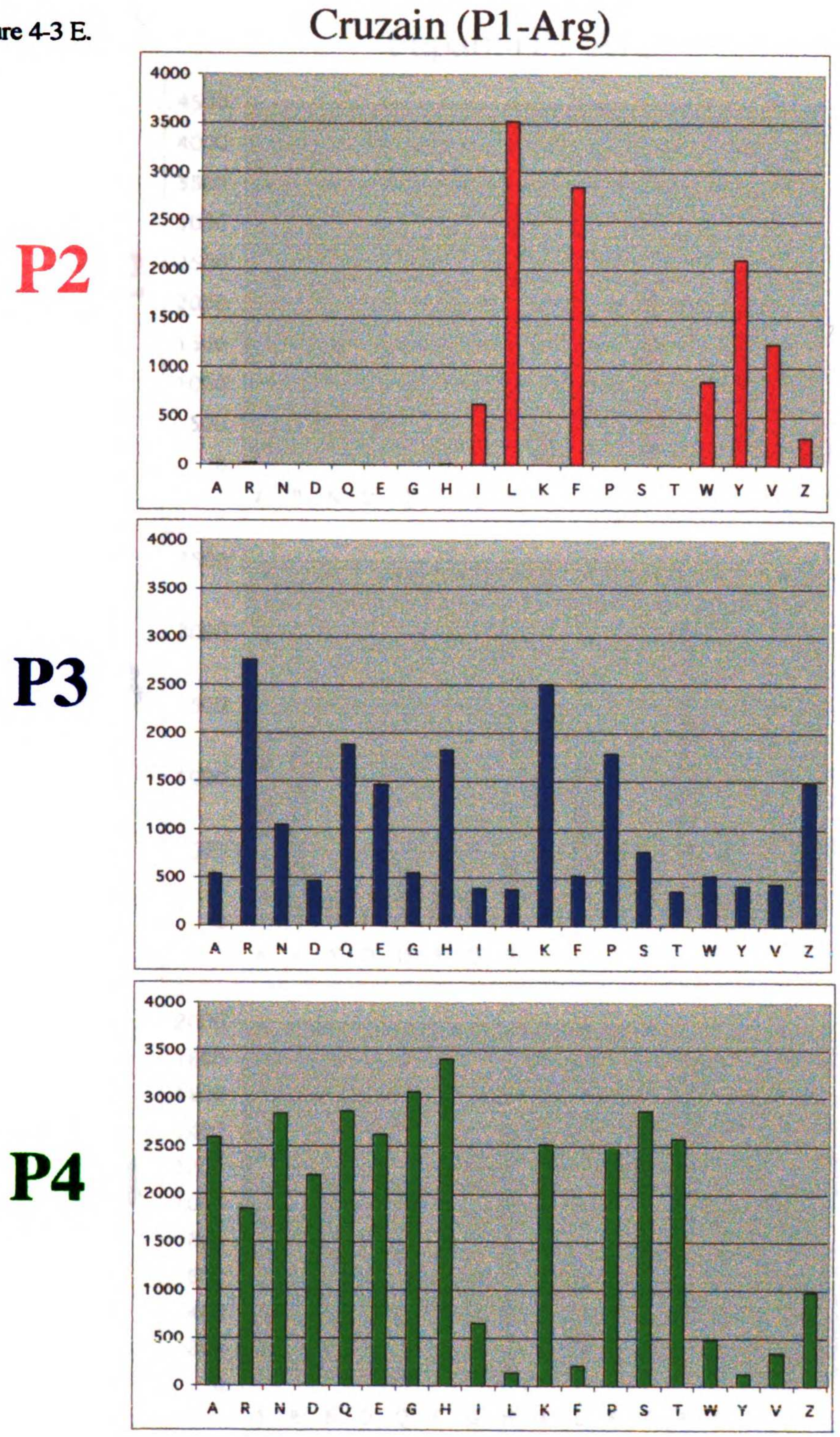

Factor Xa (P1-Arg)
P2
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
A R N D Q E G H I L K F P S T W Y V Z
P3
5000
4000
3000
2000
1000
0
A R N D Q E G H I L K F P S T W Y V Z
P4
5000
4000
3000
2000
1000
0
A R N D Q E G H I L K F P S T W Y V Z

Figure 4-3 E.


Cruzain (P1-Arg)
P2
P3
P4
4000
3500
3000
2500
2000
1500
1000
500
0
A R N D Q E G H I L K F P S T W Y V Z

[illegible]

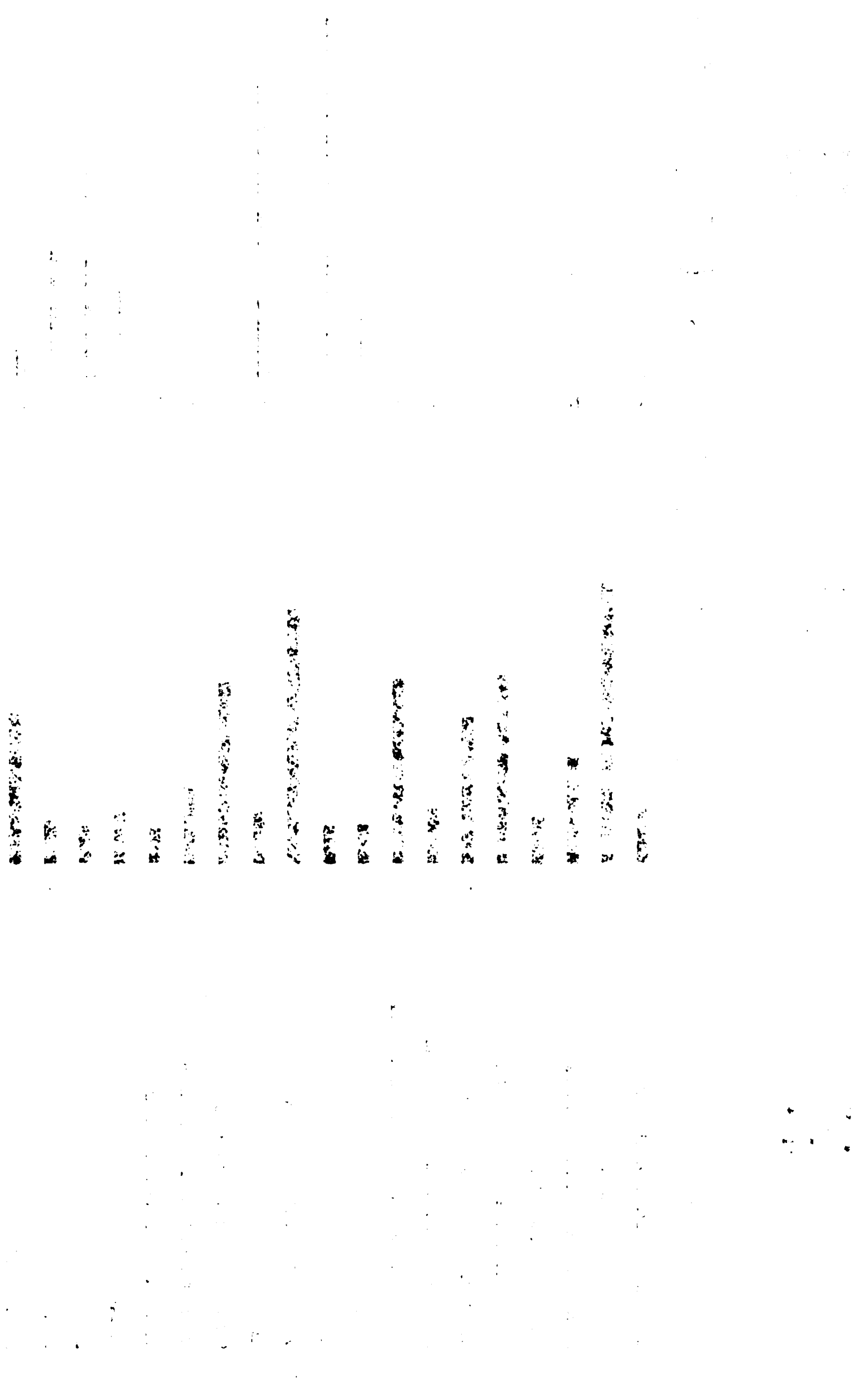
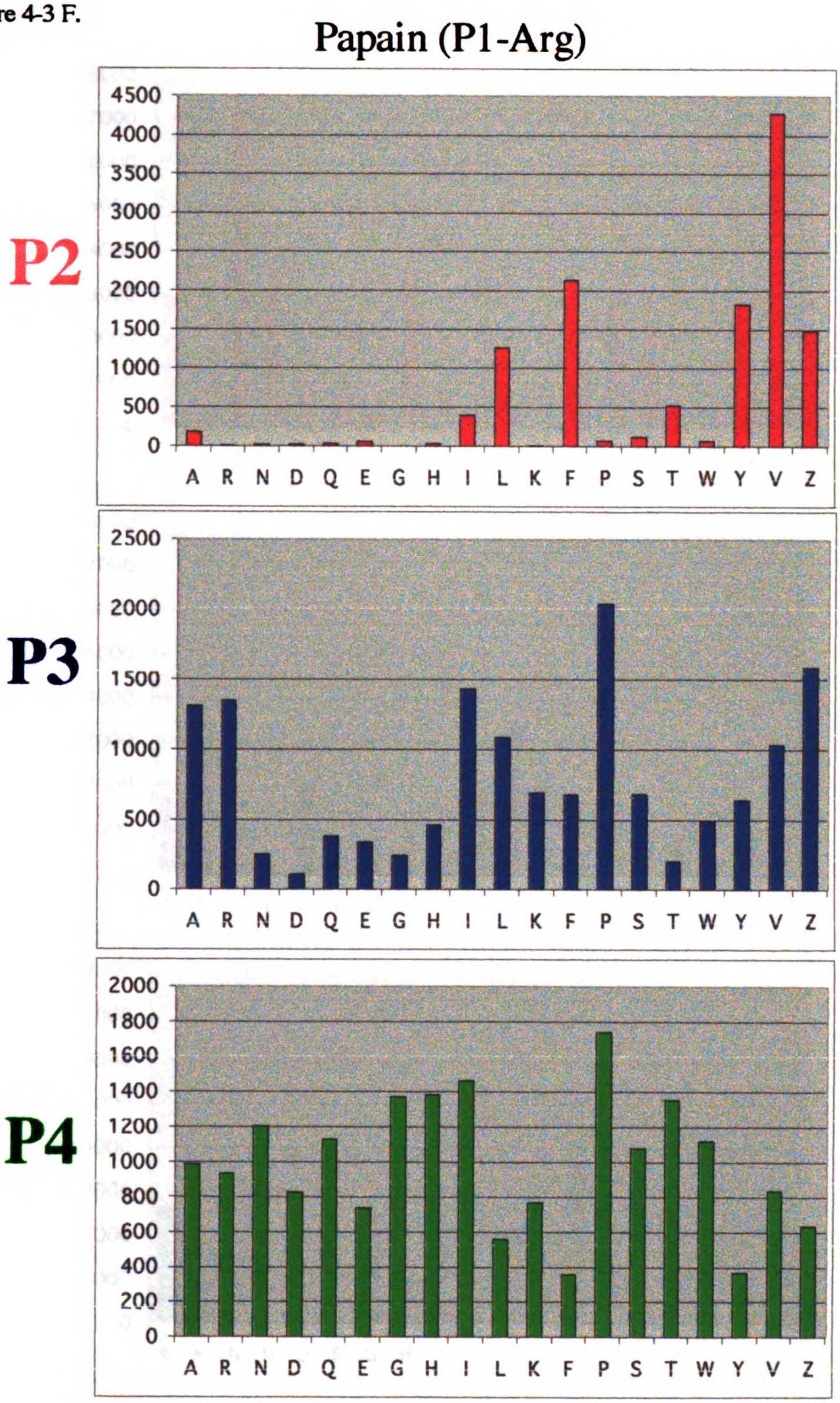


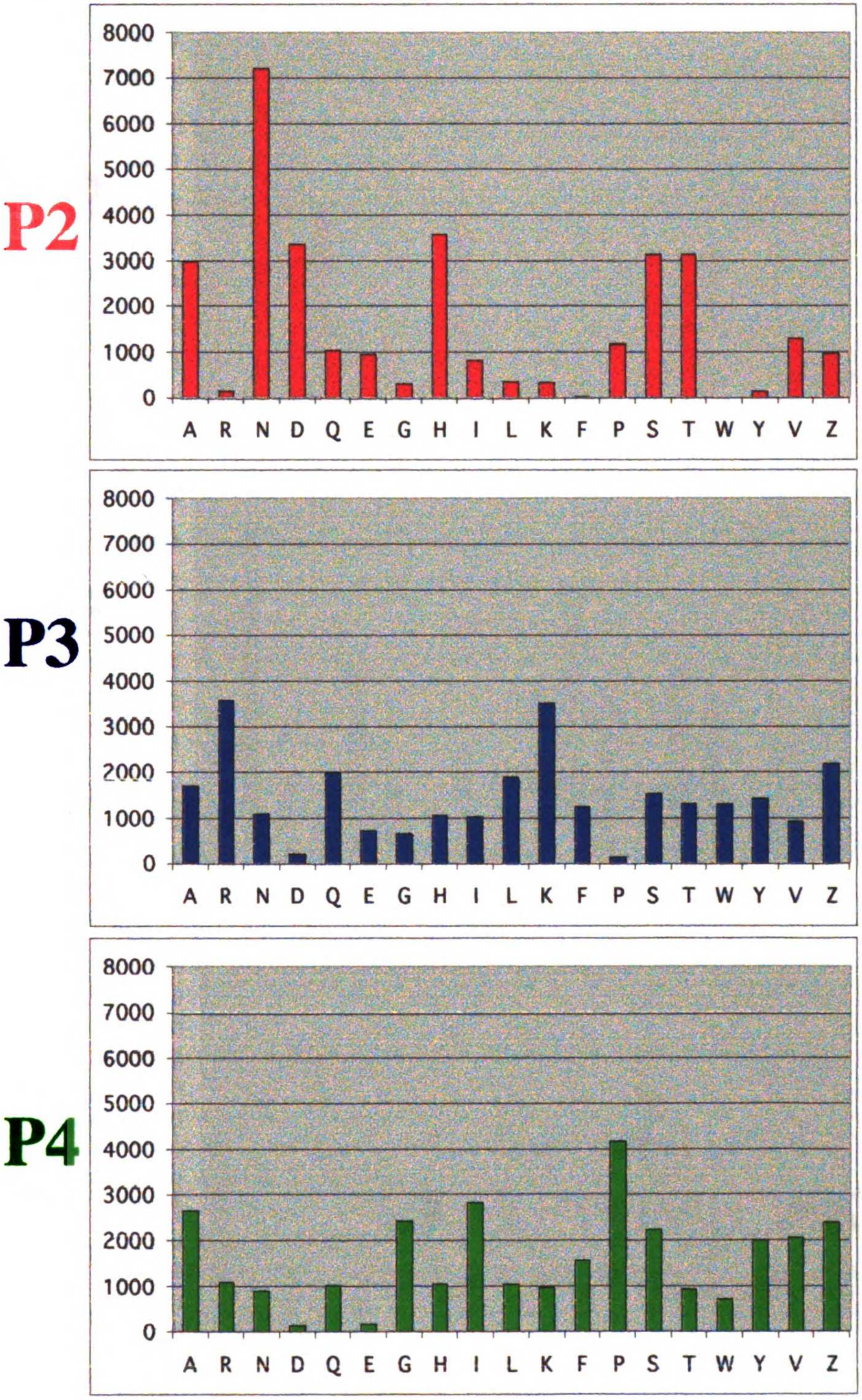
Figure 4-3 F.



Papain (P1-Arg)
P2
4500
4000
3500
3000
2500
2000
1500
1000
500
0
A R N D Q E G H I L K F P S T W Y V Z
P3
2500
2000
1500
1000
500
0
A R N D Q E G H I L K F P S T W Y V Z
P4
2000
1800
1600
1400
1200
1000
800
600
400
200
0
A R N D Q E G H I L K F P S T W Y V Z

UCSF LIBRARY

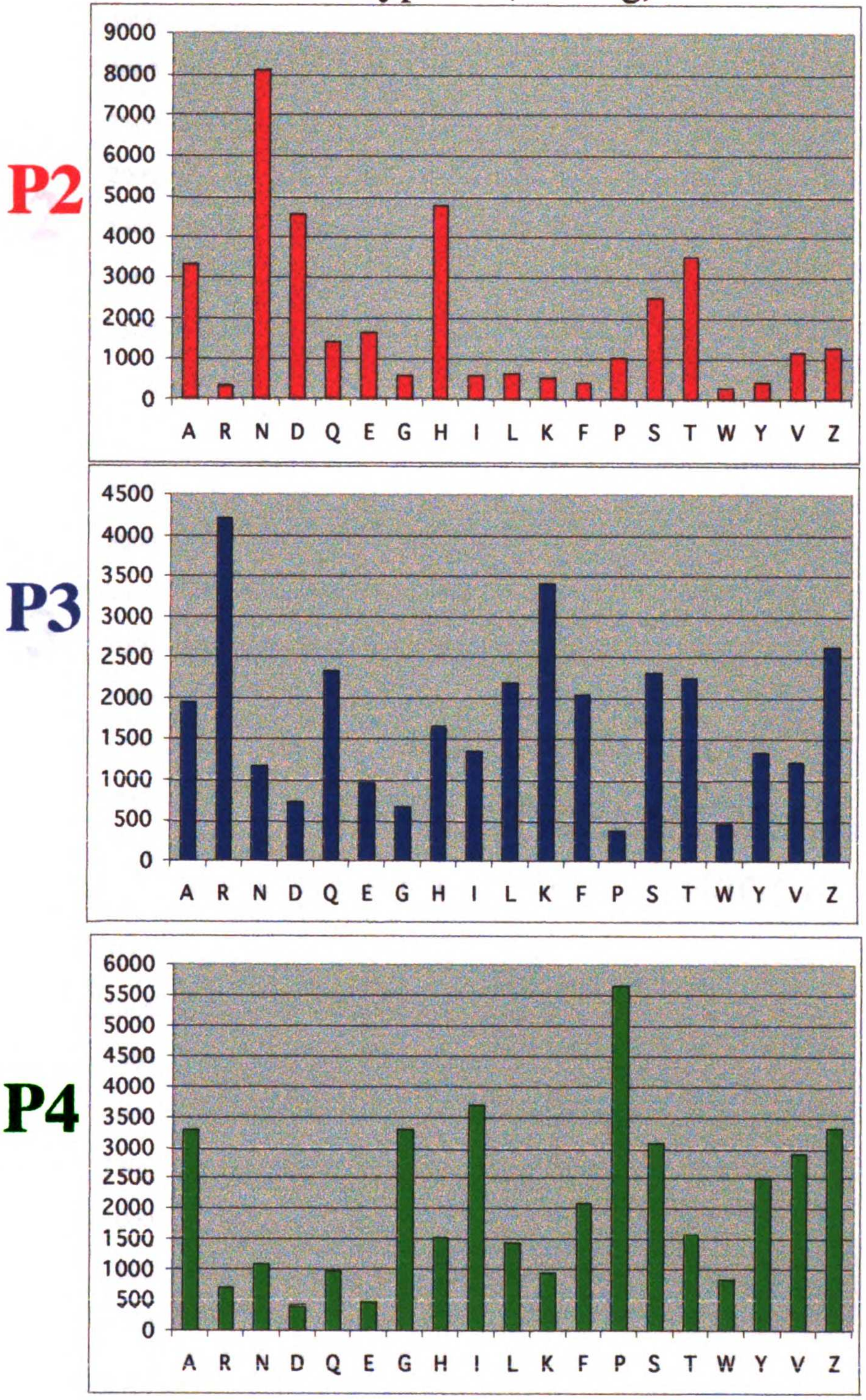
Figure 4-3 G.



Lung Tryptase (P1-Arg)

Figure 4-3 H.

Skin Tryptase (P1-Arg)

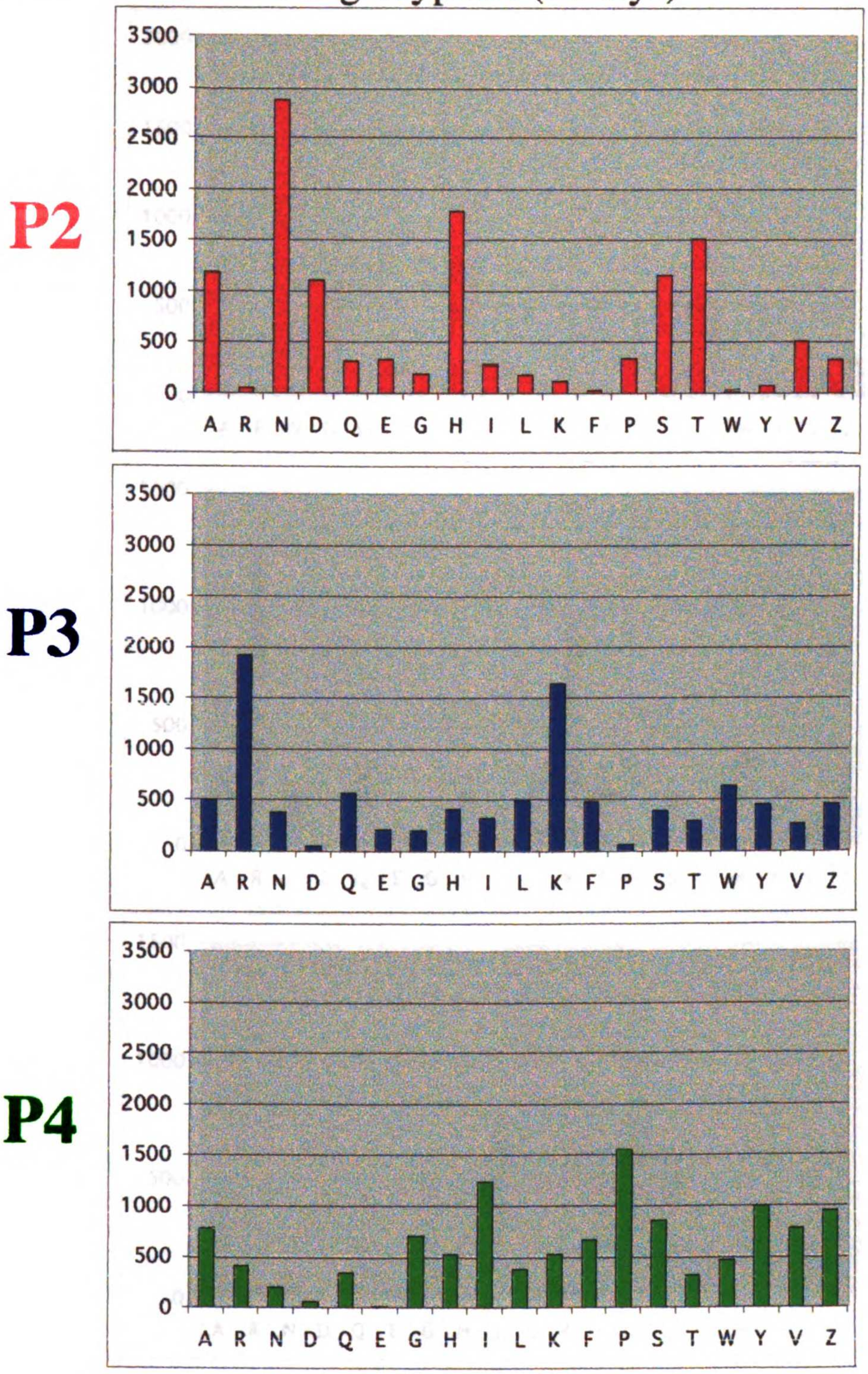


UCSF LIBRARY

[illegible]

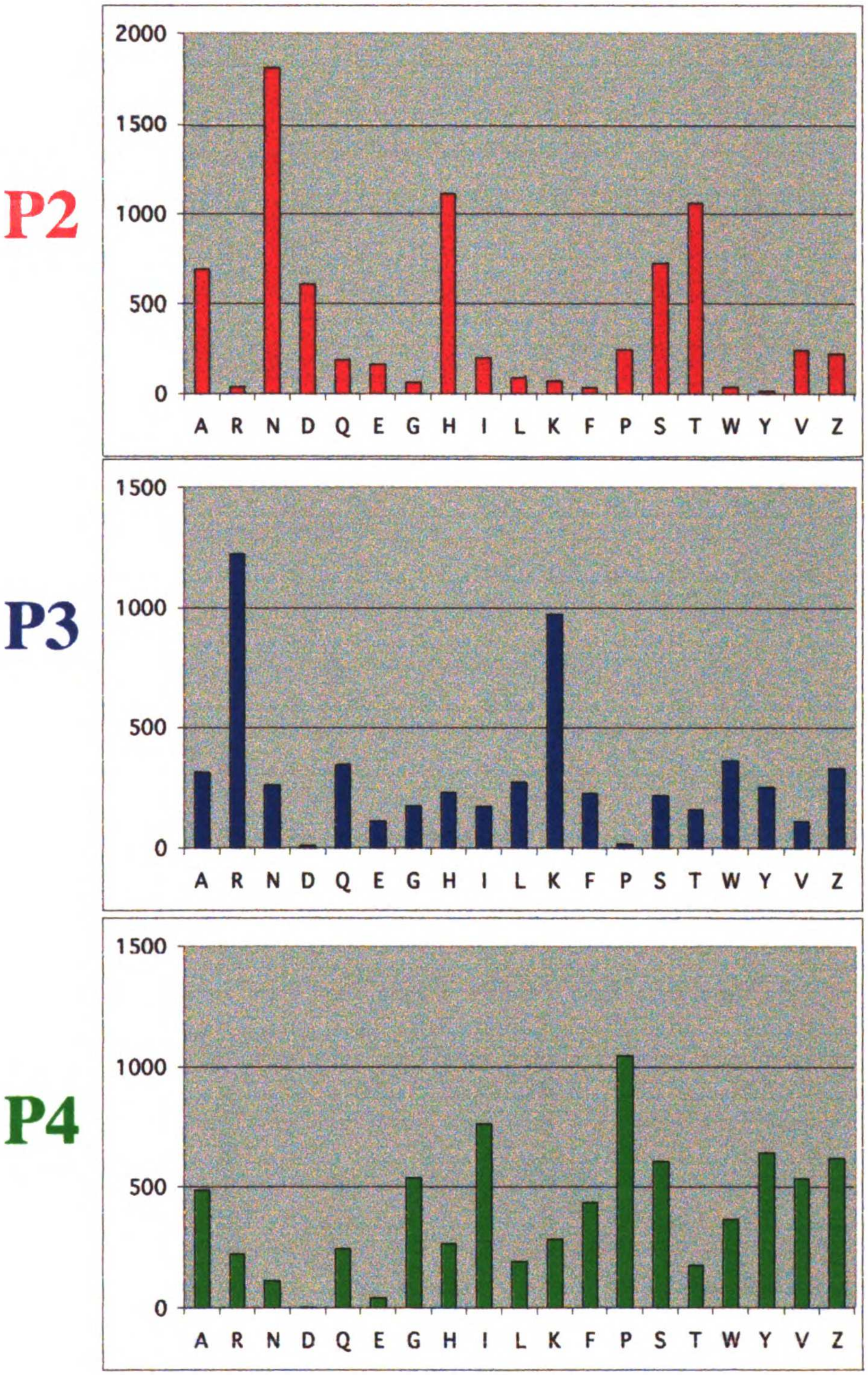
Figure 4-3I.

Lung Tryptase (P1-Lys)



UCSF LIBRARY

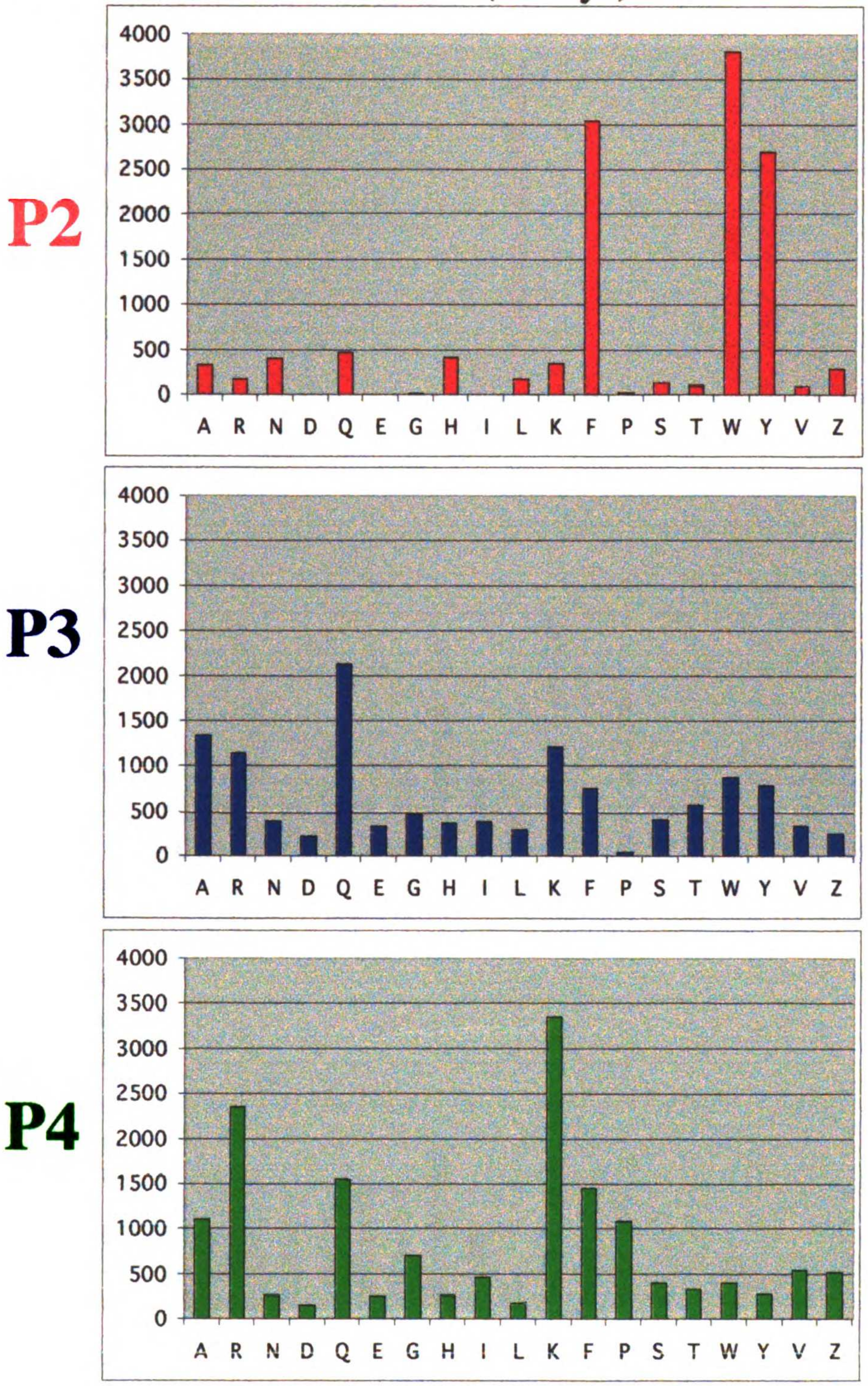
Figure 4-3 J.



Skin Tryptase (P1-Lys)

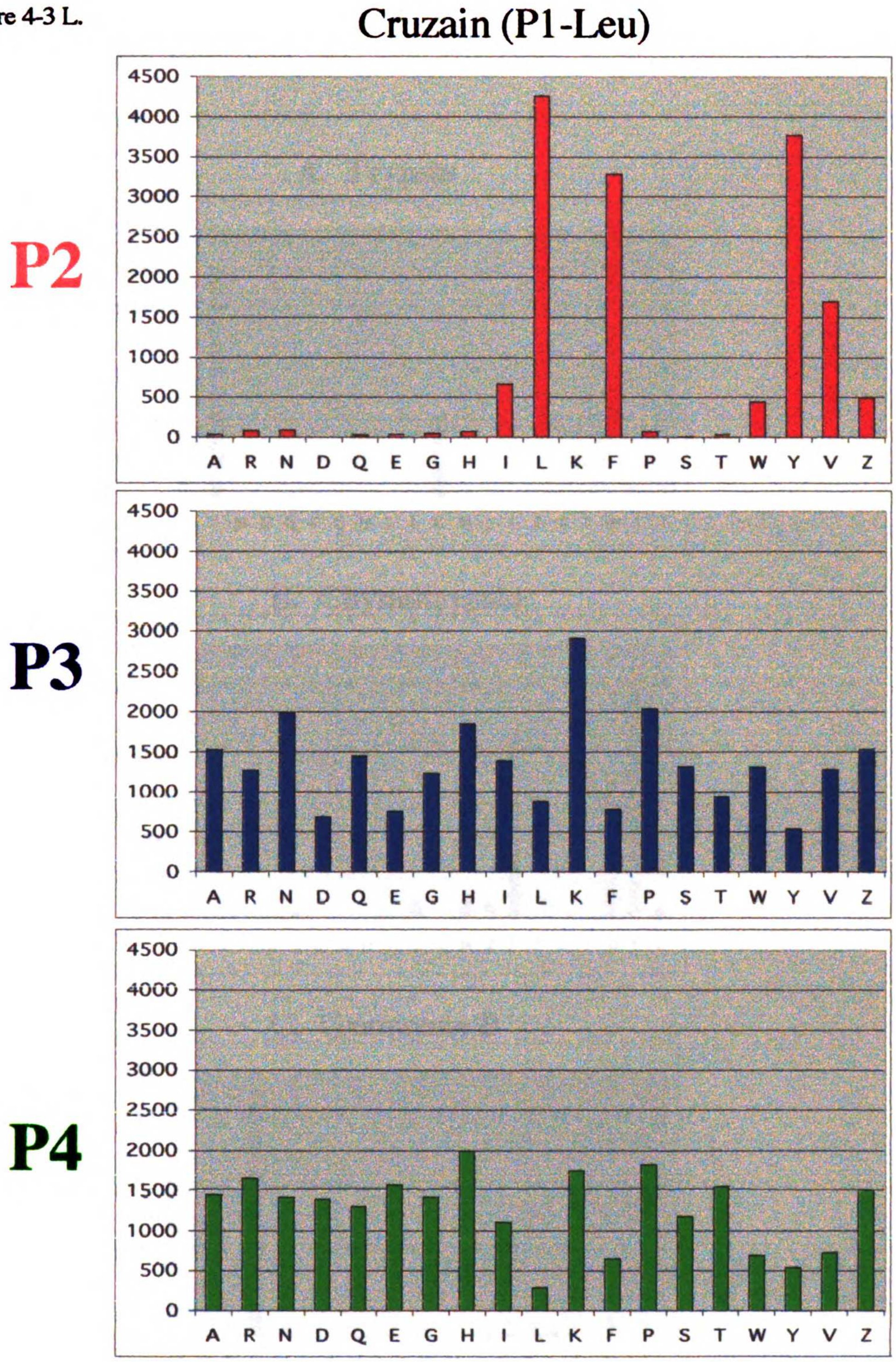
Figure 4-3 K.

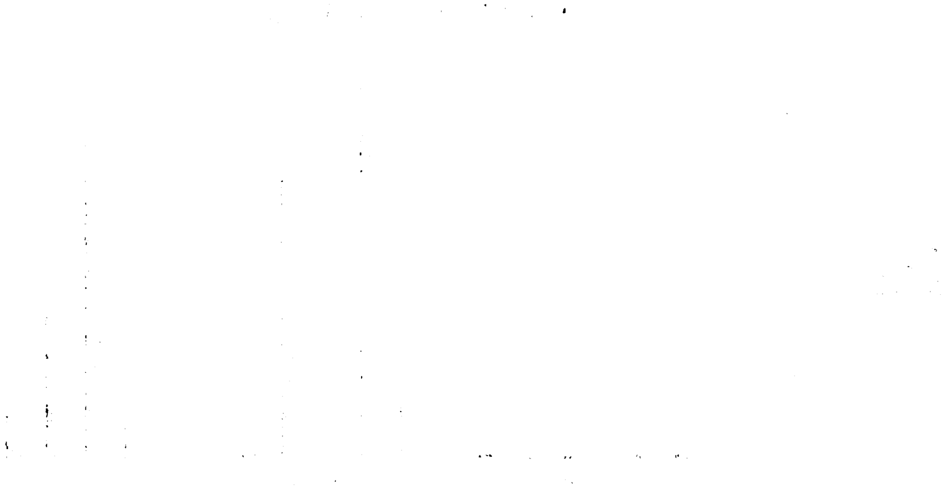
Plasmin (P1-Lys)

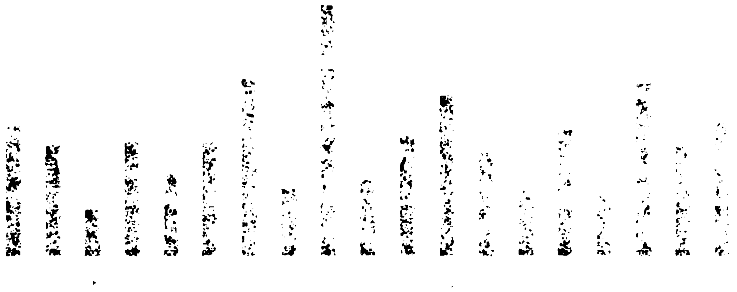


UCSF LIBRARY

Figure 4-3 L.


Cruzain (P1-Leu)
P2
P3
P4
4500
4000
3500
3000
2500
2000
1500
1000
500
0
A R N D Q E G H I L K F P S T W Y V Z





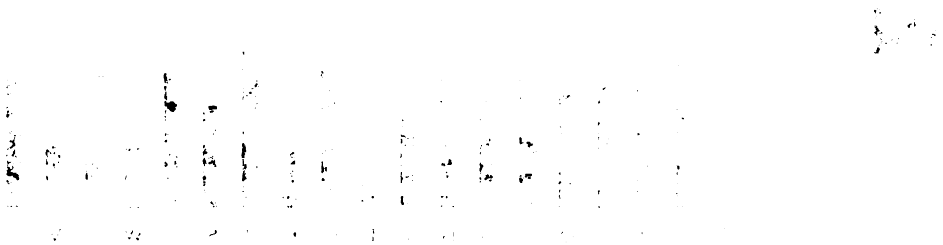
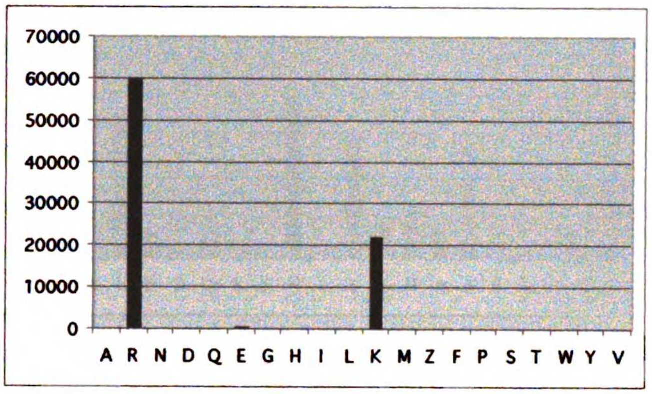
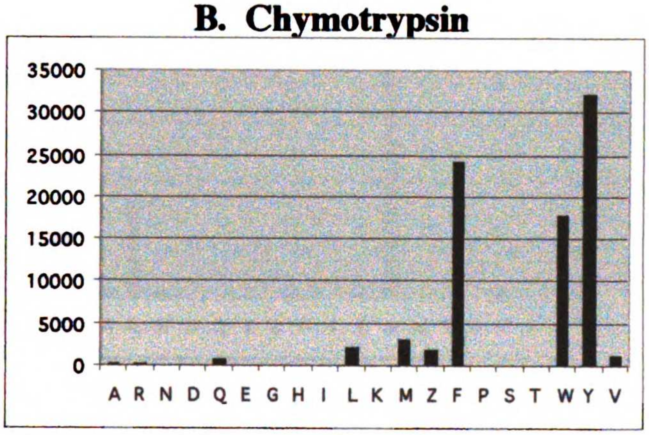


Figure 4-4.

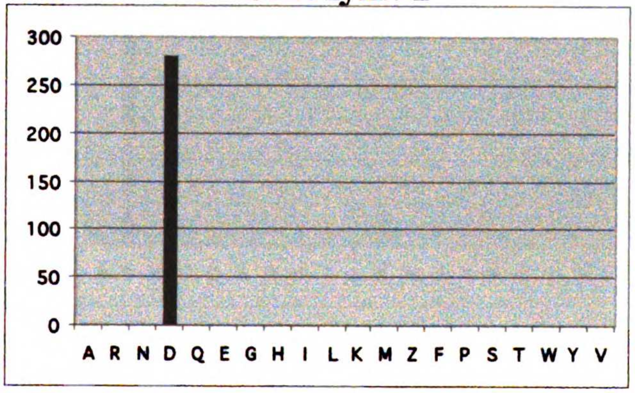
A. Trypsin



B. Chymotrypsin



C. Granzyme B



Fi

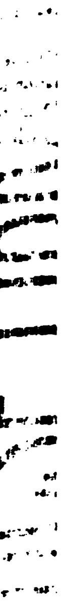
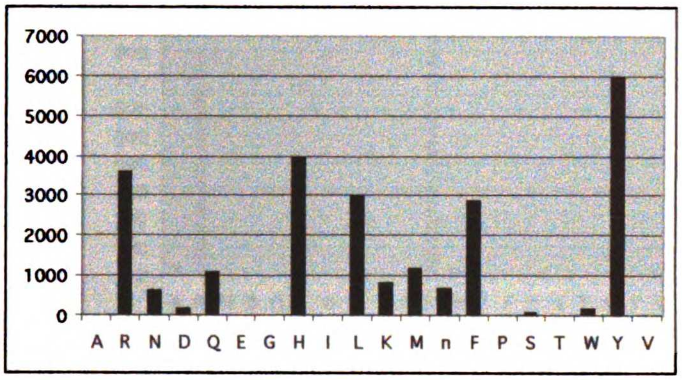
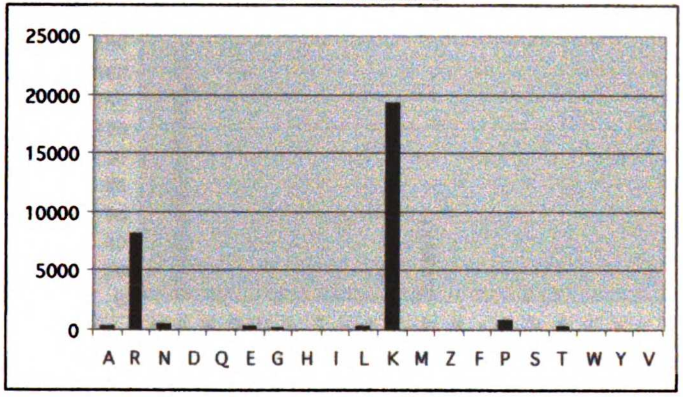


Figure 4-4 (continued).

D. Fiddler Crab Collagenase



E. Plasmin



F. Thrombin

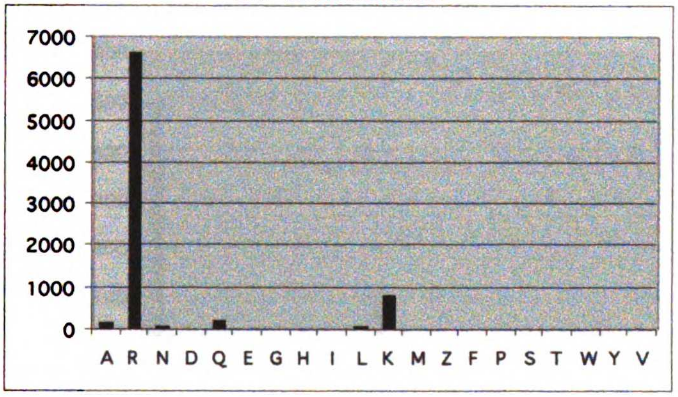
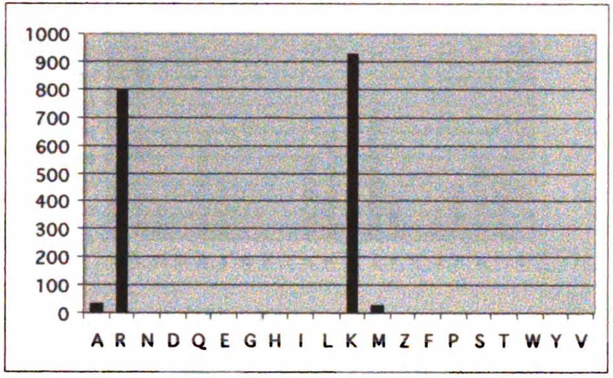
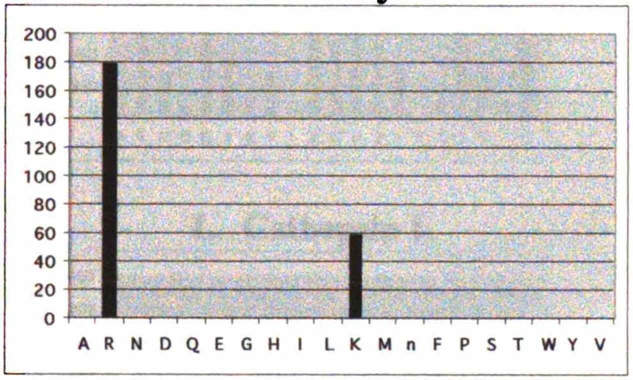


Figure 4-4 (continued).

G. Human Skin Tryptase



H. Human Granzyme A



I. Human Neutrophil Elastase

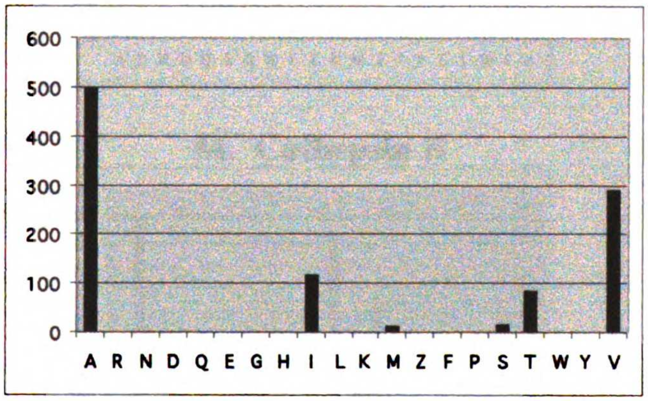
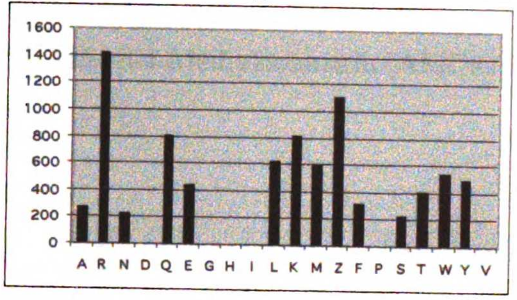
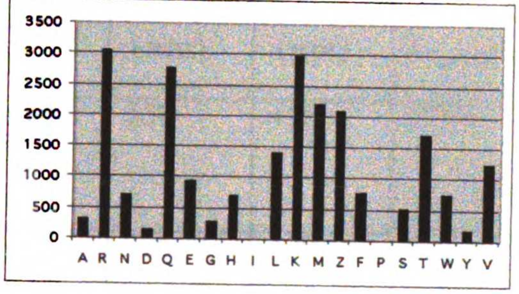


Figure 4-4 (continued).

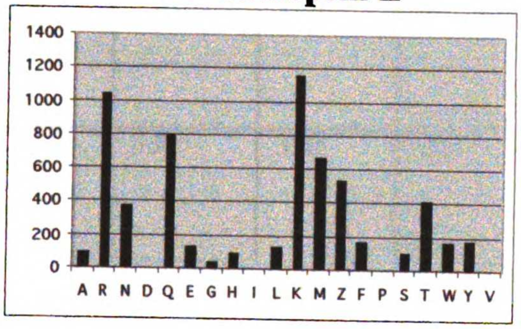
J. Papain



K. Cruzain



L. Cathepsin L



M. Cathepsin B

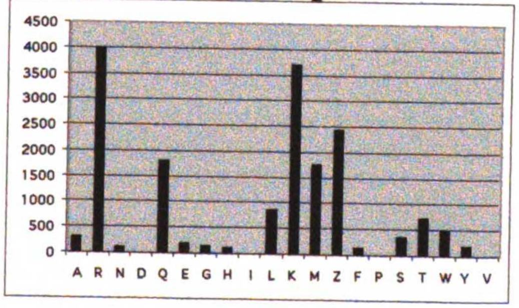


Table 4-1.

Fluorescence Properties of 7-Amino-4-Methylcoumarin (AMC) and 7-Amino-4-Acetamidecoumarin (AAC).

Compound	$\lambda_{max,\ Ex}$ (nm)	$\lambda_{max,\ Em}$ (nm)	**RFU/nM** λ_{ex} = 380 nm λ_{em} = 450 nm	**RFU/nM** λ_{ex} = 380 nm λ_{em} = 470 nm
7-Amino-4-Acetamidecoumarin (AAC)	350	450	5750	4390
7-Lys-Thr-Ser-Lys-AAC	325	400	6.6	4.8
7-Amino-4-Methylcoumarin (AMC)	340	440	2600	1550
7-Lys-Thr-Ser-Lys-AMC	330	390	3.6	2.6

Table 4-2.

Steady state kinetic constants, k_{cat} , K_m , and k_{cat}/K_m for amino-acetamide coumarin substrates for lung and skin tryptase.

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m x 10^6 (s^{-1} M^{-1})
		Lung Tryptase	
Ac-PRNK-AAC	16.84 ± 0.27	8.9 ± 0.9	1.89 ± 0.17
Ac-PANK-AAC	20.27 ± 0.48	110.5 ± 9.8	0.18 ± 0.01
Ac-PRTK-AAC	18.67 ± 0.30	14.7 ± 1.4	1.27 ± 0.12
Ac-PRNR-AAC	21.75 ± 0.67	16.5 ± 2.7	1.31 ± 0.19
		Skin Tryptase	
Ac-PRNK-AAC	17.84 ± 0.40	14.5 ± 1.9	1.23 ± 0.15
Ac-PANK-AAC	19.06 ± 0.64	133.3 ± 15.6	0.14 ± 0.01
Ac-PRTK-AAC	18.34 ± 0.33	23.4 ± 2.3	0.78 ± 0.07
Ac-PRNR-AAC	20.94 ± 0.57	18.6 ± 2.6	1.12 ± 0.14

Chapter Five:

Definition and Redesign of the extended substrate specificity of granzyme B

Abstract:

Granzyme B is a protease involved in the induction of rapid target cell death by cytotoxic lymphocytes. Definition of the substrate specificity of granzyme B allows for the identification of *in vivo* substrates in this process. Using the combinatorial methods of synthetic substrate libraries and substrate-phage display, an optimal substrate for granzyme B that spans over six subsites was determined to be Ile-Glu-Xxx-Asp/Xxx-Gly, with cleavage of the Asp/Xxx peptide bond. Granzyme B proteolysis was shown to be highly dependent on the length and sequence of the substrate, supporting the role of granzyme B as a regulatory protease. Arginine 192 was identified as a determinant of P3-Glu and P1-Asp substrate specificity. Mutagenesis of arginine 192 to glutamate reversed the preference for negatively charged amino acids at P3 to positively charged amino acids. The preferred substrate sequence matches the activation sites of caspase 3 and caspase 7 and thus is consistent with the role of granzyme B in activation of these proteases during apoptosis. The caspase substrate poly (ADP)-ribose polymerase, is cleaved by granzyme B in a cell free assay at two sites that resemble the granzyme B specificity determined by the combinatorial methods. Many caspase substrates contain granzyme B cleavage sites and are proposed as potential granzyme B targets, suggesting a redundant function with certain caspases. (This work was published in Harris, J.L., Peterson, E.P., Hudig, D., Thornberry, N.A. and Craik, C.S. (1998) Definition and redesign of the extended substrate specificity of granzyme B. *J Biol Chem*, **273**, 27364-27373.)

Abbreviations:

AMC, 7-amino-4-methyl coumarin; pNA, *p*-nitro-anilide; SBzl, thiobenzyl ester; Ac-IEPD-AMC, N-acetyl-isoleucyl-glutamyl-prolyl-aspartyl-AMC; Ac-IEPD-pNA, N-acetyl-isoleucyl-glutamyl-prolyl-aspartyl-pNA; Ac-IKPD-pNA, N-acetyl-isoleucyl-lysyl-prolyl-aspartyl-pNA; Suc-AAPX-pNA, N-succinyl-alanyl-alanyl-prolyl-Xxx-pNA (Xxx= alanyl, aspartyl, glutamyl, phenylalanyl, leucinyl, methionyl, and arginyl; Ac-EPD-pNA, N-acetyl-glutamyl-prolyl-aspartyl-pNA; Ac-PD-pNA, N-acetyl-prolyl-aspartyl-pNA; Ac-IEPD-SBzl, N-acetyl-isoleucyl-glutamyl-prolyl-aspartyl-SBzl; Boc-AAD-Sbzl, t-Butyloxycarbonyl-alanyl-alanyl-aspartyl-SBzl; Ac-IEPDW(G or N)A-NH_2, N-acetyl--isoleucyl-glutamyl-prolyl-aspartyl-tryptophanyl-(glycyl or asparaginyl)-alanyl-amide; Z-DEVD-FMK, Benzyloxycarbonyl-aspartyl-glutamyl-valyl-aspartyl-fluoromethyl ketone; DTDP, 4,4'-dithiodipyridine; MES, 2-(N-morpholino)ethanesulfonic acid; Tris, tris(hydroxymethyl) aminomethane; MUGB, 4-methylumbelliferyl *p*-guanidinobenzoate; HEPES, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid); CL, cytotoxic lymphocyte; PS-SCL, positional scanning synthetic-combinatorial library; n, norleucine; d-A, d-alanine; PARP, poly(ADP-ribose) polymerase.

Introduction:

Cytotoxic lymphocytes (CLs), which include cytotoxic T-cells and natural killer cells, are one of the most potent host defenses against tumor and virus-infected cells. Upon recognition by the CLs, apoptosis of the target cell is initiated by the granule-exocytosis mechanism and the Fas-FasL mechanism (Podack, 1995). Both pathways result in the activation of intracellular proteolysis which mediates apoptotic death of the target cell. The granule-exocytosis model proposes that granules are released from the CLs after recognition of the target cell. The major components of the granules are perforin, a putative-pore forming protein, and the granzymes, a subclass of serine proteases displaying the chymotrypsin fold (Smyth *et al.*, 1996). Perforin is believed to facilitate entry and localization of the granzymes within the target cell to effect death (Froelich *et al.*, 1996b; Shi *et al.*, 1997). Multiple granzymes have been identified and cloned from the granules of cytotoxic lymphocytes. While granzymes are strongly implicated as key mediators of cell death (Heusel *et al.*, 1994; Hudig *et al.*, 1987), the mechanisms by which individual granzymes carry out their functions have not been fully elucidated. For example, it is unclear why multiple granzymes are used and if the granzymes mediate cell death through unrestricted proteolysis of the target cell or through a more selective cleavage of target cell proteins (Williams and Henkart, 1994).

Recent studies have focused on granzyme B because of its unusual preference for cleaving after aspartate residues (Poe *et al.*, 1991). While granzyme B is the only known mammalian serine protease to have this P1-proteolytic specificity, it is shared with the caspases, a family of cysteine proteases which are also activated during apoptosis (Nicholson and Thornberry, 1997). The link between granzyme B and the caspases has been strengthened by studies indicating that granzyme B can cleave and activate certain members of the caspases (Darmon *et al.*, 1995), and it has been suggested that this is one of the mechanisms by which granzyme B mediates apoptosis *in vivo* (Froelich *et al.*, 1996b).

Participation of granzyme B in apoptosis may also involve a complementary activity with the caspases through the direct cleavage of proteins that are the proteolytic targets of caspases (Andrade *et al.*, 1998). Poly (ADP-ribose) polymerase (PARP) is a nuclear enzyme that is thought to function in numerous nuclear processes such as DNA repair and transcription (Oei *et al.*, 1997). Cleavage of PARP by caspase 3 occurs rapidly upon induction of apoptosis (Lazebnik *et al.*, 1994). *In vitro* cleavage of PARP by granzyme B suggests a cleavage site distinct from that of caspase 3 (Froelich *et al.*, 1996a).

To better understand the mechanism by which granzyme B mediates apoptosis of target cells, we have undertaken the identification of its extended substrate specificity. Although individual granzymes have been isolated from different species, the possibility of contamination from homologous granzymes also present in the granules has hindered substrate specificity studies. We therefore developed a recombinant expression system in the yeast *Pichia pastoris* to produce reagent quantities of pure, catalytically active rat granzyme B. In this study, we have used two combinatorial methods to extend the definition of the substrate specificity of granzyme B to six subsites, from P4 to P2'. Individual amino acids responsible for determining the stringent substrate specificity of granzyme B were identified through the construction of a three-dimensional model of granzyme B complexed to substrate. Variant granzyme B enzymes with altered P1 and P3 substrate recognition properties were created to define the molecular determinants of specificity. The elucidated substrate specificity was shown to be relevant within a macromolecular context by locating cleavage sites in defined molecular targets.

Experimental Methods

Materials.

Ac-IEPD-AMC, Ac-IKPD-AMC, Ac-IEPD-pNA, Ac-EPD-pNA, Ac-PD-pNA, Ac-D-pNA, Ac-IKPD-pNA, Ac-IEPD-SBzl, IKPD-SBzl, Ac-IEPDWGA-NH_2 and Ac-IEPDWNA-NH_2 were purchased from SynPep (Dublin, CA). Suc-AAPX-pNA (X=A,D,E,F,L,M,R) substrates and the Z-DEVD-FMK inhibitor were purchased from

Bachem Bioscience Incorporated (Torrance, CA). Boc-AAD-SBzl was purchased from Enzyme Systems Products (Livermore, CA). N-Glycosidase F was purchased from Boehringer Mannheim (Indianapolis, IN). All DNA modifying enzymes were purchased from New England Biolabs (Beverly, MA) or Stratagene (La Jolla, CA) and were used according to manufacturer's guidelines. TNT® coupled reticulocyte lysate system was purchased from Promega (Madison, WI) and used according to manufacturer's guidelines. Protein assay Bradford reagent was purchased from Bio-rad (Hercules, CA) and used according to manufacturer's guidelines. Substrates in the positional scanning synthetic combinatorial library (PS-SCL) were prepared as previously described (Rano *et al.*, 1997). Oligonucleotides were synthesized with an Applied Biosystems 391 DNA synthesizer (Foster City, CA). The *Pichia pastoris* expression system was purchased from Invitrogen (San Diego, CA). Recombinant rat granzyme B antiserum was produced and purchased from Berkeley Antibody Company (Richmond, CA).

Heterologous expression of rat granzyme B in yeast.

The Xho I and Not I DNA recognition sites were introduced by PCR onto the 5' and 3' ends, respectively, of the 681base-pair cDNA encoding mature rat granzyme B, amino acids 16 to 245 (Zunino *et al.*, 1990). The resulting fragment was subcloned into the Xho I and Not I sites of the yeast vector, pPICZαA (Invitrogen, San Diego, CA). This construct permitted the fusion of the mature granzyme B sequence to immediately follow the Kex2 signal cleavage site of the *Saccharomyces cerevisiae* α-factor secretion signal (Brake *et al.*, 1984). The vector was linearized with Sac I and transformed into the X33 strain of *Pichia pastoris*. Clones with the integrated rat granzyme B cDNA were selected for by resistance to Zeocin™ (Wenzel *et al.*, 1992). A granzyme B-expressing clone was isolated and used for large scale protein expression in a 12 liter B. Braun Biostat-E fermentor (Allentown, PA). Yeast growth and protein expression were maintained at pH 6.0, 30°C, and dissolved oxygen ≥ 20%. Ten milliliter aliquots were harvested at 24, 36, 48, 60, and 72 hours after induction with methanol and the cell density was determined by wet cell separation and weight determination. The total protein concentration in the culture supernatant was determined by Bradford analysis (Bradford, 1976) using the Bio-rad protein assay reagent (Livermore, CA) and bovine serum albumin as the protein standard. The concentration of granzyme B in the culture supernatant was determined at the time points by both active site titration and V_{max} measurements of granzyme B.

Purification of recombinant granzyme B.

After 48 hours of induction with methanol, the supernatant from the granzyme B expressing culture was harvested. The supernatant was adjusted to 50 mM NaCl and loaded onto a 100 mL SP-sepharose cation exchange column (Pharmacia Biotech, Uppsala Sweden). The column was washed with five column volumes of 50 mM MES, pH 6.0 and 50 mM NaCl and eluted with a linear salt gradient of 50 mM to 1000 mM NaCl. Active granzyme B eluted at 600 mM NaCl and was approximately 10% pure based on SDS-PAGE visualization with Coomassie brilliant blue, Bradford analysis of total protein, and determination of granzyme B concentration by V_{max} measurement. The fractions from the SP-sepharose column were pooled and dialyzed against 50 mM MES pH 6.0, 100 mM NaCl and loaded onto a 1 mL Mono-S® cation exchange column (Pharmacia Biotech, Uppsala, Sweden). The Mono-S column was washed with eight column volumes of buffer containing 50 mM MES pH 6.0, 100 mM NaCl. The column was then treated with a salt gradient from 100 mM to 800 mM to elute active granzyme B at a salt concentration of 580 mM NaCl. The final product was ≥98% pure as judged by SDS-PAGE Coomassie brilliant blue staining as described by Shägger and Jacow (Schagger and von Jagow, 1987), Bradford analysis of total protein, and V_{max} measurement of granzyme B.

The concentration of granzyme B protein was determined by absorbance measured at 280 nm and based on an extinction coefficient of 13,000 M^{-1} cm^{-1} (Gill and von Hippel, 1989). The proportion of catalytically active protein was quantitated as follows: active sites of trypsin solutions were titrated with MUGB (Jameson, 1973b). Assuming a 1:1 stoichiometry (McGrath *et al.*, 1994), ecotin solutions were then quantitated with active site titrated trypsin. Concentration of granzyme B was then quantitated with the active site titrated ecotin, again assuming a 1:1 stoichiometry, using Ac-IEPD-pNA as the substrate. The percentage of catalytically active protein was >95%.

To determine the glycosylation state of recombinant granzyme B, 20 μg of granzyme B was denatured by boiling for 10 minutes in the presence of 0.5% SDS and 1% β-mercaptoethanol. The denatured granzyme B was separated into two aliquots and one aliquot was incubated at 37°C for three hours with ten units of N-glycosidase F in 50 mM Tris, pH 7.5 and 1% Triton X-100. The products were analyzed by SDS-PAGE and Coomassie brilliant blue staining.

Construction of granzyme B variants.

Site-directed mutagenesis reactions were performed by the method of Kunkel (Kunkel, 1985) using single stranded template containing the coding sequence of granzyme B. The oligonucleotides used were as follows (mismatches are underlined): [R192A] CC TCC AGA GTC CCC CGC AAA GCT AGC ACG TTT GAT CTT TGG G; [R192E] CC TCC AGA GTC CCC CTC AAA GCT AGC ACG TTT GAT CTT TGG G. Conditions of expression and purification of the granzyme B variants were identical to those described for the wild-type enzyme.

Expression of macromolecular inhibitors of granzyme B.

Conditions of expression and purification of ecotin and $\text{Ecotin}_{\text{IEPD}}$ were identical to those previously described (Wang *et al.*, 1995). Site-directed mutagenesis was performed by the method of Kunkel using the single stranded template of the coding sequence of ecotin. The oligonucleotide used to incorporate the four-amino acid replacement of the ecotin active site loop, from VSTM to IEPD, was as follows (mismatched are underlined):

[$\text{Ecotin}_{\text{IEPD}}$] GTC AGT TCC CCG ATT GAA CCG GAT ATG GCA TGC CCG GAT GGC.

Positional scanning synthetic combinatorial library.

Preparation and screening of the positional scanning synthetic combinatorial library (PS-SCL) was carried out as previously described (Rano *et al.*, 1997; Thornberry *et al.*, 1997). The concentration of substrates was 0.25 μM, making the activity directly proportional to the specificity constant, k_{cat}/K_m. Enzyme activity of the PS-SCL was assayed in 100 mM HEPES, pH 7.5, 10 mM DTT at 25°C in a Tecan Fluostar (Research Triangle Park, North Carolina) at excitation and emission wavelengths of 380 nm and 460 nm respectively.

Single substrate kinetic assays.

Enzyme activity was monitored at 25°C in assay buffer containing 50 mM Tris pH 7.4 and 100 mM NaCl. Substrate stock solutions were prepared in DMSO. The final concentration of substrate ranged from 0.005 to 4 mM, the concentration of DMSO in the assay was less than 5%. Enzyme concentrations ranged from 5-50 nM. Hydrolysis of pNA substrates was monitored spectrophotometrically at 410 nm on a UVIKON 860 spectrophotometer. Hydrolysis of SBzl substrates was monitored spectrophotometrically at 324 nm in the presence of 4,4'-dithiodipyridine. Hydrolysis of AMC substrates was monitored fluorometrically with an excitation wavelength of 380 nm and emission wavelength of 460 nm on a Fluoromax-2 spectrofluorimeter.

Hydrolysis of Ac-IEPDWGA-NH_2 and Ac-IEPDWNA-NH_2 by granzyme B was monitored at 220 nm by reverse phase HPLC using a C18 column (Vydac, 5μ, 4.6 x 250 mm) with a 0-70% gradient of 0.1% trifluoroacetic acid and 0.08% trifluoroacetic acid/95% acetonitrile. Substrate concentrations ranged from 0.01-2 mM and enzyme concentrations from 9-90 nM. The enzymatic hydrolysis was quenched by the addition of trifluoroacetic acid to 0.4%. Extent of hydrolysis was determined from the calculated area of the product peak.

Creation of P3, P1', P2' His-tagged substrate phage library.

The phagemid pHisX3P3 was constructed based on the phagemid pBSeco-gIII (26). The vector contains the following amino acid sequence inserted between the ecotin secretion signal and residues 198-406 of pIII coat protein of M13 bacteriophage: A-E-S-V-Q-P-L-G-P-G-H-H-H-H-H-H-H-G-H-A-G-**I-*X*-P-D-*X*-*X***-A-G-P-G-G-G. The inserted sequence produces a histidine-tag (underlined) linked to a substrate sequence (bolded) randomized in the P3, P1', and P2' positions (bolded and italicized) followed by a GPGGG linker to pIII. The degenerate oligonucleotides synthesized to create the library consisted of the following sequence (where N indicates equimolar concentrations of G, C, A and T; S indicates equimolar concentrations of G and C): CAT GGG CAT GCA GGA ATT NNS CCA GAC NNS NNS GCA GGG CCC GGA GGC GGT CCA TTC GTT. This oligonucleotide was used in combination with a primer that annealed to the HindIII sequence 3' to the pIII coding sequence for PCR. Digestion of the 680 bp PCR product with SphI and HindIII was followed by ligation into the SphI/HindIII cut pHisX3P3 vector. The library of substrate phage has 32,768 possible DNA sequences which translates into 8000 possible protein sequences.

Phage particles expressing the engineered pIII-substrate fusion protein were prepared as previously described (Wang *et al.*, 1995). Briefly, the library plasmid DNA was transformed into Epicurian Coli® XL2-Blue MRF' cells (Stratagene, La Jolla, CA). The transformation efficiency was determined by plating a portion of the transformed cells onto LB plates containing 60 μg/mL ampicillin (LB-AMP). The transformation efficiency was 1 x 10^6 individual clones, which allowed for >99.9% completeness of the library. The rest of the transformed cells were allowed to grow at 37°C with shaking in 2YT containing 60 μg/mL ampicillin to an A_{600} of 0.25 and were then infected with VCSM13 helper phage at a multiplicity of infection of 100 phage/cell for the production of recombinant phage. The culture was allowed to grow for six hours. Phage particles were precipitated by addition of polyethylene glycol-8000 to 5% and NaCl to 500 mM. Phage were resuspended in 50 mM Tris pH 7.4, 100 mM NaCl.

His-tagged substrate phage cleavage assay:

Two hundred microliters of nickel (II)-nitrilo-tri-acetic acid resin (Qiagen, Santa Clarita, CA) was washed with 10 mL of activity buffer (50 mM Tris, pH 7.4, 100 mM NaCl, 0.1% Tween-20). Phage particles (10^9) were added to the washed Ni (II) resin and allowed to bind with gentle agitation for three hours. The Ni (II) resin was then washed with 40 mL of activity buffer to remove unbound phage. Recombinant granzyme B was added to a final concentration of 10 nM. After an incubation period of six hours, the cleaved phage were separated from the resin in a total volume of five milliliters of buffer. The cleaved phage were amplified by the infection of male strain JM101 (F') cells with the addition of VCSM13 helper phage to form recombinant phage which were then used for the next round of cleavage selection. After four rounds of cleavage selection, JM101 (F') cells were infected with the eluted phage and plated onto LB-AMP. Twenty individual colonies were selected and grown in three milliliters LB-AMP and plasmid DNA was isolated and sequenced in the region of the cleavage site.

Molecular modeling of Granzyme B-substrate complex.

Rat mast cell protease-2 (RMCP-2) (Remington *et al.*, 1988) was used as the framework to model rat granzyme B since they share 49% sequence identity. The coordinates for RMCP-2 were obtained from the Brookhaven Protein Data Bank (3RP2). The Biopolymer module of the Biosym (San Diego, CA) molecular modeling package was used to replace the amino acids of RMCP-2 with those of granzyme B. The structure was minimized with the Discover module of the Biosym molecular modeling package. The substrate structure is based on the trypsin-ecotin inhibitor complex (McGrath *et al.*, 1994) and docked onto the granzyme B structure.

Poly (ADP-Ribose) Polymerase mutagenesis and cleavage assay.

The 3045 bp cDNA for human PARP was cloned downstream of the T7 RNA polymerase promoter into the SacI and XbaI sites of pBluescript II KS(+) vector (Stratagene). Site-directed mutagenesis was performed by the method of Kunkel (Kunkel, 1985) to introduce an alanine substitution at aspartate 536. The oligonucleotides used were as follows (mismatches are underlined): [D536A] GCA GCT GTG GAT CCT GCA TCC GGA CTG GAA CAC TC and [D644A] CCC CTG GAG ATT GCG TAC GGC CAG GAT GAA GAG. The TnT® coupled reticulocyte lysate system (Promega) was used to transcribe and translate the PARP and PARP [D536A] genes. The translation products were then incubated with 100 nM granzyme B in the reticulocyte lysate at 25°C in 50 mM Tris, pH 7.4 and 100 mM NaCl.

Results

Expression and purification of rat granzyme B.

The active form of rat granzyme B was expressed and secreted from the methylotropic yeast *Pichia pastoris* following induction with methanol. Cleavage activity of a P1-Asp substrate was detected 24 hours into induction, and increased to a plateau at 48 hours (Figures 5-1A). A 150-fold purification of recombinant granzyme B was achieved by cation exchange chromatography using SP-sepharose and Mono-S column resins (Fig 5-1B, Lane 1 and 2 and Table 5-1). The final yield of purified (>98%) active granzyme B was 0.7 milligrams per liter of initial culture. Recently, a similar expression system for mouse granzyme B was designed utilizing *his4* rather than zeocin selection (Pham, 1998).

Two species of granzyme B are detected by Coomassie brilliant blue staining (Figure 5-1B). The apparent molecular mass of the major species is ~30 kDa, which exceeds that of 25 kDa based on the amino acid sequence of granzyme B. This difference between the actual and the expected molecular mass suggests glycosylation of granzyme B at the N-glycosylation consensus site, Asn 66. Treatment of recombinant granzyme B with N-Glycosidase F resulted in a decrease of apparent molecular mass to that of the minor species, approximately 27 kDa (Figure 5-1C, Lane 1 and 2). Removal of the N-linked glycosylation site by site directed mutagenesis (N66Q) resulted in a catalytically active enzyme. However, the yield of secreted protein was decreased 100-fold (data not shown).

The direct detection of granzyme B hydrolytic activity in the culture supernantant indicates that the α-factor leader sequence is correctly processed immediately before Ile16 of granzyme B. Like most enzymes of this class, granzyme B is synthesized as an inactive zymogen with an N-terminal pro-peptide. Another protease is required for cleavage of the propeptide for mature processing of granzyme B. Our expression system was designed to bypass the addition of another protease to activate granzyme B thereby eliminating the potential of complicating subsequent substrate specificity studies.

Primary (P1) substrate specificity.

Purified recombinant granzyme B was tested for hydrolytic activity against a panel of tetrapeptide substrates of the form Suc-AAPX-pNA which contained various amino acids at the P1 position (P1-Ala, Glu, Phe, Leu, Met, or Arg). The only substrate with detectable activity was that with P1-Asp, with a k_{cat}/K_m of 62 M^{-1} s^{-1}. Cleavage of the thiobenzyl ester substrate, Boc-AAD-SBzl, by recombinant rat granzyme B demonstrates that the enzyme is kinetically indistinguishable from natively purified human granzyme B ($k_{cat,\ recombinant}$= 11.4 s^{-1}, $K_{m,\ recombinant}$ = 140 µM; $k_{cat,\ native}$ = 11 s^{-1}, $K_{m,\ native}$ = 145 µM (13)).

Extended (P4-P2) substrate specificity.

A positional scanning combinatorial substrate library (PS-SCL) was used to elucidate the specificity of purified, recombinant, rat granzyme B. This library, of the general structure Ac-X_{P4}-X_{P3}-X_{P2}-D-AMC, has been previously used to identify the amino acid preferences of the caspases and human granzyme B purified from cultured natural killer leukemia YT cells (Thornberry *et al.*, 1997). The PS-SCL is composed of three libraries, each of which consists of 20 sublibraries, for a total of 8000 compounds. In each sublibrary, one position (P4, P3, or P2) contains a defined amino acid and the other two positions contain an equimolar mixture of amino acids (two unnatural amino acids, d-alanine (d-A) and norleucine (n), are included; cysteine and methionine are excluded). Thus analysis of the three libraries affords a complete understanding of enzyme preferences for amino acids at P4, P3, and P2. This approach has been previously validated as providing an accurate measure of protease specificity using caspase-1 (Rano *et al.*, 1997).

Using this method we determined that the preferred tetrapeptide substrate recognition sequence for recombinant, rat granzyme B is (I>V)(E>Q=n)XD (Figure 5-2A). Granzyme B prefers the beta-branched aliphatic amino acids, isoleucine and valine, in the P4 position. Glutamate, glutamine and norleucine are the preferred amino acids in the P3 position. The PS-SCL indicates that granzyme B can accept a broad range of amino acids in the P2 position, although proline is the preferred amino acid.

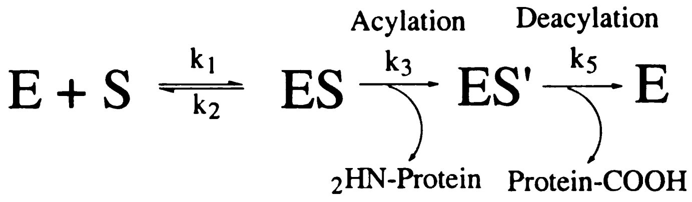
Kinetic consequences of extended (P4-P2) substrate specificity.

The positional scanning-synthetic combinatorial library suggests that granzyme B exhibits unique and extended substrate specificity. To quantitiate dependence on extended interactions and to explore the mechanism by which granzyme B utilizes extended binding energy to achieve catalysis, kinetic parameters were determined for various substrates (Table 5-2). While catalysis by the familiar pancreatic serine proteases is influenced by the length of the peptide-substrate, the majority of substrate specificity is observed in P1.

Granzyme B, on the other hand, is absolutely dependent on extended binding of substrate for efficient catalysis. Granzyme B is not capable of cleaving the dipeptide substrate, Ac-PD-pNA or the single residue substrate, Ac-D-pNA, at concentrations as high as 4 mM. Reducing the optimal tetrapepide, Ac-IEPD-pNA, to the tri-peptide, Ac-EPD-pNA, results in a >100-fold decrease in activity, from 666.0×10^2 M^{-1} s^{-1} in k_{cat}/K_m to 6.5×10^2 M^{-1} s^{-1}. Primary sequence recognition is also an important requirement for catalysis. Hydrolysis of the sub-optimal P4-P3 substrate Ac-AAPD-pNA is disfavored by greater than 1000-fold as compared to the optimal substrate Ac-IEPD-pNA, decreasing from 666.0×10^2 M^{-1} s^{-1} for the latter substrate to 0.62×10^2 M^{-1} s^{-1} for the former. The sub-optimal P3 substrate Ac-IKPD-pNA is disfavored by 10-fold when compared to the optimal substrate (Table 5-2).

The dependence of granzyme B catalysis on recognition of a particular tetrapeptide sequence becomes more manifest when the scissile bond of the substrate contains a poorer leaving group (i.e. more peptide-like). The sub-optimal P3 substrate Ac-IKPD-AMC results in a 140-fold decrease in catalysis, from 33.3×10^2 M^{-1} s^{-1} for Ac-IEPD-AMC to 0.23×10^2 M^{-1} s^{-1} for Ac-IKPD-AMC (Table 5-2). Whereas under conditions of ester hydrolysis, Ac-IEPD-SBzl and Ac-IKPD-SBzl are almost indistinguishable, k_{cat}/K_m values of 3233×10^2 M^{-1} s^{-1} to 2683×10^2 M^{-1} s^{-1} respectively (Table 5-2).

The demonstration that granzyme B does not display substrate specificity in ester hydrolysis is consistent with what is known about other enzymes of this class (Hedstrom *et al.*, 1992). Serine proteases, including granzyme B, catalyze the cleavage of substrates through the following mechanism:



$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow[{}_2\text{HN-Protein}]{k_3} ES' \xrightarrow[\text{Protein-COOH}]{k_5} E$$

The equations used to define the steady state macroscopic constants in terms of microscopic constants for the above mechanism are:

$$k_{cat} = \frac{k_3\,k_5}{k_3 + k_5} \qquad K_m = K_d\left(\frac{k_5}{k_3 + k_5}\right)$$

Typically, the deacylation step is completely rate limiting for ester hydrolysis by serine proteases (i.e. $k_3 >> k_5$ and therefore $k_{cat,\,ester} = k_5$) (Zerner, 1964). Extending this assumption to recombinant rat granzyme B, the data obtained from ester hydrolysis of Ac-IEPD-SBzl and Ac-IKPD-SBzl can be used to derive the mechanistic constants for the amide hydrolysis of Ac-IEPD-AMC and Ac-IKPD-AMC from the following equations:

$$k_3 = \frac{k_{cat,ester} \bullet k_{cat}}{k_{cat,ester} - k_{cat}} \qquad K_d = K_m \bullet \frac{k_3 + k_{cat,ester}}{k_{cat,ester}}$$

The observation that K_d increases approximately 5-fold, from 167 μM for Ac-IEPD-AMC to 805 μM for Ac-IKPD-AMC (Table 5-3), demonstrates that granzyme B utilizes extended substrate binding sites to enhance formation of the ground state Michaelis complex. However, while ground state stabilization is observed, an even greater stabilization occurs in the acylation transition state, k_3 decreases 24-fold from 0.558 s^{-1} for Ac-IEPD-AMC to 0.023 s^{-1} for Ac-IKPD-AMC (Table 5-3).

Extended (P3, P1', P2') substrate specificity determined by substrate phage display and single peptide kinetics.

To elucidate the granzyme B substrate preference C-terminal to the scissile bond (prime-side), the method of monovalent "substrate-phage" display was used (Matthews *et al.*, 1994; Matthews and Wells, 1993). One million individual clones representing 32,768 nucleotide sequences and 8000 protein substrate sequences were displayed on phage particles between a histidine-tag affinity anchor and the M13 phage coat protein, pIII. The histidine-tag allows the "substrate" displaying phage to be immobilized on Ni (II)-resin at an approximately nanomolar concentration (Figure 5-3A). Upon incubation with granzyme B, the phage that are displaying substrate sequences susceptible to cleavage by granzyme B lose their histidine-tags; the phage displaying sequences refractory to cleavage remain bound to the Ni(II)-resin (Figure 5-3B). The eluted phage are subsequently used to infect *E. coli* cells (Figure 5-3C) and are amplified by the addition of helper phage for another round of cleavage selection or are isolated for plasmid isolation and DNA sequencing (Figure 5-3D). After four rounds of cleavage selection, twenty of the phagemid DNA plasmids were sequenced and analyzed. Translation of the resulting DNA sequences (Table 5-4) from the substrate phage confirms the PS-SCL result for preference of glutamate at the P3 position. Interestingly, methionine, which was not included in the positional scanning synthetic combinatorial library, is also represented at this position. While the amino acid preference of rat granzyme B at the P1' position is not strict, there is a distinct absence of charged amino acids. The results of the substrate phage library also indicate that rat granzyme B has specificity for glycine at the P2' position. The specificity for glycine at P2' was tested through cleavage of two P4-P3' extended peptides, one with a glycine and the other with an asparagine in the P2' position (Table 5-2). The resulting kinetic data indicated that in the context of the peptides tested, P2'-glycine is 5-fold more efficiently cleaved than asparagine at the same position.

Structural determinants of granzyme B substrate specificity.

To understand the structural determinants of the extended substrate specificity of granzyme B, a homology-built three-dimensional model of the enzyme-substrate complex was made (Figure 5-4). This model can provide a structural framework to rationalize our experimental results as well as guide future studies. In this model a hydrophobic pocket is formed from Ile 99, Tyr 215 and Tyr 175 around the P4 position of the substrate. Acidic P3 and P1 specificity appears to be a result of the positive electrostatic surface formed by Arg 192 and Arg 226. Although the overall sequence identity between human and rat granzyme B is greater than 70%, there is a conservative amino acid change at position 192 from an arginine in the rat to a lysine in the human granzyme B. The subtle differences in P3 substrate specificity between the human (E>>G>S>D) and the rat (E>>Q=n>A>S>D) granzyme B may be dictated by the difference at position 192. A hydrophobic pocket is formed by loop A which may allow for accommodation of large hydrophobic amino acids. This pocket is formed by the extended conformation of loop A, which contains a two amino acid insertion when aligned with other chymotrypsin-like proteases. A tryptophan modeled at the P1' position of substrate can make favorable hydrophobic interactions with Ile 35, the Cys 42-Cys 58 disulfide bond, and the aliphatic portion of Lys 41. The specificity for glycine at P2' may be a result of glycine's unique absence of a side chain, allowing for the maximization of backbone hydrogen bonding between P2'-Gly and Lys 41, while not sterically clashing with the Arg 192 side-chain.

To explore the mechanism of substrate discrimination by granzyme B, mutations of the arginine 192 side chain were made. Substitution of arginine 192 for alanine resulted in a mutant with only 1.7% to 36% of specific activity versus the wild-type enzyme for Ac-IEPD-AMC and Ac-IKPD-AMC, 55 s^{-1} M^{-1} versus 3330 s^{-1} M^{-1} and 8.3 s^{-1} M^{-1} versus 23.3 s^{-1} M^{-1} (Table 5-5). The smaller differences between these two substrates for the [R192A] mutant versus the wild-type enzyme, indicates that the decreased activity is not due solely to P3 interactions but may also be affecting P1 interactions. Indeed, the P4-P2 substrate specificity profile for GrB[R192A], as screened by the PS-SCL, is identical to the wild-type enzyme, indicating that there are additional determinants for P3-Glu specificity (Figure 5-2B). Support for the role of arginine 192 as a determinant for P3 specificity occurs when arginine 192 is mutated to glutamate. The specificity profile indicates that GrB[R192E] has decreased activity for acidic amino acids at P3, while the activity against basic amino acids has significantly increased (Figure 5-2C). This mutant, GrB [R192E], retains only 0.3% of the wild-type activity for the optimal substrate, Ac-IEPD-AMC, while rescuing the majority, 71%, of the wild type activity for the complementary, though non-optimal, substrate Ac-IKPD-AMC (Table 5-5).

Analysis of the mechanistic rate constants for the arginine 192 mutants provides insight into how substrate selectivity is attained by granzyme B. The wild type enzyme is able to utilize extended P3 interactions to preferentially stabilize the acylation transition state versus the Michaelis complex ground state as indicated by the 5-fold increase in K_d for Ac-IEPD-AMC versus Ac-IKPD-AMC and the 24-fold decrease in k_3 for these same substrates (Table 5-3). For the optimal substrate, Ac-IEPD-AMC, there is a sequential decrease in both the affinity and acylation rate upon mutation of arginine 192, indicated by increased K_d and decreased k_3, respectively (Table 5-3). For the less-optimal substrate, Ac-IKPD-AMC, there are smaller differences in K_d upon mutation. While approximately the same decrease in acylation exists between the two substrates for the [R192A] variant, acylation is completely rescued in the [R192E] variant for the Ac-IKPD-AMC substrate (Table 5-3).

Cleavage of Poly(ADP-Ribose) polymerase by granzyme B.

Based on the current specificity studies with granzyme B, two potential granzyme B cleavage sites were identified in poly (ADP-ribose) polymerase (PARP), VDPD/SG at position 536 in PARP and LEID/YG at position 644. While neither of these sites is absolutely ideal for granzyme B, they contain a majority of the defined specificity determinants for granzyme B. Cleavage at Asp 536 would yield 59 and 54 kDa fragments (Figure 5-5B). Cleavage at Asp 644 would yield 72 and 41 kDa fragments (Figure 5-5B). To establish if these two sites were processed by granzyme B, wild-type PARP and the cleavage site mutants [D536A] and [D644A] were constructed and produced in rabbit reticulocyte lysate. Upon incubation of PARP with granzyme B in the lysate, bands at 89, 59, 54, 41, 35, and 24 kDa were observed (Figure 5-5C Lane 2). The 89 and 24 kDa fragments were reminiscent of the cleavage of PARP by caspase-3 at the site DEVD/EV. Cleavage at this caspase site would also explain the appearance of the 35 kDa fragment (Figure 5-5B). Additional experiments were undertaken to determine if the caspase-like cleavage of PARP was due directly to granzyme B or if the reticulocyte lysate contained a latent caspase that was activated upon incubation with granzyme B. We observed that in the presence of a macromolecular inhibitor that is specific for granzyme B, Ecotin$_{IEPD}$ (K_i = 1nM, Harris and Craik, unpublished data), cleavage of PARP was abolished (Figure 5-5C Lane 3). However, in the presence of the specific caspase-3 inhibitor, Z-DEVD-FMK, the predominant cleavage bands at 59 and 54 kDa were observed (Figure 5-5C Lane 4). Incubation of reticulocyte lysate alone with granzyme B followed by inhibition of the granzyme B with Ecotin$_{IEPD}$ resulted in an activity in the lysate that was capable of cleaving PARP to the 89 and 24 kDa fragments (Figure 5-5C Lane 5). Indeed, incubation of purified PARP with granzyme B resulted in the 59 and 54 kDa, and to a lesser extent 41 kDa cleavage products (data not shown). Confirmation that VDPD/SG was the major cleavage site in PARP was obtained upon removal of the cleavage site through mutation of aspartate 536 to alanine to give the altered sequence VDPA/SG. Cleavage to the 59 and 54 kDa fragments was not observed with PARP [D536A] in the reticulocyte lysate (Figure 5-5C Lanes 6-10). However, 72 and 41 kDa fragments appeared, suggesting that the LEID/YG is a second cleavage site for granzyme B. Upon removal of the secondary cleavage site by mutation of the aspartate 644 to alanine to give the sequence LEIA/YG, the 41 kDa fragment was no longer observed (Figure 5-5C Lanes 11-15).

Discussion

Cytotoxic lymphocytes express multiple serine proteases that appear to be essential to targeted cell death administered through the granule exocytosis pathway. One of the most abundant serine proteases in the granules is granzyme B. While cytotoxic lymphocytes derived from mice with a disruption in the granzyme B gene are impaired in inducing rapid target cell death (Heusel *et al.*, 1994), the mechanism of granzyme B's involvement in this process has not been fully elucidated. A current model is that granzyme B enters the cell and mediates proteolytic activation of the apoptotic class of caspases (Froelich *et al.*, 1996b), although it may also act directly on cellular substrates. Recent studies have indicated that granzyme B is transported to the nucleus in the presence of

UCSF LIBRARY

perforin (Jans *et al.*, 1996; Shi *et al.*, 1997). This suggests that granzyme B may have nuclear substrate targets that it cleaves directly, rather than indirectly through caspase activation. Indeed, granzyme B has recently been shown to efficiently cleave several caspase substrates (Andrade *et al.*, 1998). Definition of the substrate specificity of granzyme B may help uncover the role it plays in CL mediated cell death.

Multiple granzymes are stored in the granules of CLs and each may have an individual or redundant function. We have focused the present study on granzyme B since it may serve as an important structural model for the related granzymes. To avoid contamination by other granzymes, a recombinant expression system in yeast was developed. In addition to producing reagent quantities of granzyme B, the heterologous expression system has the advantage of making variant versions of the enzyme including the mature enzyme, circumventing the need for an additional activating protease. Expression in yeast also allows for correct post-translational modification, such as disulfide bond formation and glycosylation. We believe that glycosylation may aid in stabilizing or solubilizing granzyme B, and this may explain why our attempts at expression in prokaryotes failed (data not shown).

Like native granzyme B, the recombinant protease is rigorously selective for cleavage after aspartate residues. Previously reported studies with ester substrates indicated that granzyme B has the ability to cleave after Asn, Met and Ser (Odake *et al.*, 1991; Pham, 1998). While we have observed the esterolytic activity of granzyme B on these substrates, cleavage of cognate amide substrates was not seen. These observations are a direct result of the catalytic mechanism of granzyme B. Under conditions of ester hydrolysis, deacylation of the acyl-enzyme intermediate is rate limiting (i.e. $k_{cat}=k_5$). Since the enzyme does not distinguish between multiple ester substrates, substrate specificity is not manifested in deacylation. Granzyme B is an efficient, but not very specific, esterase. Presumably, once the acyl-enzyme intermediate is formed, the covalent nature of this interaction may be sufficient for correct positioning of the substrate in relation to the catalytic groups for the deacylation reaction to occur.

In contrast to ester hydrolysis, amide hydrolysis by granzyme B is very specific and dependent on extended enzyme-substrate binding interactions. For amide hydrolysis, granzyme B activity increases with the ability of the leaving group to delocalize the developing negative charge in the acylation transition state (i.e. pNA>AMC>peptide). This trend in leaving group dependence is consistent with acyl-enzyme formation being rate determining for amide hydrolysis. Using a synthetic combinatorial approach in which the hydrolysis of a fluorogenic amide was monitored, the extended non-prime specificity of rat granzyme B was defined from P4 to P2. The optimal P4-P1 substrate sequence for granzyme B was determined to be Ile-Glu-Pro-Asp. Enzymological characterization of extended substrates showed that granzyme B is critically dependent on favorable extended interactions and is not capable of hydrolyzing amide substrates less than three amino acids in length. Truncating the substrate from a tetrapeptide to a tripeptide resulted in a dramatic 100-fold decrease in catalysis. One possible explanation for this data is that extensive substrate interactions are necessary to stabilize the enzyme active site in a catalytically productive orientation.

Specificity of an enzyme for a substrate is dependent on the ability of an enzyme to utilize substrate binding energy to promote catalysis. Implicit in this definition is the recognition that specificity requires the stabilization of the rate determining transition state complex. In the case of the pancreatic serine proteases, the importance of maximizing k_{cat}/K_m is observed by the preferential stabilization of the transition state over stabilization of the ground state Michaelis complex. Loss of catalytic efficiency through ground state stabilization is not desirable in proteases whose main biological function is digestion. In the case of the granzymes, whose biological role may be more regulatory in nature, the pressure to maximize catalytic rate may be secondary to the pressure to maximize control. Therefore, stabilization of the transition state would still be necessary for the manifestation of specificity, but binding energy could also be used to stabilize ground state interactions. The large increases in K_m (and K_d where determined) for sub-optimal substrates indicate instability of the ground state Michaelis complexes. In addition to the decreased ground state stabilization, transition state stabilization is significantly decreased for the amide hydrolysis of sub-optimal substrates as observed by k_{cat} (and k_3 where determined).

To examine the substrate specificity on the prime-side of the substrate (C-terminal to the scissile bond) and to verify the P3 specificity results from the synthetic library in a protein context, the combinatorial method of substrate phage was used. This method has been used successfully by other groups to determine the substrate specificities of multiple proteases (Ke *et al.*, 1997b; Matthews *et al.*, 1994; Matthews and Wells, 1993). Strict selectivity by granzyme B was not observed in the P1' position, but there is a general preference for large hydrophobic amino acids at this position, demonstrated by the presence of Trp, Leu, Phe, and Ile. There also appears to be a general aversion towards charged residues at the P1' position as demonstrated by the absence of Arg, Lys, Asp, and Glu. The presence of serine at this position may be reflective of the broad P1' specificity, coupled with the increased occurrence of serine in the initial library (serine is represented by three of the 32 possible codons at this position). Another explanation is that rat granzyme B has multiple binding modes and can accommodate both serine and tryptophan at the S1' subsite. In contrast to the broad specificity at the P1' position, results from the substrate-phage cleavage assay reflect a strong preference for glycine at the P2' position. We have shown that enzyme-substrate interactions C-terminal to the scissile bond are catalytically significant and play a role in determining substrate specificity. The presence of prime-side interactions are consistent with the mechanism we have described. If deacylation determined specificity, non-primed interactions would be of little consequence.

An advantage of the substrate-phage assay is that cleavage of the substrate occurs in a protein context. This allows us to evaluate how the results obtained from cleavage of the tetrapeptide substrates in the positional scanning combinatorial library reflect cleavage of protein substrates. The agreement between the two libraries for the preferential cleavage of substrates that contain glutamate at the P3 position supports the synthetic approach in elucidating protein substrate specificity. The correspondence of P3-Glu also confirms that granzyme B is predominantly cleaving the substrate-phage within the designed substrate sequence and not elsewhere in the pIII-His fusion protein. The appearance of methionine at the P3 position in the substrate phage library indicates that rat granzyme B has the ability to cleave substrates with P3-Met. Although methionine and cysteine are excluded from the PS-SCL for synthetic reasons, granzyme B has significant activity against norleucine, an unnatural amino acid that is approximately isosteric with methionine. Hence, these combinatorial approaches complement each other in the determination of substrate specificity.

The homology-built three-dimensional molecular model of granzyme B bound to a hexapeptide substrate with the optimal sequence provides additional insight into the structural determinants of specificity. One of the most striking structural differences between granzyme B and the pancreatic serine proteases is the lack of a disulfide bond at Cys191-Cys220. This may result in granzyme B having a less rigid binding site, and therefore requiring multiple substrate-enzyme interactions to overcome this entropic factor for efficient catalysis. In addition, the loss of the disulfide bond may allow the residue at position 192 to have more interactions with substrate. We postulate that arginine 192 plays a synergistic role with arginine 226 in determining the P1-aspartate and P3-glutamate substrate specificity. Qualitative evidence of this model was observed when arginine 226 was replaced with glycine, which resulted in decreased hydrolysis of a P1-Asp thiobenzyl ester substrate in the crude cell lysates in which the enzyme was expressed (Caputo *et al.*, 1994). We have demonstrated and quantitated the interaction of arginine 192 in determining the P3 and P1 specificity for amide substrates. Arginine 192 appears to be important in translating the extended substrate binding interactions into specific catalysis. We hypothesize that this position in the enzyme is important for determining the substrate specificity of the other granzymes. Indeed, granzyme C contains a glutamate at position 192. Loops A and B are capable of interacting directly with substrate and are likely to provide substrate specificity with residues C-terminal to the scissile bond. Recently, it has been shown that these loops act together to determine the S1' specificity in trypsin (Kurth *et al.*, 1997).

The recognition that granzyme B is a protease with extended specificity allows speculation on its biological role during cytotoxic lymphocyte mediated cell death. The amino acid preference of granzyme B has been defined for six subsites, P4 to P2', as (I>V)(E>Q=M)XD/XG (Figure 5A). *In vitro,* the substrate specificity of granzyme B can be kinetically defined; *in vivo*, other factors such as substrate availability, cellular localization and potential co-factors must be considered. However, there is a functional relationship between the preferential substrate sequence of granzyme B and the activation site of members of the caspases (Figure 5-5D). Indeed, studies have shown that granzyme B cleaves and activates several caspases involved in apoptosis (Chinnaiyan *et al.*, 1996; Darmon *et al.*, 1995; Duan *et al.*, 1996; Fernandes-Alnemri *et al.*, 1996; Muzio *et al.*, 1996; Orth *et al.*, 1996). However, granzyme B does not activate caspase 1 (ICE) or caspase 4 (Van de Craen *et al.*, 1997), both caspases that are involved in inflammation (Darmon *et al.*, 1994). Although these studies demonstrate that granzyme B has the ability to cleave and activate multiple caspases, the kinetic efficiencies have not yet been determined. Our data on the substrate specificity of granzyme B suggests that caspase 3 and caspase 7 are preferentially activated during apoptosis.

Knowledge of the extended substrate specificity of granzyme B allows for the proposal of additional targets of granzyme B during apoptosis. The substrate specificity of caspase 6 matches that of granzyme B (Thornberry *et al.*, 1997), suggesting that both enzymes cleave the same substrates. Several proteins known to be cleaved during apoptosis, such as nuclear lamin A (Orth *et al.*, 1996) and nuclear poly (ADP)-ribose polymerase (Lazebnik *et al.*, 1994), contain the potential granzyme B cleavage sites

VEID/NG and VDPD/SG, respectively (Figure 5-5B). In support of this, PARP has been shown to be cleaved directly by granzyme B and indirectly through the granzyme B activation of caspases. Mutation of the major and minor granzyme B cleavage sites in PARP abolishes cleavage at those sites. This demonstrates that the kinetically determined specificity of granzyme B is relevant in the context of a protein. Furthermore, other proteins that contain the preferred substrate sequence, but have not yet been shown to be cleaved during cytotoxic lymphocyte mediated cell death, may be targets for granzyme B.

In conclusion, we have demonstrated that granzyme B displays extended substrate specificity and that there is a significant dependence on these extended interactions for catalysis. Structural determinants of substrate specificity have been identified through construction of a structural model of granzyme B. This model has been tested through the redesign of the substrate specificity of the enzyme. Definition of the preferred substrate cleavage sequence has led us to propose a model in which granzyme B can activate members of the caspases as well as cleave other intracellular proteins. While we have limited this study to granzyme B, we expect that other members of the granzymes will display extended substrate specificities. The identification of their specificity will further expand our knowledge of the role that granzymes play in cytotoxic lymphocyte mediated cell death.

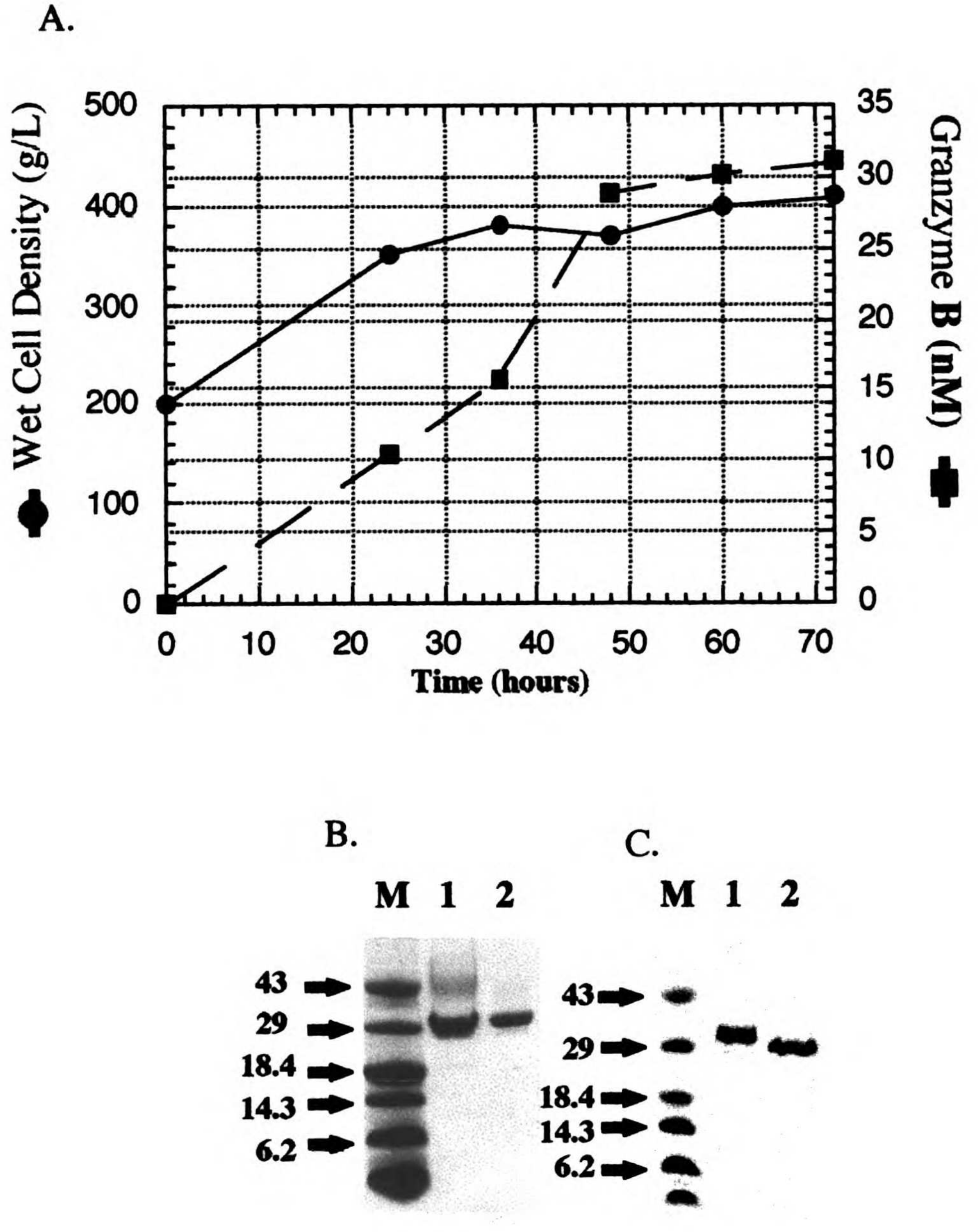
Figure 5-1. Expression and purification of granzyme B in *Pichia pastoris*.

A. A.Expression time course profile. Cell density and granzyme B production were monitored over 72 hours to determine optimal expression time.

B. Recombinant granzyme B purification. M, molecular mass markers. Lane 1, aliquot of pooled fractions eluted from SP-sepharose column. Lane 2, aliquot of pooled fractions eluted from Mono-S column. Samples were subjected to 10% SDS-PAGE as described in materials and methods and stained with Coomassie brilliant blue.

C. Recombinant granzyme B deglycosylation. M, molecular mass markers. Lane 1, control of granzyme B subjected to deglycosylation conditions, except that N-glycosidase F was not added. Lane 2, result of N-glycosidase F treatment of granzyme B. Samples were subjected to 10% SDS-PAGE as described in materials and methods and stained with Coomassie brilliant blue.

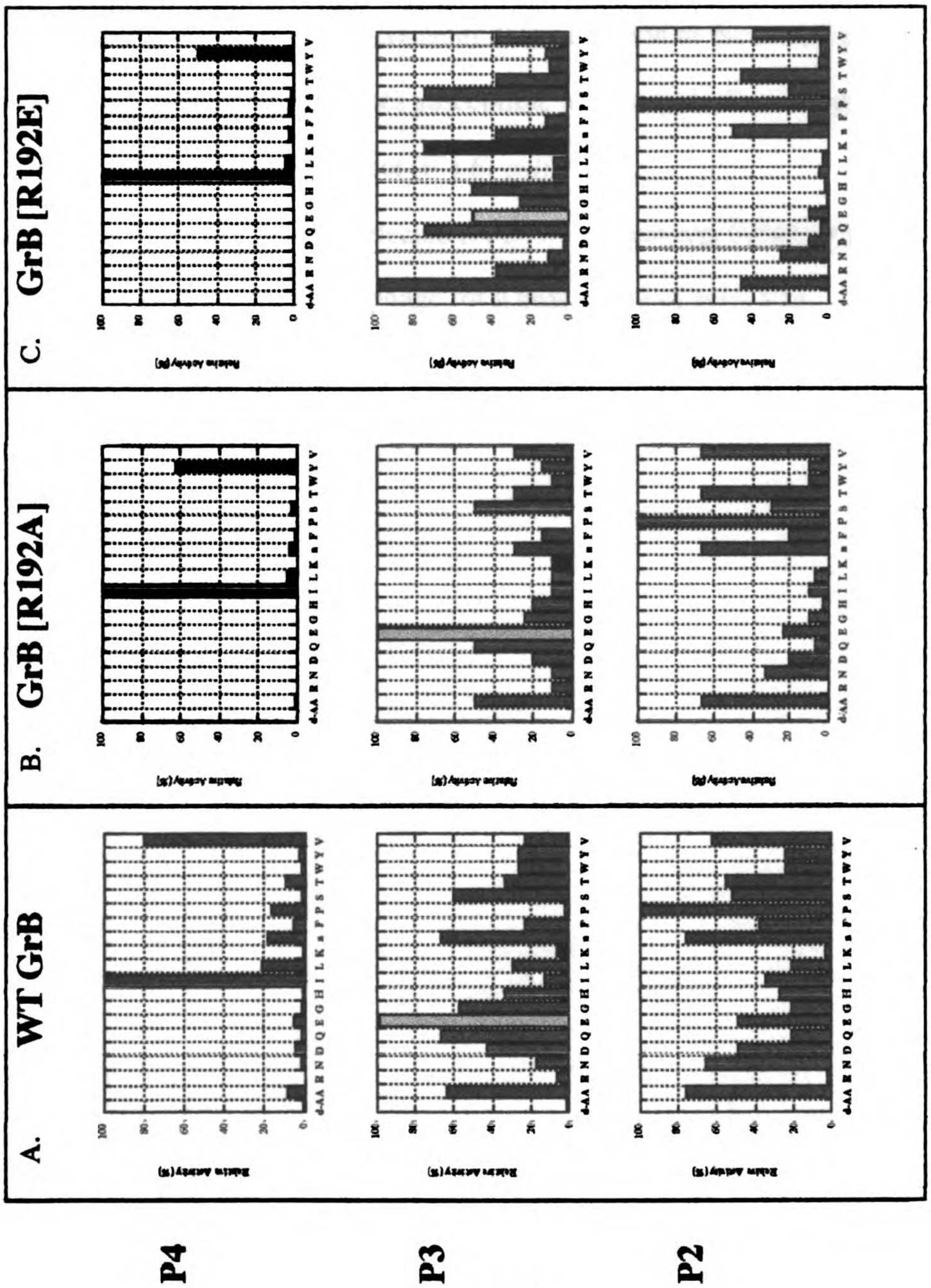
Figure 5-1.

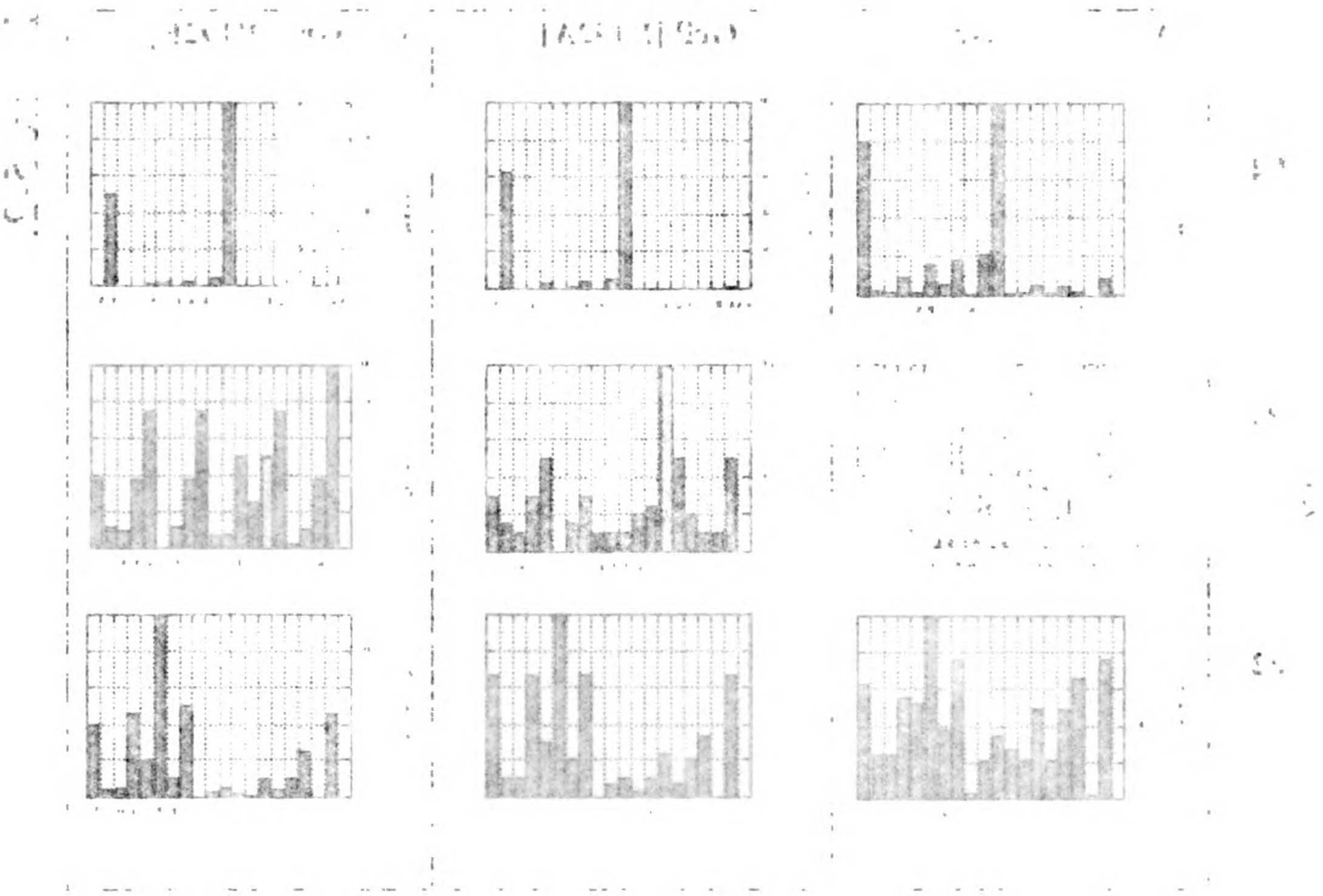

A.
500
400
300
200
100
0
● Wet Cell Density (g/L)
35
30
25
20
15
10
5
0
Granzyme B (nM) ■
0
10
20
30
40
50
60
70
Time (hours)
B.
M 1 2
43
29
18.4
14.3
6.2
C.
M 1 2
43
29
18.4
14.3
6.2

UCSF LIBRARY

Figure 5-2. P4-P2 substrate specificity profile of recombinant granzyme B and the granzyme B variants [R192A] and [R192E]. A positional scanning-synthetic combinatorial library was used to determine the amino acid preference at each of the subsites as described in material and methods. The y-axis indicates the rate of substrate hydrolysis relative to the maximum hydrolysis observed. The one letter code is used to denote the twenty natural amino acids, minus cysteine and methionine and plus d-alanine (d-A) and norleucine (n), on the x-axis. P3-Glu activity is indicated by a gray bar and P3-Lys activity is indicated with a black bar.

Figure 5-2.


A.
WT GrB
B.
GrB [R192A]
C.
GrB [R192E]
Relative Activity (%)
100
80
60
40
20
0
d-AA A R N D Q E G H I L K n F P S T W Y V
P4
P3
P2



UCSF LIBRARY

Figure 5-3. Schematic representation of substrate-phage display used to determine the P3, P1', and P2' substrate specificity profile for granzyme B.

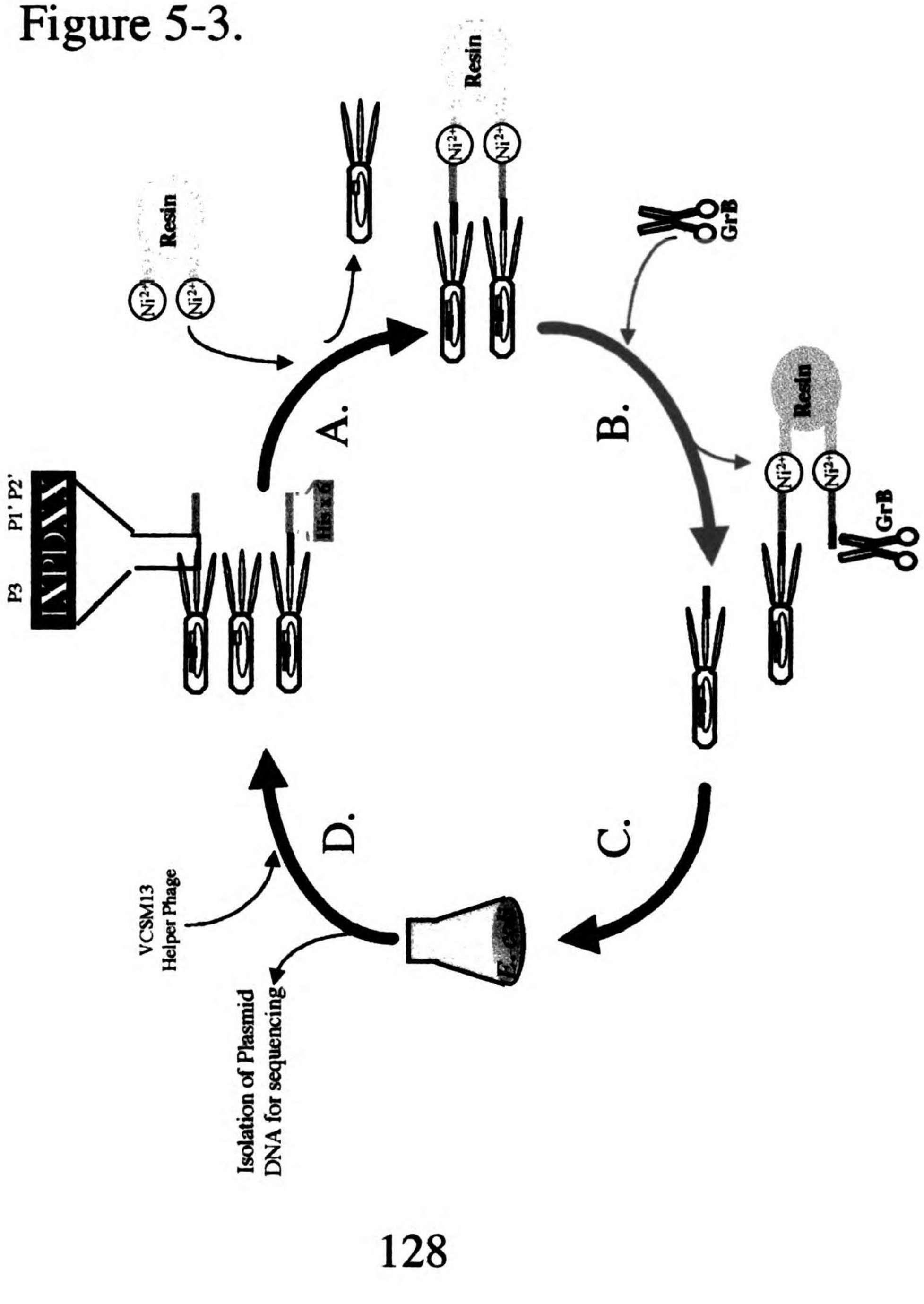
A. Phage displaying randomized substrate fused to pIII and a histidine-tag are immobilized on Ni(II) resin. Background phage or non-substrate displaying phage are washed away.

B. Granzyme B is added at a final concentration of 10 nM to the bound phage.

C. Those phage displaying substrates susceptible to granzyme B cleavage are eluted from the Ni(II) resin and used to infect (F') *E. coli.*

D. Individual clones can then be selected for DNA sequencing or helper phage can be added to produce recombinant phage for a next round of selection.

Figure 5-3.

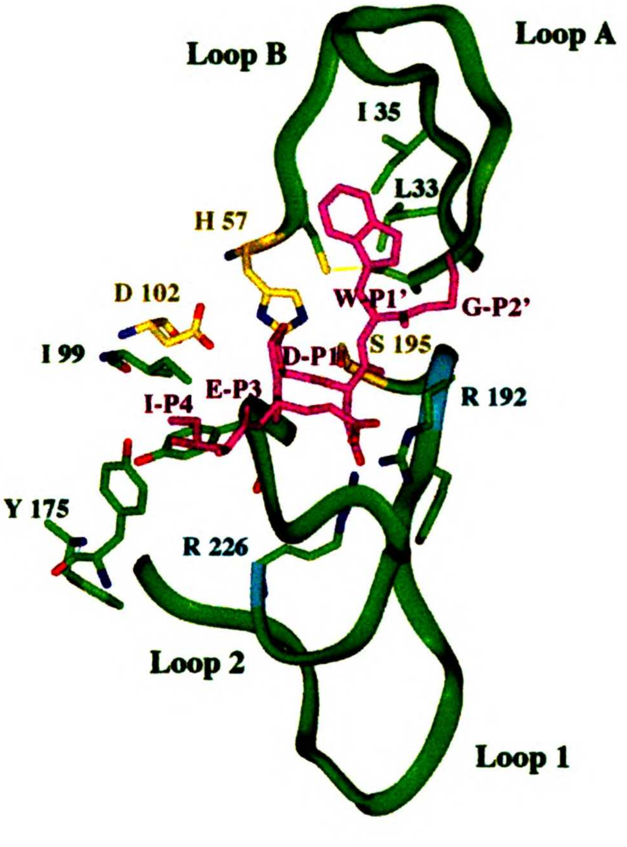


P3
P1' P2'
IXPDXX
His x 6
A.
Ni2+
Ni2+
Resin
Resin
Ni2+
Ni2+
GrB
B.
Resin
Ni2+
Ni2+
GrB
C.
D.
VCSM13
Helper Phage
Isolation of Plasmid
DNA for sequencing

UCSF LIBRARY

Figure 5-4. Three dimensional model of granzyme B complexed to IEPDWG substrate. Shown is the active site of granzyme B (green, catalytic triad in yellow) complexed to the determined optimal substrate (magenta). The model highlights potential interactions that determine P4, P3, P1, P1', and P2' specificity.

Figure 5-4.


Loop B
Loop A
I 35
L33
H 57
W-P1'
G-P2'
D 102
S 195
I 99
D-P1
R 192
E-P3
I-P4
Y 175
R 226
Loop 2
Loop 1

USGS LIBRARY

Figure 5-5. Biological significance of granzyme B P4-P2' substrate specificity.

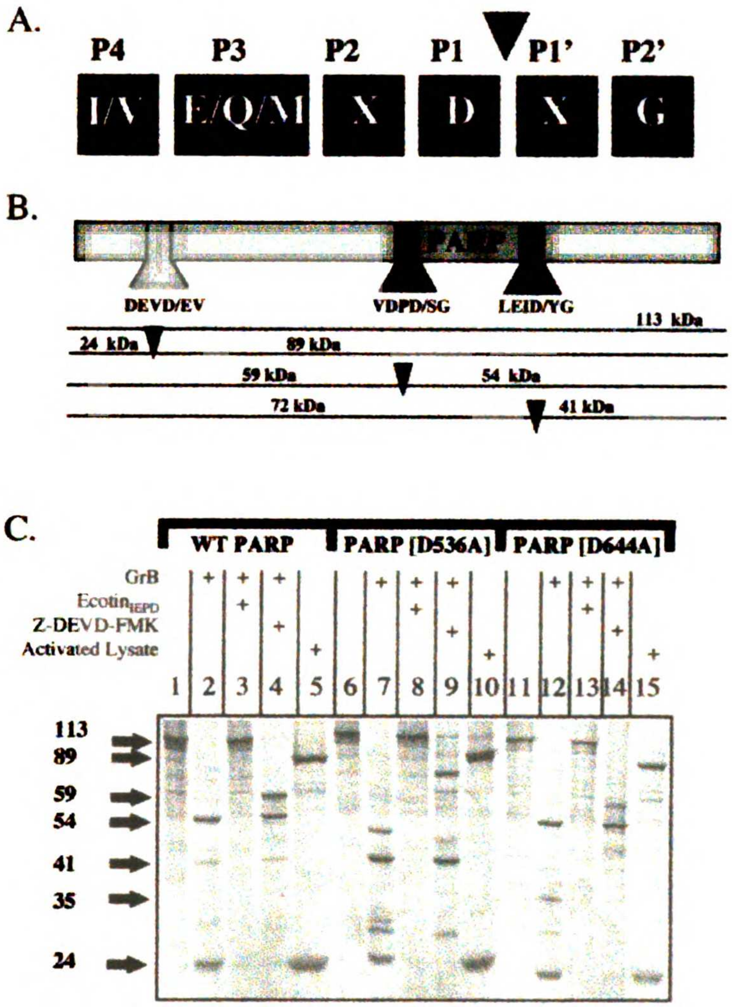
A. The optimal substrate specificity as determined by this study is shown.

B. Poly (ADP-ribose) polymerase (PARP) cleavage patterns resulting caspase and granzyme B incubation (see Results).

C. Cleavage of wild type PARP, PARP [D536A], and PARP [D644A] by granzyme B in rabbit reticulocyte lysate. $\text{Ecotin}_{\text{IEPD}}$, a specific granzyme B inhibitor and Z-DEVD-FMK, a specific caspase inhibitor was used to differentiate the various protease activities in the lysate (see Results).

D. The activation cleavage sequence of caspases 1-10 is shown. The correspondence of the activation sequence of several of the caspases to the optimal substrate specificity of granzyme B provides evidence that caspase activation may be a crucial role for granzyme B.

Figure 5-5.


A.
P4
P3
P2
P1
P1'
P2'
I/V
E/Q/M
X
D
X
G
B.
PARP
DEVD/EV
VDPD/SG
LEID/YG
113 kDa
24 kDa
89 kDa
59 kDa
54 kDa
72 kDa
41 kDa
C.
WT PARP
PARP [D536A]
PARP [D644A]
GrB
Ecotin$_{IEPD}$
Z-DEVD-FMK
Activated Lysate
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
113
89
59
54
41
35
24

D.

Pro-Caspase activation sequence:

	P4	P3	P2	P1	P1'	P2'
Caspase 2	D	E	T	D	R	G
Caspase 3	I	E	T	D	S	G
Caspase 6	T	E	V	D	A	A
Caspase 7	I	Q	A	D	S	G
Caspase 8	V	E	T	D	S	E
Caspase 9	P	E	P	D	A	T
Caspase 10	I	E	A	D	A	L
Caspase 1	W	F	K	D	S	V
Caspase 4	W	V	R	D	S	P
Caspase 5	W	V	R	D	S	P

UCSF LIBRARY

Table 5-1
Purification of recombinant rat granzyme B

	Volume (mL)	Total Protein (mg)	Total Granzyme B (mg)	Purity (%)
Culture Supernatant	10,000	1020	7	0.7
SP-Sepharose Fractions	20	75.2	6.7	8.9
Mono-S Fractions	2	6.8	6.7	98.5

Table 5-2
Steady-state kinetic parameters for the hydrolysis of substrates by granzyme B

Substrate	k_{cat} (s^{-1})	K_m x 10^{-6} (M)	k_{cat}/K_m (s^{-1} M^{-1})
Ac-IEPD-pNA	4.16 ± 0.05	57 ± 4	66600 ± 3200
Ac-EPD-pNA	1.20 ± 0.07	1430 ± 220	650 ± 130
Ac-PD-pNA	No Detectable Activity		
Ac-D-pNA	No Detectable Activity		
Suc-AAPD-pNA	≥ 0.2[a]	≥ 2000[a]	62 ± 5
Ac-IKPD-pNA	2.95 ± 0.33	560 ± 120	5100 ± 50
Ac-IEPD-AMC	0.536 ± 0.033	160 ± 30	3330 ± 60
Ac-IKPD-AMC	0.0226 ± 0.0003	802 ±16	23.3 ± 1.7
Boc-AAD-SBzl	11.4 ± 1.8	140 ± 30	82600 ± 2600
Ac-IEPD-SBzl	13.8 ± 0.3	43 ± 3	323300 ± 33300
Ac-IKPD-SBzl	7.67 ± 0.40	29 ± 6	268300 ± 15000
Ac-IEPDWGA-NH_2	0.00353 ± 0.00031	70 ± 20	52.7
Ac-IEPDWNA-NH_2	0.00287 ± 0.00022	300 ± 60	9.4

[a] Saturation was not observed at 2 mM substrate concentration.

Table 5-5

Steady-state kinetic parameters for the hydrolysis of substrates by granzyme B and the variants [R192A] and [R192E]

	k_{cat} (s^{-1})	K_m x 10^{-6} (M)	k_{cat}/K_m (s^{-1} M^{-1})
	Ac-IEPD-SBzl		
GrB WT	13.8 ± 0.3	43 ± 3	323300 ± 33300
GrB [R192A]	6.0 ± 1.5	95 ± 6	62900 ± 7800
GrB [R192E]	7.3 ±1.1	200 ± 67	36800 ±1100
	Ac-IKPD-SBzl		
GrB WT	7.67 ± 0.40	29 ± 6	268300 ± 15000
GrB [R192A]	2.67 ± 0.08	13 ± 2	208700 ± 3900
GrB [R192E]	4.47 ± 0.12	32 ± 4	140000 ± 6600
	Ac-IEPD-AMC		
GrB WT	0.536 ± 0.033	160 ± 30	3330 ± 60
GrB [R192A]	0.067 ± 0.005	1210 ± 130	55 ± 4
GrB [R192E]	0.029 ± 0.002	3280 ± 320	8.8 ± 1.3
	Ac-IKPD-AMC		
GrB WT	0.0226 ± 0.0003	802 ±16	23.3 ± 1.7
GrB [R192A]	0.0062 ± 0.0010	530 ± 100	8.3 ± 0.9
GrB [R192E]	0.0234 ± 0.0008	1480 ± 70	16.7 ± 0.8

Table 5-3
Mechanistic kinetic parameters for the hydrolysis of amide substrates by granzyme B and variants[a]

	K_d x 10^{-6} (M)	k_3 (s^{-1}) acylation	k_5 (s^{-1}) deacylation
		Ac-IEPD-AMC	
WT GrB	167	0.558	13.8
GrB [R192A]	1220	0.068	6.0
GrB [R192E]	3290	0.028	7.3
		Ac-IKPD-AMC	
WT GrB	805	0.023	7.7
GrB [R192A]	530	0.006	2.7
GrB [R192E]	1490	0.024	4.5

[a] Mechanistic parameters for granzyme B hydrolysis of amide substrates are derived from the steady-state kinetic constants in Table II. Assuming that for the hydrolysis of the thiobenzyl ester substrates, Ac-IEPD-SBzl and Ac-IKPD-SBzl, the deacylation step is completely rate determining, the microscopic constants for hydrolysis of the corresponding amide substrates can be determined (see Results).

Table 5-4
P3, P1', and P2' substrate specificity of Granzyme B determined by substrate phage

Clone	P3	P1'	P2'
1	Glu	Leu	Gly
2	Ser	Ser	Gly
3	Glu	Trp	Gly
4	Ser	Phe	Gly
5	Met	Phe	Gly
6	Glu	Met	Gly
7	Glu	Ile	Gly
8	Glu	Leu	Gly
9	Glu	Ser	Gly
10	Glu	Trp	Gly
11	Met	Ser	Gly
12	Glu	Trp	Ala
13	Val	Ser	Gly
14	Thr	Ser	Thr
15	Met	Ala	Ala
16	Asp	Val	Gly
17	Ile	Trp	Gly
18	Glu	Leu	Gln
19	Met	Ser	Gly
20	Glu	Val	Gly

Chapter Six:

Identification of macromolecular substrates of granzyme B by a small pool cDNA expression cleavage screen

Abstract:

Intracellular proteolysis is a hallmark of all types of apoptosis, regardless of the inducer. Proteolysis of target substrates is typically due to the activity of caspases, intracellular cycteine proteases. However, recent studies have indicated a role for a caspase-independent mechanism in the case of cytotoxic lymphocyte-induced cell death. Granzyme B, a serine protease found in the granules of cytotoxic lymphocytes, has been implicated in mediating both caspase-dependent and caspase-independent pathways of target-cell death. The current study presents the method of small pool cDNA expression screening for the rapid identification of macromolecular substrates for granzyme B. Multiple potential macromolecular substrates were isolated from this screening process including nuclear lamin B, c-abl, polyA-binding protein, Notch, Br140, fibroblast growth factor receptor I, heterogenous nuclear ribonucleoprotein H', heat shock protein-70kD. Deconvolution of the cDNA library pool and characterization of the granzyme B cleavage site in lamin B is presented. Lamin B has a cleavage site that correlates to the optimal cleavage sequence determined for granzyme B. Cleavage of lamin B is not dependent on caspase activation but can be cleaved by granzyme B directly.

Abbreviations:

GrB, granzyme B; CL, cytotoxic lymphocyte; c-abl, abelson kinase; FGFR1, fibroblast growth factor receptor 1; HNRNP H'; heterogenous nuclear ribonuclear protein H'; hsp70, heat shock protein 70 kDa.

Introduction:

Cytotoxic lymphocytes initiate the death of virus infected- and tumor-cells through the release of granule contents at the cellular interface or through the association of fas ligand with fas receptor (Podack, 1995). Both mechanisms lead to activation of the caspases, intracellular cysteine proteases. The caspases mediate cell apoptosis through the cleavage of selected cellular substrates.

Recently, the substrate specificity of granzyme B, a cytotoxic lymphocyte granule protease, has been determined to six subsites, from P4-P2' (Harris *et al.*, 1998). Based on this preferred cleavage sequence, Caspase 3 and Caspase 7 are believed to be the main caspase activation targets for granzyme B. However, other cellular proteins contain the preferred granzyme B cleavage sequence and may also be target substrates of granzyme B.

A systematic method was used to screen the proteome for potential granzyme B substrates. In this method, a cDNA library is divided into smaller libraries that are transcribed and translated *in vitro* in the presence of ^{35}S labeled methionine (Lustig *et al.*, 1997). The labeled protein pools are then incubated with granzyme B. Those pools containing proteins specifically cleaved by granzyme B are deconvoluted to a single cDNA sequence. Using this *in vitro* method we have identified potential macromolecular substrates of granzyme B.

Experimental Materials and Methods:

Reagents: The coupled transcription-translation reticulocyte lysate system was purchased from Promega (Madison, WI). The caspase inhibitor, Z-DEVD-Fmk, was purchased from Enzyme Systems Products (Livermore, CA). The caspase inhibitor Z-VAD-Fmk was purchased from Calbiochem (San Diego, CA). The caspase inhibitors, Ac-DEVD-fmk and Ac-IETD-fmk were purchased from Enzyme Systems Products (Livermore, CA). The activating monoclonal antibody to fas (CH-11) was purchased from Kamiya Biomedical

Company (Seattle, WA). Antibodies to Lamin B (C-20) and PARP (H-250) were purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). Antibodies to caspase-3 (65906E) were purchased from PharMingen. Rat granzyme B was heterologously expressed in the yeast *Pichia pastoris* as previously described (Harris *et al.*, 1998).

Small pool c-DNA screening: A c-DNA library was constructed from RNA isolated from a 14-day postcoitus mouse embryo as previously described (Kothakota *et al.*, 1997). Briefly, the c-DNA library was cloned into the high copy number plasmid pCS2$^+$ (Turner and Weintraub, 1994) to yield 200,000 independent clones. The library was divided into pools of approximately 100 unique plasmids.

Methionine (^{35}S)-labeled protein pools were prepared directly from the cDNA pools using the coupled transcription-translation reticulocyte lysate system (Lustig *et al.*, 1997). Briefly, 5 µL of transcription-translation mix was incubated at 30°C with 0.1 µg of pooled DNA. After two hours, half of the labeled-protein reaction mix was incubated with 1 nM recombinant granzyme B, 50 µM Z-VAD-Fmk, 50 µM Ac-DEVD-fmk, and 50 µM Ac-IETD-fmk in a buffer containing 100 mM Hepes, pH 7.4 and 100 mM NaCl. After 1 hour, loading dye was added to each reaction mix and heated at 65°C for 10 minutes. The samples were loaded side by side onto 10-20% Tris-Glycine gels and separated by SDS-PAGE. Separated proteins were visualized by autoradiography and scored for cleavage by the disappearance or appearance of bands in the granzyme B incubated samples that were not found in the control samples.

Whole cell substrate cleavage by recombinant granzyme B: Jurkat T cells (clone E6-1, ATCC, Rockville, MD) were grown in RPMI 1640 supplemented with 10% fetal bovine serum. On the day of the experiment the cells were harvested and washed 3 times in buffer containing 2.7 mM KCl, 1.5 mM KH_2PO_4, 137 mM NaCl, 8mM Na_2HPO_4, 1 mM $CaCl_2$. The cells were resuspended at a concentration of 3×10^7 per mL in the absence or presence of 50 μM caspase inhibitors, Z-DEVD-Fmk and Z-VAD-Fmk. Cell death through the fas pathway was induced by the addition of the CH-11 activating antibody to fas at a concentration of 100 ng/mL. Cell death through the granzyme B pathway was induced by the addition of recombinant granzyme B at a concentration of 50 nM. Aliquots of were removed at 0, 1, 5, 20, 60, and 120 minutes. An equal volume of RIPA-SDS buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.0), 5 mM Na_2EDTA, 100 mM NaF, 200 _M Na_3VO_4 (ortho), 1% (vol/vol) Triton X-100, 0.1% (wt/vol) deoxycholic acid, 1 mM DTT, 1 mM Phenylmethylsulfonyl fluoride, 1 μg/mL leupeptin, 1 μg/mL pepstatin, 50 mM benzamidine-HCl) and mechanically homogenized with a 22 gauge needle. The insoluble fraction was removed by centrifugation. The resulting cellular proteins were resolved on SDS-PAGE and transferred to nitrocellulose membranes. The transferred proteins were probed with antibodies to Lamin B and Caspase-3.

Results:

Two hundred thousand genes were separated into 2000 pools of approximately one hundred genes (Figure 6-1A & B). The genes were translated to protein products in an *in vitro* transcription/translation system in the presence of ^{35}S-methionine. The labeled protein samples were then incubated with granzyme B (Figure 6-1C). Recently, we have shown that rabbit reticulocyte lysate contains a latent pro-caspase 3-like activitiy that is activated upon addition of granzyme B (Harris *et al.*, 1998). To avoid ascribing cleavage of substrates by caspase 3 to granzyme B, caspase inhibitors were included in the incubation reaction. The incubated labeled protein products were then separated by SDS-PAGE and visualized by autoradiography. Those pools displaying cleaved protein products were then transformed into *E. coli* to return to a clonal population. Individual clones and DNA were isolated and tested for cleavage by granzyme B. The plasmid DNA from the clones encoding proteins susceptible to GrB cleavage were then sequenced

(Figure 6-1D). In this screen, fourteen of the cDNA pools contained potential granzyme B substrates. Of the fourteen clones, eight were identified proteins in the protein databank. These proteins were: nuclear lamin B, c-abl, Br140, FGFR1, polyA-binding protein, heterogenous nuclear ribonucleoprotein H', notch, and hsp70. Here we describe in detail the isolation and characterization of one of these proteins, nuclear lamin B.

The concentration of recombinant granzyme B used in the cleavage screen was determined by cleavage of the known granzyme B substrate, caspase 3, at the cleavage site IETD/SG (Figure 6-2A). Caspase 3 was transcribed and translated *in vitro* and incubated with varying amounts of granzyme B. The concentration of *in vitro* translated products is approximately femtomolar, which is below the K_m for granzyme B. Therefore, assuming that the substrate concentration is constant, cleavage of the translated products by granzyme B is determined by the enzyme concentration and the specificity constant, k_{cat}/K_m. Under these conditions, 1nM granzyme B processed approximately half of the caspase 3 translated product. One nanomolar granzyme B was used throughout the screen and cleavage of caspase 3 was routinely monitored as a positive control.

Cleavage by granzyme B of the protein products in pool 50.13 led to the disappearance of a band at 43 kDa with the appearance of a band at 40 kDa (Figure 6-3A--Pool 50.13). The plasmids containing the cDNA from pool 50.13 were transformed into *E. coli.* Single clones were selected and grown in a 96-well plate. Twelve of the colonies were combined into eight pools and tested for cleavage with granzyme B. The cleaved protein was found in pool 50.13.7 (Figure 6-3B). The twelve clones from pool 50.13.7 were amplified and retested for cleavage with granzyme B. This led to the isolation of a single gene product in pool 50.13.7.6 (Figure 6-3C). The plasmid containing the isolated cDNA insert was sequenced (Figure 6-3D) and determined to be a partial transcript of nuclear lamin B.

The lamin B transcript lacked the N-terminal portion of the protein with initiation taking place at methionine 212 instead of the start methionine (Figure 6-4, internal start methionine underlined). Based on the sequence, the resulting protein product should migrate with a molecular mass of 43kDa, which corresponds to that seen on the SDS-PAGE (Figure 6-3). Analysis of the protein sequence revealed an optimal granzyme B cleavage site of VEVD/SG at position 232 (Figure 6-4, large font). Cleavage at this site would result in a loss of 3 kDa from the partial transcript protein product as seen in the cleavage screen (Figure 6-3).

Proteolytic degradation of nuclear lamins has been seen in cells induced to undergo apoptosis with a variety of stimuli. The intracellular protease believed to be responsible for lamin cleavage is caspase 6 (Figure 6-2B). Caspase 6 shares a similar substrate specificity profile with granzyme B, as determined by tetrapeptide substrate analysis. A redundant function for lamin degradation may be displayed by granzyme B in cells induced to undergo apoptosis by cytotoxic lymphocytes. To test the ability of granzyme B to directly cleave lamin B, Jurkat cell lysate was incubated with granzyme B in the presence of several fluoromethyl ketone caspase inhibitors. Activation of of the fas/fasL mechanism resulted in the degradation of lamin B (Figure 6-5), as previously reported by others (Orth *et al.*, 1996). Fas/fasL activated cleavage of Lamin B was abolished in the presence of the broadly specific caspase inhibitor, Z-VAD-Fmk, the effector-specific caspase inhibitor, Z-DEVD-Fmk, and the activator-specific caspase inhibitor Z-IETD-Fmk (Figure 6-5). In contrast to the absence of cleavage of lamin B in the presence of caspase inhibitors by fas/fasL activation, granzyme B efficiently cleaved lamin B (Figure 6-5).

While granzyme B cleaves *in vitro* transcribed/translated caspase-3 as efficiently as *in vitro* transcribed/translated lamin B (Figure 6-6), in whole cell lysate, granzyme B cleaves caspase-3 more rapidly than lamin B (Figure 6-7).

Discussion:

Identifying downstream macromolecular substrates for granzyme B will expand the definition of its role in the induction of target-cell death. Interaction between a protease and its substrates is a special subset of general protein-protein interactions. The same methods used to detect protein-protein interactions will not necessarily be sucessful with protease-substrate interactions. This is because most of the methods depend on a strong ground state binding potential. Serine proteases have evolved to destabilize ground state interactions with substrate in preference to transition state stabilization. Preferential transition-state binding to tetrapeptide substrates has been previously demonstrated with granzyme B (Harris *et al.*, 1998). Therefore it is necessary when searching for granzyme B substrates to use a method that monitors protease catalysis.

The method described here uses the proteolytic activity of granzyme B to screen a mouse embryonic cDNA expression library. This macromolecular screening method resulted in the identification of several granzyme B target proteins. These proteins can be grouped into three categories: (1) proteins that have previously been identified as caspase substrates and share the same cleavage site with granzyme B (lamin B); (2) cleavage of proteins previously identified as modulating apoptosis, but not shown to be caspase or granzyme B substrates (c-abl (Agami *et al.*, 1999; Daniel *et al.*, 1998; Daniel *et al.*, 1996), hsp70 (Dix *et al.*, 1996; Mosser *et al.*, 1997), FGFR1 (Kanda *et al.*, 1999; Legeai-Mallet *et al.*, 1998), notch (Jehn *et al.*, 1999; Shelly *et al.*, 1999)) and (3) proteins that have not yet been shown to be involved in or degraded during apoptosis (Br140, PolyA binding protein, HNRNP H'). Here we describe granzyme B's proteolytic activity against lamin B, a substrate that has previously been shown to be a caspase-6 substrate.

The nuclear lamina forms a lattice of lamin intermediate filament proteins on the inner surface of the nuclear membrane. Lamins have been suggested to serve as major chromatin achoring sites and to possibly be involved in organizing higher order chromatin domains. The nucleus is a site of major morphological changes during apoptosis with nuclear condensation and internucleosomal fragmentation of DNA being a hallmark of apoptosis. Destruction of the nuclear lamina is postulated to be important in promoting nuclear breakdown during apoptosis (Neamati *et al.*, 1995; Rao *et al.*, 1996).

Recent studies have demonstrated that granzyme B is localized to the nucleus in target cells during cytotoxic lymphocyte mediated cell death (Trapani *et al.*, 1998). The observation that granule mediated cytotoxicity can result in cell death in the absence of caspase activity (Sarin *et al.*, 1997; Trapani *et al.*, 1998) is consistent with granzyme B acting directly on nuclear substrates rather than indirectly through caspase activation. It is interesting to note that the majority of the macromolecular proteins identified in this screen are nuclear proteins or shuttle from the cytoplasm to the nucleus. Yet to be determined are whether these proteins are true substrates of granzyme B during cytotoxic lymphocyte mediated-cell death and what the consequences of that cleavage are.

Figure 6- 1.

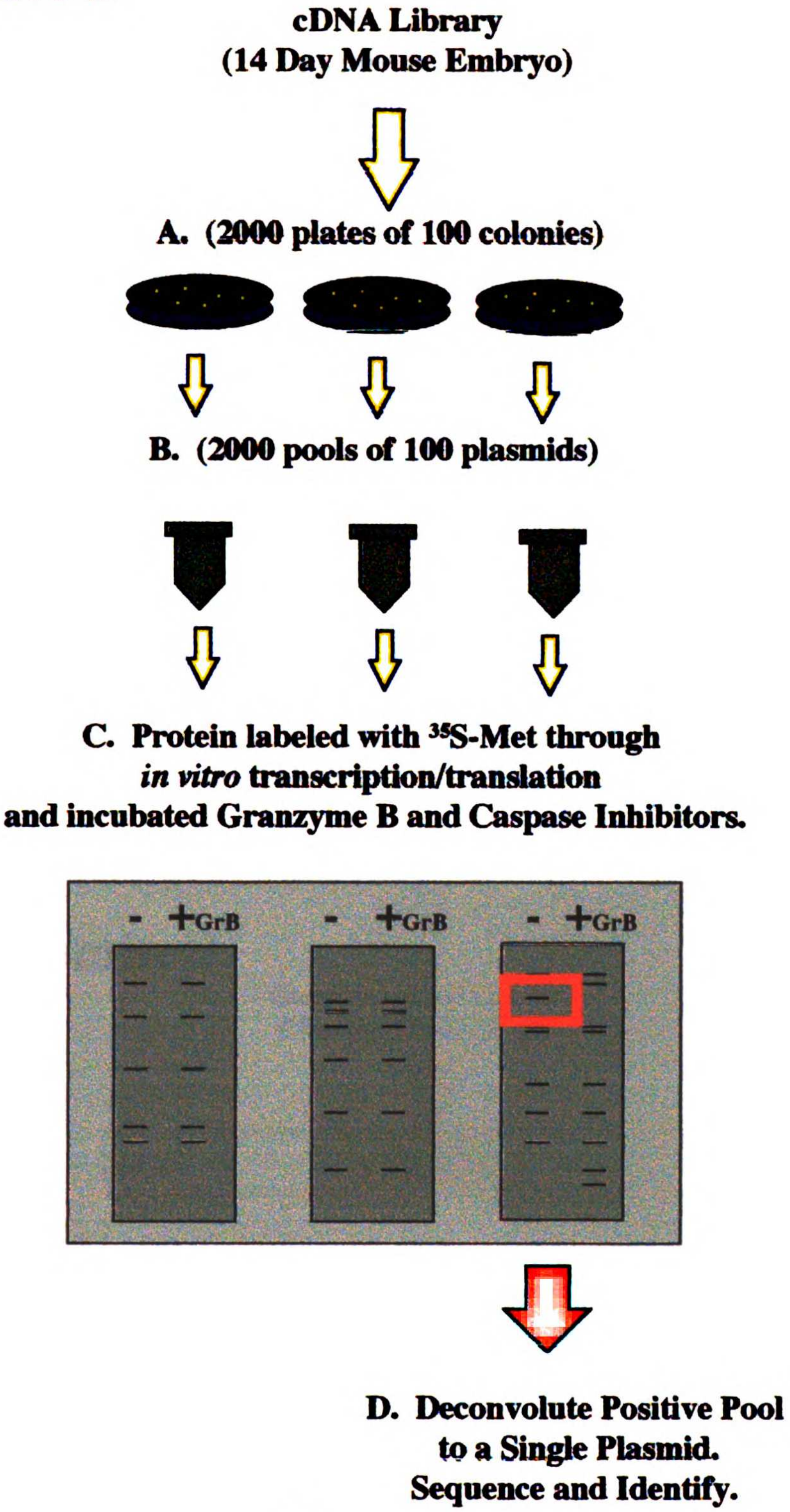

cDNA Library
(14 Day Mouse Embryo)
A. (2000 plates of 100 colonies)
B. (2000 pools of 100 plasmids)
C. Protein labeled with 35S-Met through
in vitro transcription/translation
and incubated Granzyme B and Caspase Inhibitors.
- +GrB
- +GrB
- +GrB
D. Deconvolute Positive Pool
to a Single Plasmid.
Sequence and Identify.

Figure 6- 2.
Cleavage Fragments of Lamin B
and Caspase 3 by Granzyme B and Caspases.

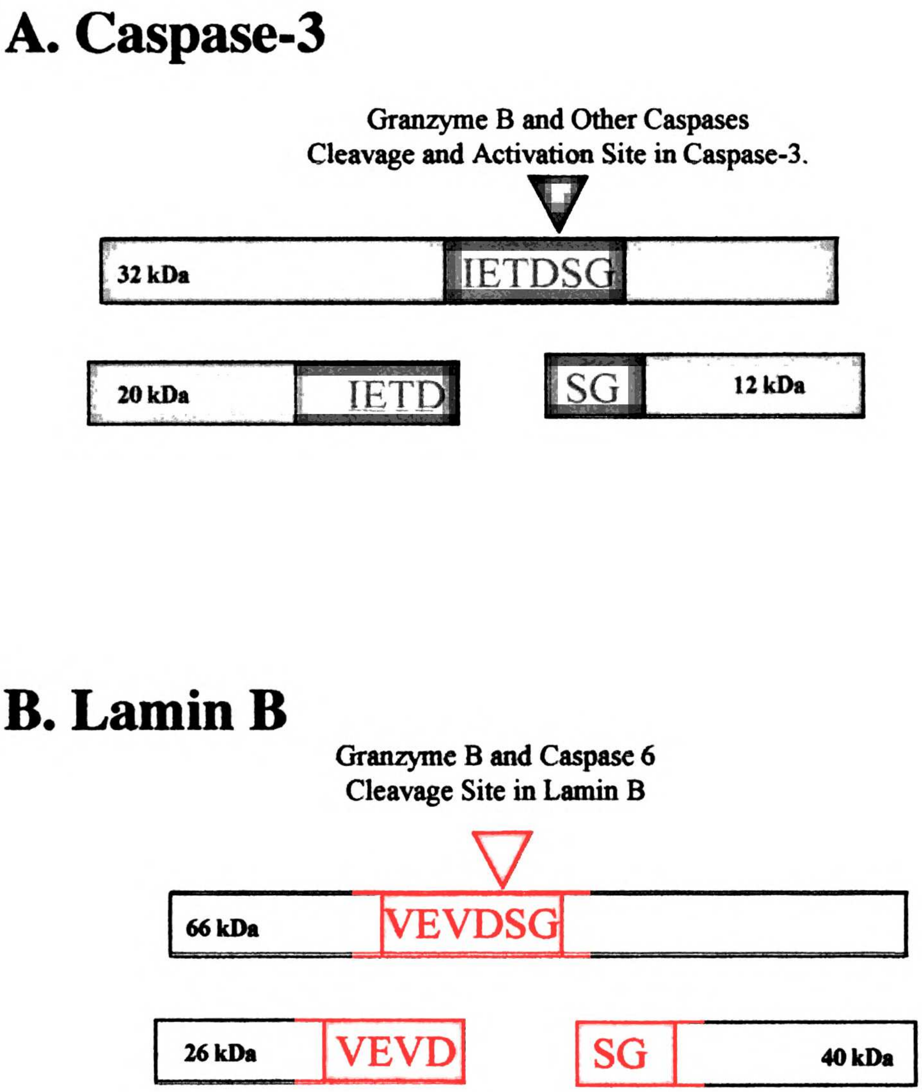

A. Caspase-3
Granzyme B and Other Caspases
Cleavage and Activation Site in Caspase-3.
32 kDa
IETDSG
20 kDa
IETD
SG
12 kDa
B. Lamin B
Granzyme B and Caspase 6
Cleavage Site in Lamin B
66 kDa
VEVDSG
26 kDa
VEVD
SG
40 kDa

Figure 6-3.

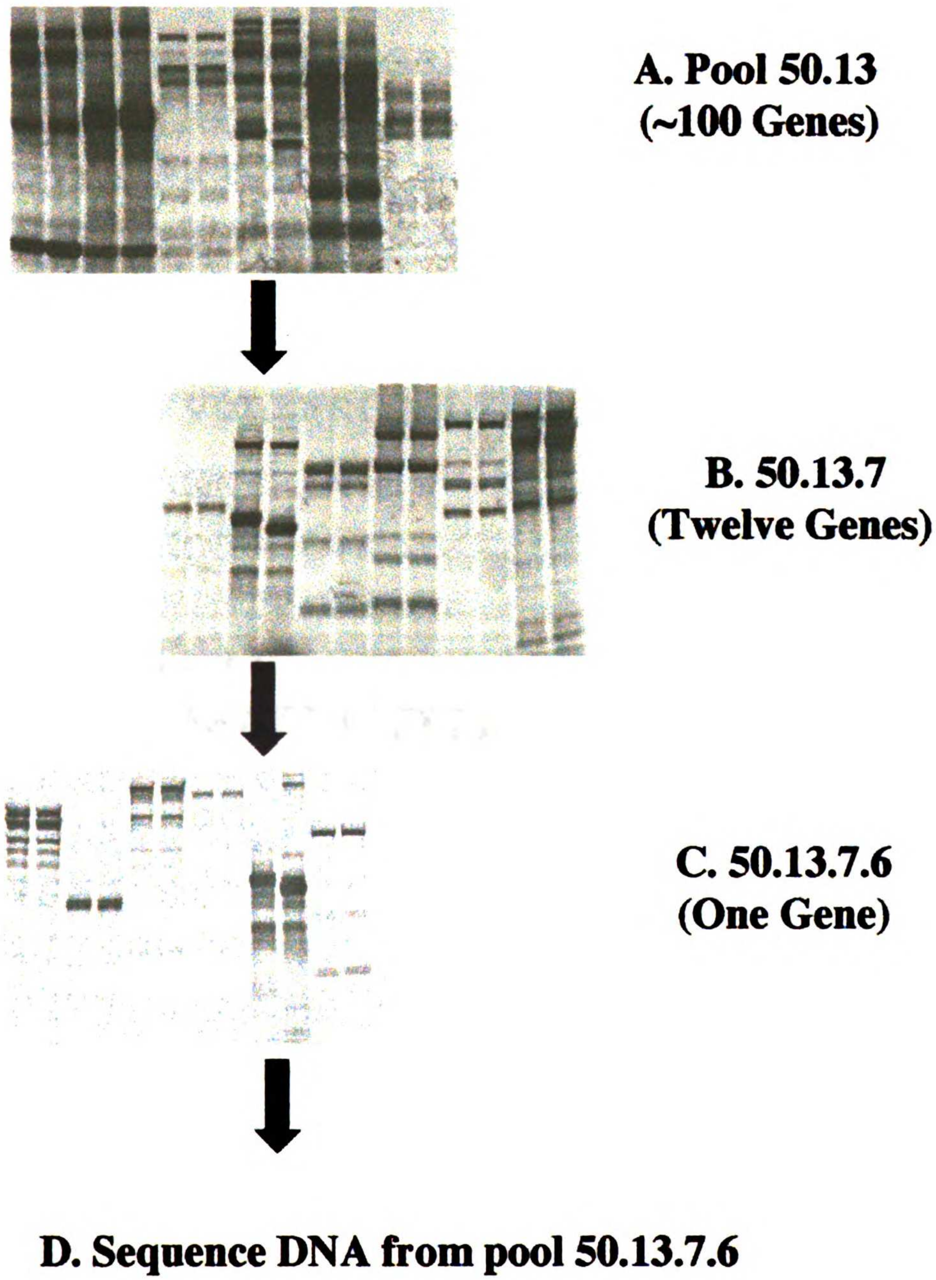

A. Pool 50.13
(~100 Genes)
B. 50.13.7
(Twelve Genes)
C. 50.13.7.6
(One Gene)
D. Sequence DNA from pool 50.13.7.6

Figure 6-4. Granzyme B Cleavage Site in Nuclear Lamin B:

MATATPVQQQRAGSRASAPATPLSPTRLSRL
QEKEELRELNDRLAVYIDKVRSLETENSALQ
LQVTEREEVRGRELTGLKALYETELADARRA
LDDTARERAKLQIELGKFKAEHDQLLLNYAK
KESDLSGAQIKLREYEAALNSKDAALATALG
DKKSLEGDLEDLKDQIAQLEASLSAAKKQLA
DETLLKVDLENRCQSLTEDLEFRKNMYEEEI
NETRRKHETRLVEVD▽SGRQIEYEYKLAQA
LHEMREQHDAQVRLYKEELEQTYHAKLENAR
LSSEMNTSTVNSAREELMESRMRIESLSSQL
SNLQKESRACLERIQELEDMLAKERDNSRRM
LSDREREMAEIRDQMQQQLSDYEQLLDVKLA
LDMEISAYRKLLEGEEERLKLSPSPSSRVTV
SRASSSRSVRTTRGKRKRVDVEESEASSSVS
ISHSASATGNVCIEEIDVDGKFIRLKNTSEQ
DQPMGGWEMIRKIGDTSVSYKYTSRYVLKAG
QTVTVWAANAGVTASPPTDLIWKNQNSWGTG
EDVKVMLKNSQGEEVAQRSSVFKTTIPEEEE
EEEEPIGVAVEEERFHQQGAPRAWNKSCAIM

Figure 6- 5

***In vitro* cleavage of Lamin B**

Granzyme B as fas activation.

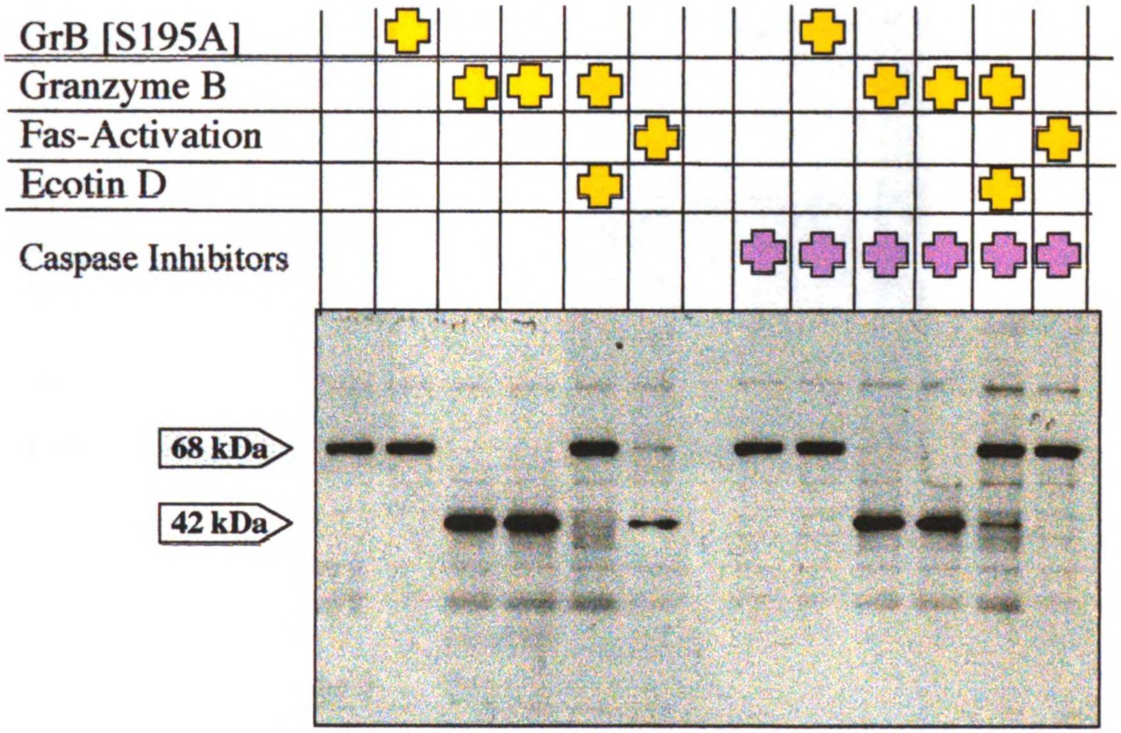
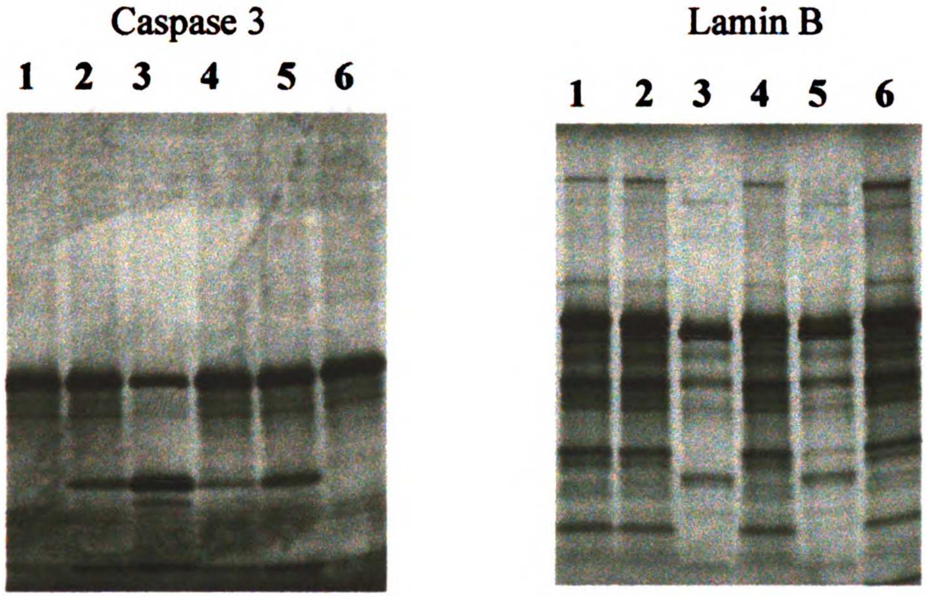

GrB [S195A]
Granzyme B
Fas-Activation
Ecotin D
Caspase Inhibitors
68 kDa
42 kDa

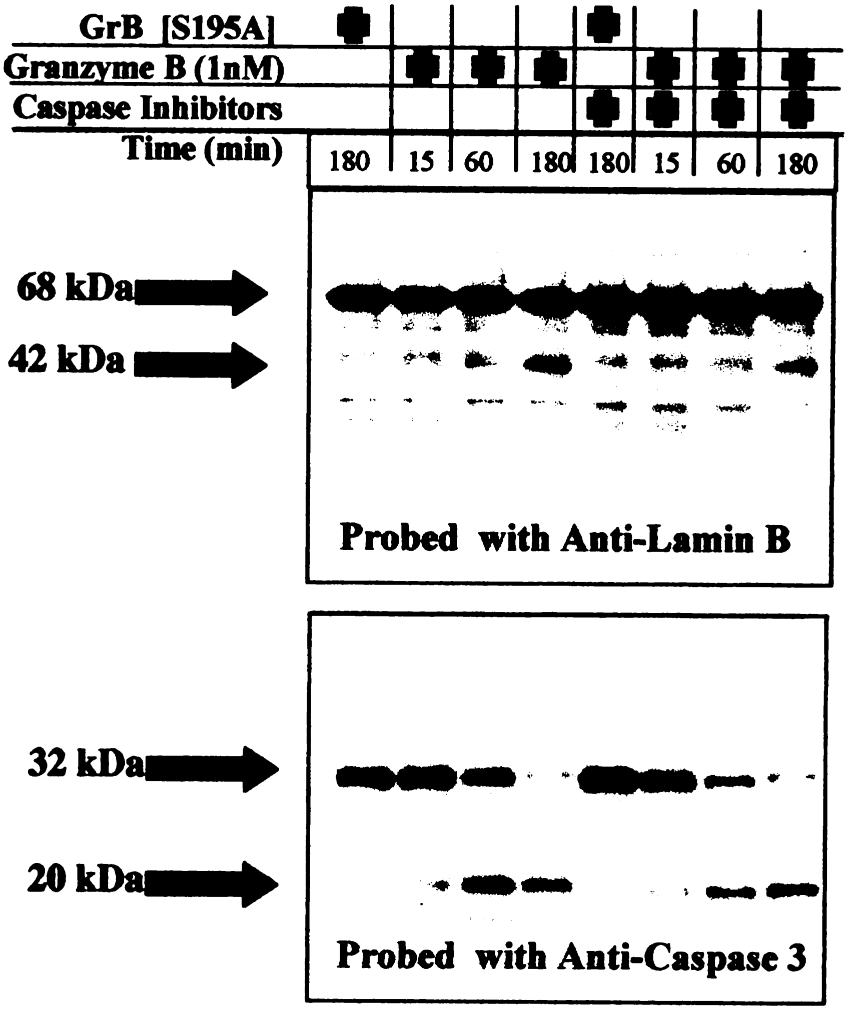
Figure 6-6. Cleavage of *in vitro* transcribed/tranlated ^{35}S-Met labeled Caspase-3 and Lamin B in presence and absence of caspase activity:


Caspase 3

Lamin B

Lane 1	**time = 0 minutes**
Lane 2	**time = 30 minutes** **1 nM Granzyme B**
Lane 3	**time = 180 minutes** **1 nM Granzyme B**
Lane 4	**time = 30 minutes** **1 nM granzyme B** **Caspase Inhibitors**
Lane 5	**time = 180 minutes** **1 nM Granzyme B** **Caspase Inhibitors**
Lane 6	**time = 180 minutes** **Granzyme B [S195A]**

Figure 6- 7.
***In Vitro* Cleavage of Lamin B**
and Caspase 3 by Granzyme B:


GrB [S195A]
Granzyme B (1nM)
Caspase Inhibitors
Time (min)
180
15
60
180
180
15
60
180
68 kDa
42 kDa
Probed with Anti-Lamin B
32 kDa
20 kDa
Probed with Anti-Caspase 3

Chapter Seven:

Mouse and human granzyme A display different extended substrate specificities

Abstract:

Granzyme A is the most abundantly expressed serine protease in the granules of cytotoxic lymphocytes (Shresta *et al.*, 1999). Human and rodent cytotoxic lymphocytes share the serine proteases, granzyme A, granzyme B, granzyme K, and Metase. However, humans have an additional B-class protease not shared with rodents, granzyme H, and rodents have several additional B-class proteases not shared with human, granzyme C-G. What accounts for this difference between species? Do equivalent granzymes have equivalent functions in each species? Or do they play species-specific roles that reflect the different evolutionary pressures between humans and rodents? One method to address the species difference is through the functional characterization of the "equivalent" enzymes. Using the method of positional-scanning synthetic combinatorial libraries, the extended substrate specificity of human and mouse granzyme A were defined. These enzymes are homologous with over 70% sequence identity between them. While both enzymes shared a P1-preference for basic amino acids, arginine over lysine, the extended P4-P2 substrate specificity profiles were not commensurate. The optimal P4-P2 substrate specificity for human granzyme A was (Ile/Val)-(Ala/Gly/Ser)-(Asn), and that for mouse granzyme A was (Gly>Val/Ile)-(Tyr/Phe)-(Phe). Single substrate kinetic analysis confirmed the preferences observed in the library. The lack of "functional-homology" demonstrated through differing substrate specificity profiles suggests a non-equivalent role for granzyme A in humans and rodents.

Materials and Methods:

N-Glycosidase F was purchased from Boehringer Mannheim (Indianapolis, IN). All DNA modifying enzymes were purchased from New England Biolabs (Beverly, MA) or Stratagene (La Jolla, CA) and were used according to manufacturer's guidelines. TNT® coupled reticulocyte lysate system was purchased from Promega (Madison, WI) and used according to manufacturer's guidelines. Protein assay Bradford reagent was purchased from Bio-rad (Hercules, CA) and used according to manufacturer's guidelines. Substrates in the positional scanning synthetic combinatorial library (PS-SCL) were prepared as previously described in chapter 4. The *Pichia pastoris* expression system was purchased from Invitrogen (San Diego, CA).

Heterologous expression of mouse and human granzyme A in yeast.

The Xho I and Not I DNA recognition sites were introduced by PCR onto the 5' and 3' ends, respectively, of the 705 base-pair cDNA encoding mature human granzyme A, and 693 base-pair cDNA encoding mature mouse granzyme A. The resulting fragments were subcloned into the Xho I and Not I sites of the yeast vector, pPICzαA (Invitrogen, San Diego, CA). This construct permitted the fusion of the mature granzyme B sequence to immediately follow the Kex2 signal cleavage site of the *Saccharomyces cerevisiae* α-factor secretion signal (Brake *et al.*, 1984). The vector was linearized with Sac I and transformed into the X33 strain of *Pichia pastoris*. Clones with the integrated human and mouse granzyme A cDNA were selected for by resistance to Zeocin™ (Wenzel *et al.*, 1992). Mouse and human granzyme A-expressing clones were isolated and used for large scale protein expression. Yeast growth and protein expression were maintained at pH 6.0, 30°C.

Purification of recombinant mouse and human granzyme A.

After 48 hours of induction with methanol, the supernatant from the granzyme A expressing culture was harvested. The supernatant was adjusted to 50 mM NaCl and loaded onto a 100 mL SP-sepharose cation exchange column (Pharmacia Biotech, Uppsala Sweden). The column was washed with five column volumes of 50 mM MES, pH 6.0 and 50 mM NaCl and eluted with a linear salt gradient of 50 mM to 1000 mM NaCl. Active granzyme B eluted at 400 mM NaCl. The fractions from the SP-sepharose column were pooled and dialyzed against 50 mM MES pH 6.0, 100 mM NaCl and loaded onto a 1 mL Mono-S® cation exchange column (Pharmacia Biotech, Uppsala, Sweden). The Mono-S column was washed with eight column volumes of buffer containing 50 mM MES pH 6.0, 100 mM NaCl. The column was then treated with a salt gradient from 100 mM to 800 mM to elute active granzyme B at a salt concentration of 400 mM NaCl. The final product was ≥98% pure as judged by SDS-PAGE Coomassie brilliant blue staining as described by Shägger and Jacow (Schagger and von Jagow, 1987). The final yield of mouse granzyme A was ~0.5 mg/L and that of human granzyme A was ~10 mg/L.

The concentration of mouse granzyme A protein was determined by absorbance measured at 280 nm and based on an extinction coefficient of 31520 M^{-1} cm^{-1} (1.22 mL mg^{-1}) and that of human granzyme A of 31520 M^{-1} cm^{-1} and 1.23 mL mg^{-1} (Gill and von Hippel, 1989).

Results and Discussion:

Granzyme A is unique within the granzyme family in that it is the only granzyme that exists as a dimer. Dimerization is facilitated by an intermolecular disulfide bond at position 93 (Figure 7-1). Mouse and human granzyme A share 70% sequence identity and 80% sequence similarity.

Both mouse and human granzyme A were expressed as active enzymes in the yeast, *Pichia pastoris*. Two major dimer bands, and two monomer bands were seen for mouse granzyme A on SDS-PAGE. Mouse granzyme A contains two potential glycosylation sites, Asn 146 and Asn 159. To determine the glycosylation state of recombinant mouse granzyme A, 20 μg of granzyme A was denatured by boiling for 10 minutes in the presence of 0.5% SDS and 1% β-mercaptoethanol. The denatured granzyme A was separated into two aliquots and one aliquot was incubated at 37°C for three hours with ten units of N-glycosidase F in 50 mM Tris, pH 7.5 and 1% Triton X-100. The incubated products were analyzed by SDS-PAGE and Coomassie brilliant blue staining (Figure 7-2A). Upon treatment with a glycosidase and under reducing conditions we see the four bands coalesce to a single band (Figure 7-2A).

To determine if granzyme A was an obligate dimer, gelatin gel zymography was run under native conditions and reducing conditions. The experiment with mouse granzyme A is shown (Figure 7-2B).

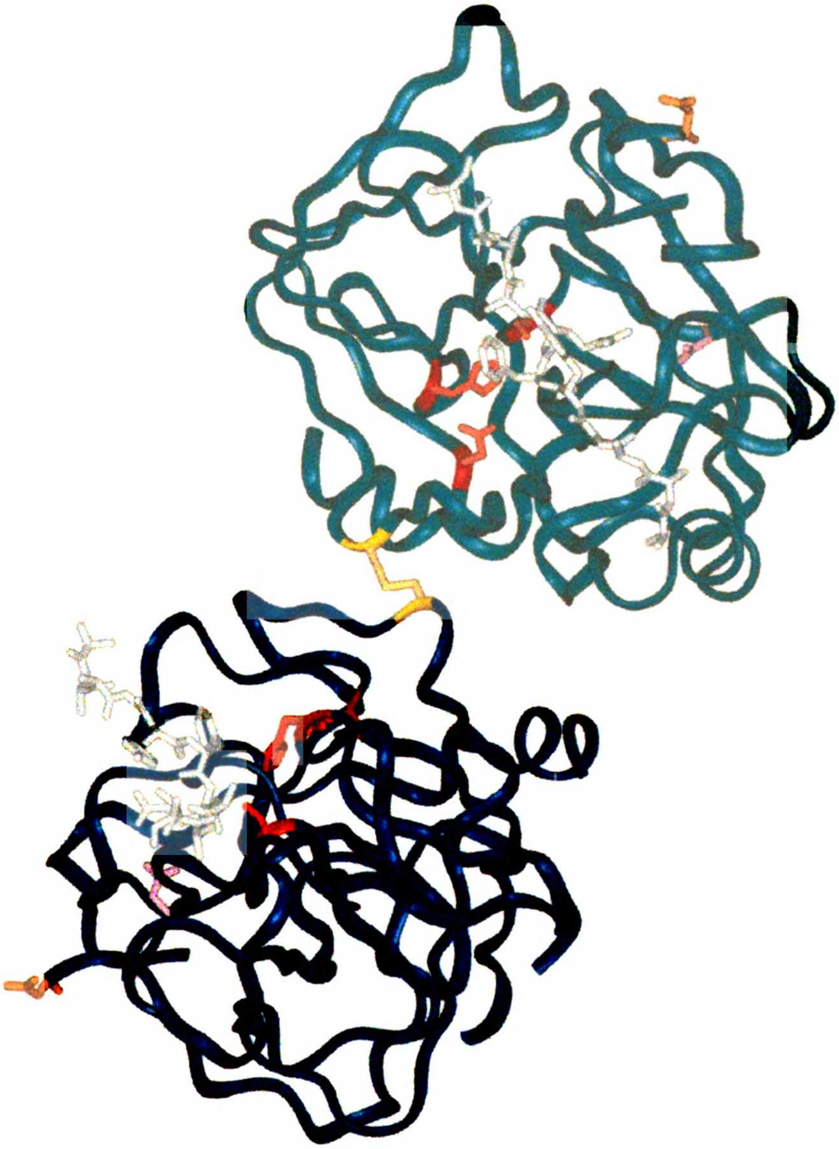
To solve the discrepancy between the actual molecular mass (25,814 g/mol/monomer) based on the human granzyme A amino acid sequence and the apparent molecular mass seen on SDS-PAGE (38 kDa/dimer and 23 kDa/monomer) (Figure 7-3), matrix assisted laser desorption ionization (MALDI) mass spectroscopy was collected on the recombinant enzyme (Figure 7-4). The mass spectrum demonstrates that human granzyme A is intact, with peaks at 25,890 g/mol and 52,032 g/mol.

To explore the functional differences between mouse and human granzyme A, specificity studies were initiated. Activity of human and mouse granzyme A against substrates in the P1-Diverse positional scanning-synthetic combinatorial library show that both enzymes possess primary specificity for basic amino acids, arginine over lysine (Figure 7-5). However, when the extended substrate specificity is probed, the enzymes display significantly different preferences (Figure 7-6 A & B). Mouse granzyme A has a P4-preference for glycine, a P3-preference for aromatic amino acids, and a P2-preference for the aromatic amino acid phenylalanine (Figure 7-6 A). Human granzyme A has a P4-preference for beta-branched aliphatic amino acids, valine and isoleucine, a P3-preference for small amino acids, alanine, glycine, and serine, and a P2 preference for asparagine (Figure 7-6 B).

What accounts for the functional substrate specificity differences in granzyme A between human and mouse? Perhaps mouse and human granzyme A have maintained their function in cytotoxic lymphocyte mediated cell death and their substrate specificities have evolved to accommodate the changing cleavage site sequences of the substrates. Perhaps mouse and human granzyme A have diverged in function to accommodate immune challenge of species-specific pathogens. Perhaps extended specificity is not important for the function of granzyme A. This functional discrepancy between human and mouse raises important questions as to the validity of mouse models to study human cytotoxic lymphocyte function. Further effort into the identification of granzyme A susbtrates, in both mouse and human, will shed light on its biological function. This data also begs for structural analysis of how enzymes that are 70% identical can have such disparate substrate specificities.

Figure 7-1. Three-dimensional dimeric model of granzyme A. The intermolecular disulfide bond (Cys 93) is shown in yellow, the catalytic triad in red, aspartic acid 189 in magenta, asparagine 161 (glycosylation site) in orange, and substrate in grey.

Figure 7- 1. Three Dimensional Model of dimeric Granzyme A.



Figu

glyc

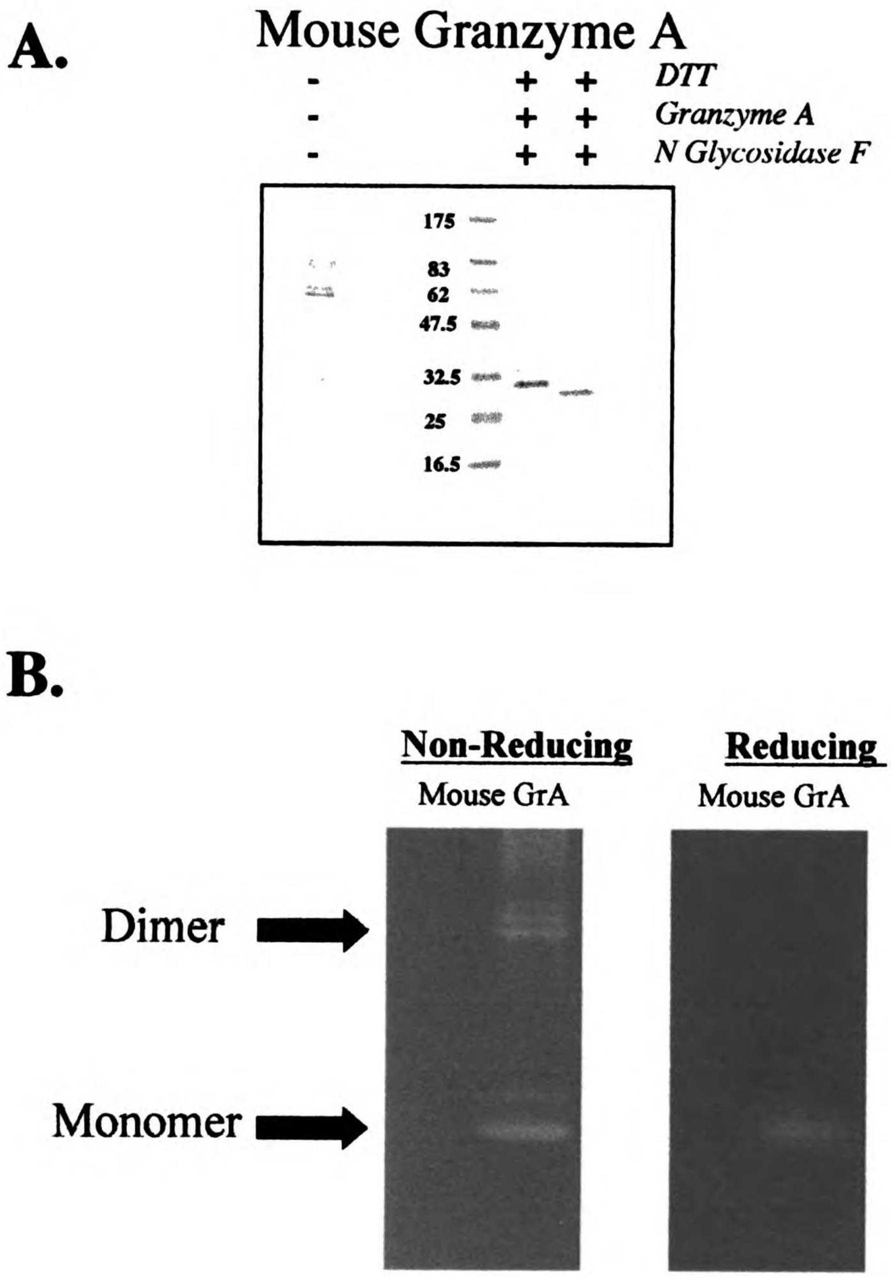
Fig

redu

Figure 7-2A. Coomassie stain of recombinant mouse GrA. Treated with and without N-glycosidase-F.

Figure 7-2 B. Gelatin Zymogram of mouse granzyme A, under non-reducing and reducing conditions.

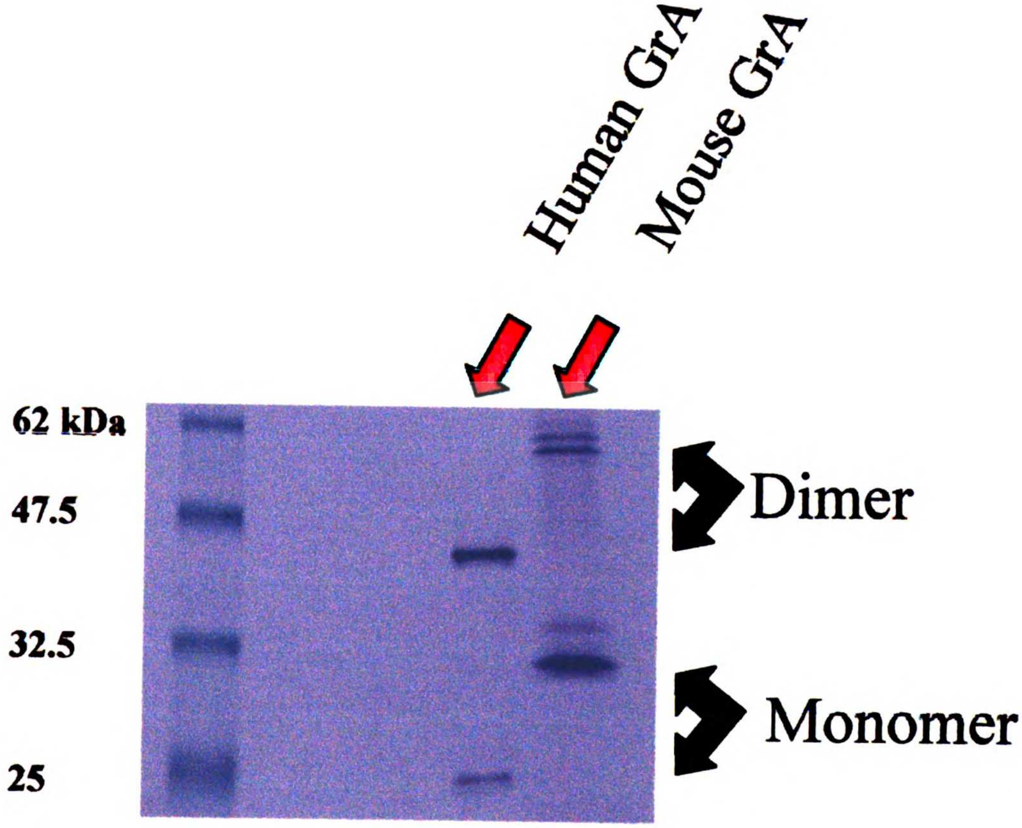
Figure 7- 2.


A.
Mouse Granzyme A
- + + DTT
- + + Granzyme A
- + + N Glycosidase F
175
83
62
47.5
32.5
25
16.5
B.
Non-Reducing
Mouse GrA
Reducing
Mouse GrA
Dimer
Monomer

UCSF LIBRARY

Figure 7-3. Coomassie stain of recombinant human and mouse granzyme A. Human granzyme A runs at 38 kDa (dimer) and 23 kDa (monomer), molecular mass based on sequence should be 25.8 kDa (monomer) and 51.6 kDa (dimer). Mouse granzyme A runs, as expected, at 56kDa (dimer) and 28 kDa (monomer).

Figure 7- 3 .


Human GrA
Mouse GrA
62 kDa
47.5
32.5
25
Dimer
Monomer

LSE LIBRARY

Figure 7-4. MALDI mass spectrum of human granzyme A, showing that the dimer and monomer have the expected molecular mass of 52 kDa and 26 kDa.

Figure 7- 4. Mass Spectrum of Human granzyme A (N161Q)

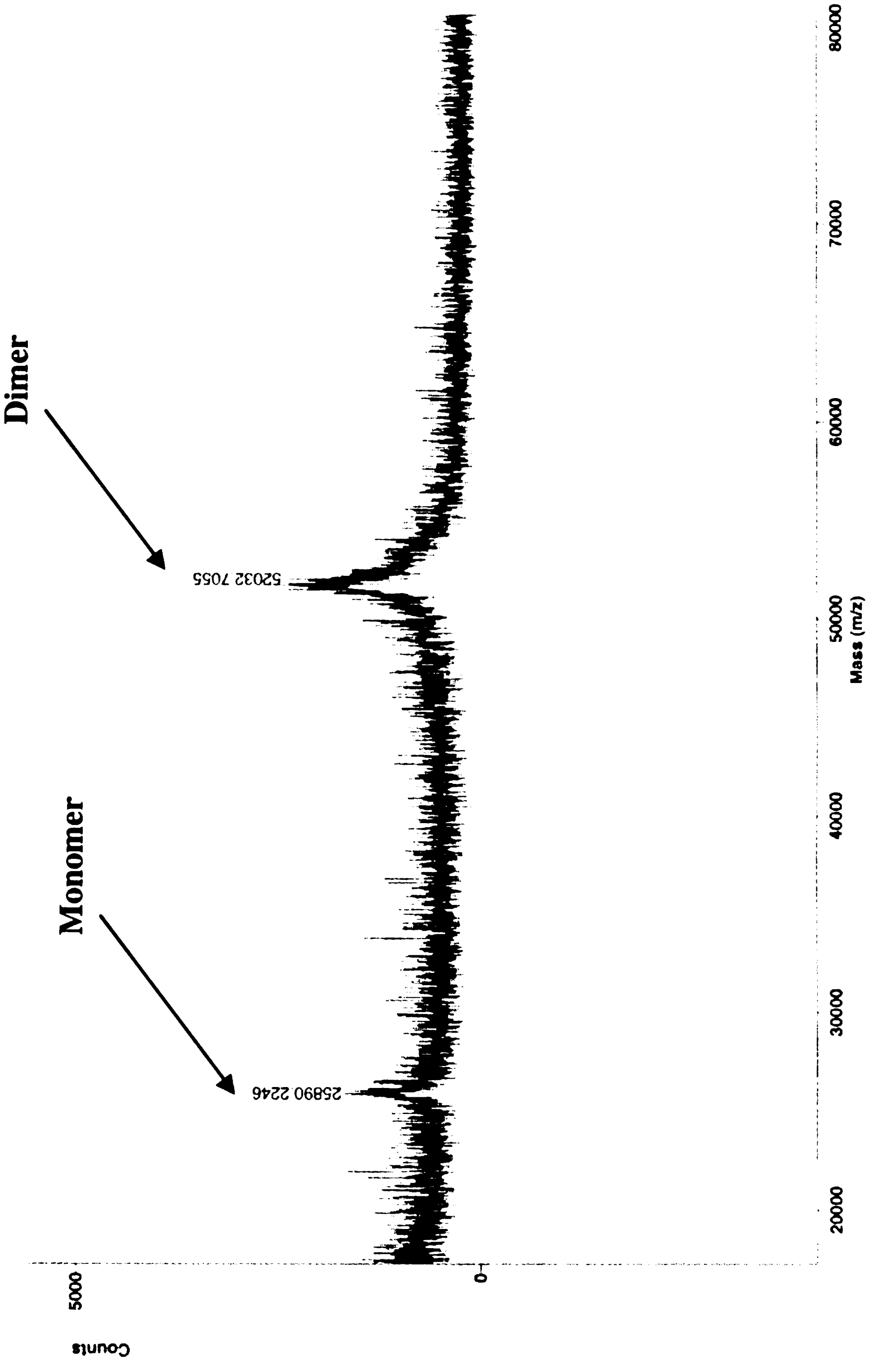
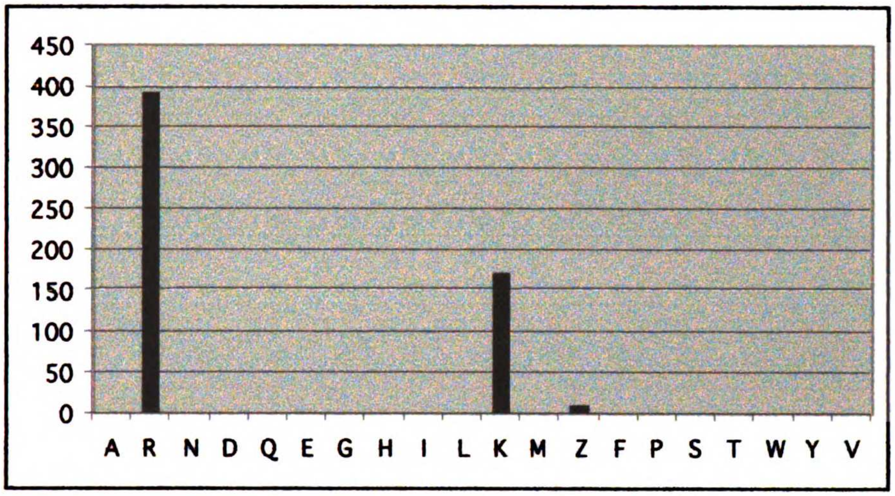
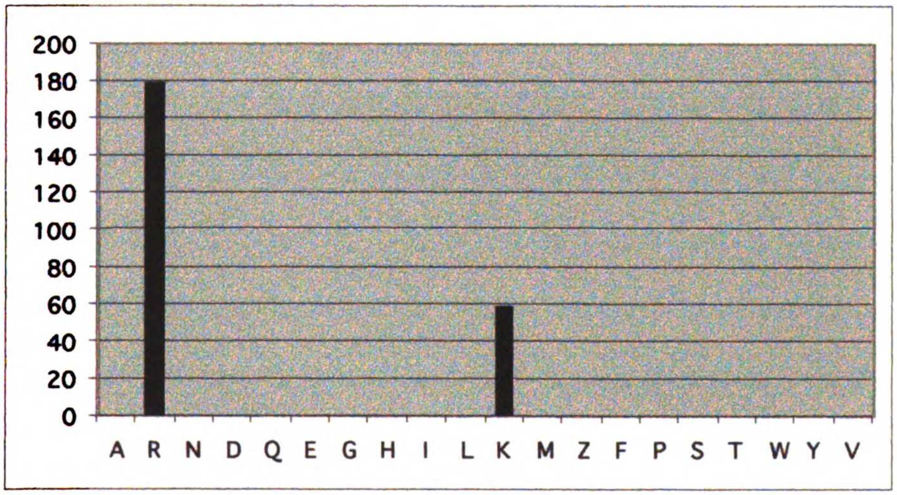


Figure 7- 5. P1-Specificity of Human and Mouse Granzyme A

P1-Diverse Library Mouse GrA



P1-Diverse Library Human GrA

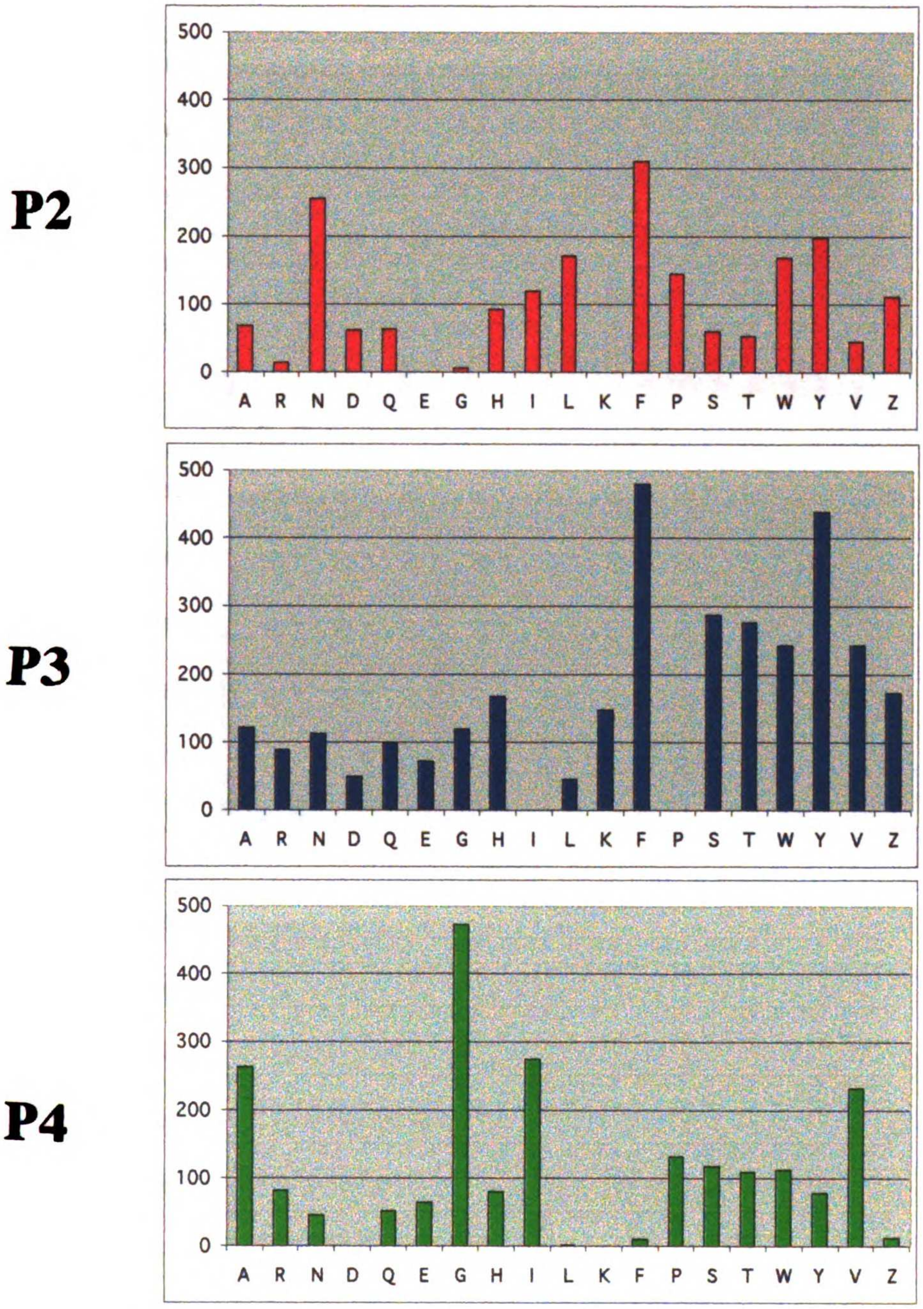


USGS LIBRARY

Figure 7-6A. Extended (P4-P2) substrate specificity of Mouse Granzyme A run against the P1-Arg PS-SCL.

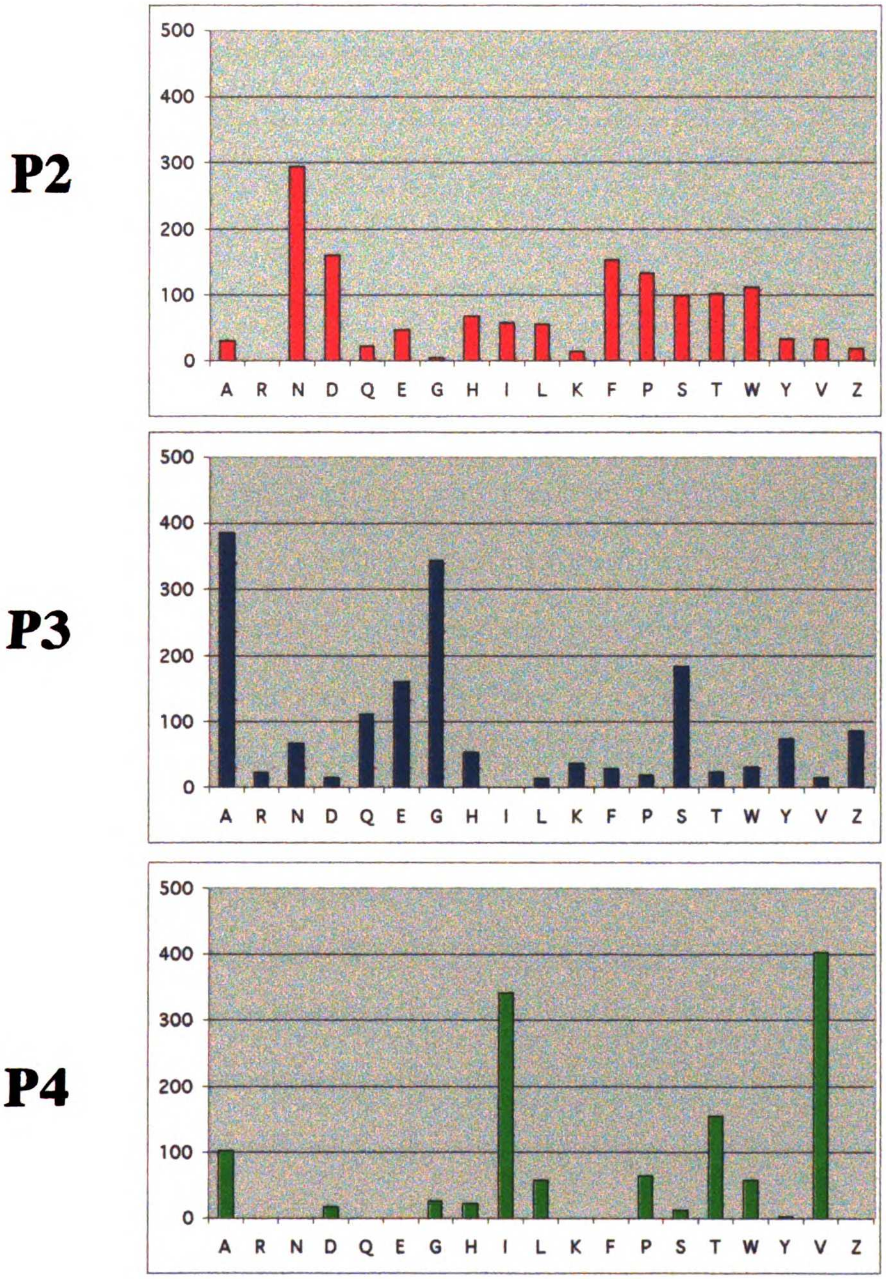
Figure 7-6B. Extended (P4-P2) substrate specificity of Human Granzyme A run against the P1-Arg PS-SCL.

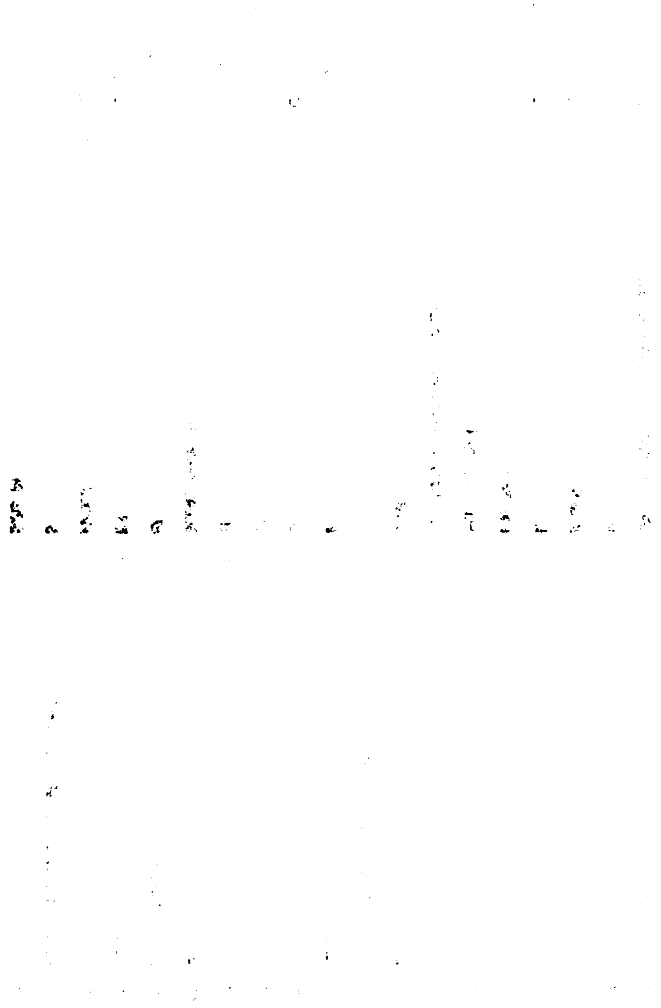
Figure 7- 6A. P4-P2 Substrate specificity of Mouse Granzyme A. (P1-Arg Library)



UCSF LIBRARY

Figure 7- 6B. P4-P2 Substrate specificity of Human Granzyme A. (P1-Arg Library)





Chapter Eight:

Future directions

The protease field is moving into exciting new areas. With the developing appreciation that proteases are involved in every biological process, the task then becomes one of characterization. The projects that I have worked on during my time at UCSF have revolved around the characterization of one major aspect of proteolytic function, that of substrate specificity. The projects have been fruitful and have provided intellectual stimulation and personal satisfaction. While scientific research is a self-correcting and dynamic process, one that adjusts its direction based on incipient experiments, I will attempt to give brief view of the future directions of my graduate projects.

Physical determinants of protease substrate specificity

My initial interest in proteolytic substrate specificity stemmed from basic questions about the physical determinants of specificity. I initially started working on dissecting the broad P1-specificity of fiddler crab collagenase through the structural analysis of the enzyme bound to various P1-residues. The resulting structure with alanine at P1 versus methionine showed that there was very little change in the S1-pocket. The exercise also taught me the fact that one's interpretations can only be as good as one's data. Crystallization of the collagenase:ecotin complex into a more tractable unit cell may allow for the collection of higher resolution data that will allow for more detailed analysis of subtle structural changes with various P1-amino acids.

The frustration with the collagenase system led me to seek out a protease system, the granzymes, that nature had adapted to have various specificities. While most of my research has led away from the physical aspects of substrate specifcity in preference of the biological aspects, the granzymes provide an excellent system for examining the structural determinants of substrate specificity. The structure of granzyme B is currently being solved by Sandra Waugh in the lab and can be immediately compared to two other enzymes of the granzyme B subclass, cathepsin G (Hof *et al.*, 1996) and rat mast cell protease II (Remington *et al.*, 1988). Additional structures of members of the granzymes, for example granzyme C, F and H, complexed to substrate-like sequences will enhance the model of required interactions for substrate turnover in this family of serine proteases. Moreover, the subtle differences between the sequences of human and mouse granzyme A (70% identity) that yield major differences in extended specificity appeal for the structural analysis of these enzymes to be pursued.

An area that I did not pursue to a large extent is the use of substrate specificity characterization to engineer "designer" proteases. One of the major benefits of using the positional scanning-synthetic combinatorial libraries to define extended substrate specificity is the ability to rapidly generate profiles for large numbers of enzymes. Coupling of the specificity information with the sequence alignments, or structural alignments where available, can lead to a general specificity model. Identification of the specificity determinants may allow for the re-engineering, or *de novo* design, of a protease to have a unique secificity profile.

Biological consequences of substrate specificity

The Holy Grail of research on regulatory proteases is the identification of downstream macromolecular substrates. I have taken two approaches to identifying downstream substrates of the granzymes, substrate specificity determination and small-pool cDNA expression screening. While I have focused on the granzymes, these approaches are completely generalizable and can be applied to other proteases. Once the substrates have been identified it is important to verify that they are cleaved by the protease in an *in vivo* (or *ex vivo*) setting. In some cases, as in the case of granzyme B for the caspases, or thrombin for PAR1, substrate cleavage sites have been identified in the substrates and specificity studies can validate their cleavage. However, for all of the substrates that I identified in the small pool cDNA expression screening, verification of their cleavage during cytotoxic lymphocyte-mediated cell death needs to be established. This could be an extremely fertile project, especially if the oncogenes, c-abl and notch, turn out to be valid targets of granzyme B. One of the major limitations of the cDNA expression screening approach is that the library is inherantly biased—the more abundant transcripts have a higher representation, incorporation of partial transcripts, inability of membrane or secreted proteins to express. The development of methods to allow for normalized representation of the proteome as well as environments to allow for the expression and screening of membrane and secreted proteins will greatly enhance the applicability of this approach.

There are many directions that this work could be pursued. While I feel that I have contributed to the identification of proteolytic activity and proteolytic substrates in an *in vitro* setting, methods for identifying proteolytic activity and proteolytic substrates in an *in vivo* setting are in need of development.

Appendix:

Publishers' permission to include Chapter Five

THE JOURNAL OF
BIOLOGICAL CHEMISTRY

Jennifer L. Harris
513 Parnassus Avenue Box 0446
Department of Pharmaceutical Chemistry
University of California San Francisco
San Francisco, CA 94143-0446

Rights and Permission
Manager
9650 ROCKVILLE PIKE
BETHESDA, MD 20814. USA

AUG 09 1999

August 4, 1999

Dear Rights and Permission Manager:

I am requesting permission to reprint in my PhD dissertation the article entitled "Definition and redesign of the extended substrate specificity of granzyme B" by Harris, JL et al. appearing in Vol 273 of the Journal of Biological Chemistry, pp. 27364-73. The article will appear in my PhD dissertation, submitted to the Graduate Division of the University of California San Francisco, and will be used for display and educational purposes only. I am willing to abide by restrictions your journal deems appropriate. Please send permission and authorization to reproduce article to the address listed above.

Sincerely,

Jennifer L. Harris

Graduate Student

PERMISSION GRANTED
contingent upon obtaining that of the author

AUG 11 1999

Jeanne Gladfelter
for the copyright owner
THE AMERICAN SOCIETY FOR BIOCHEMISTRY
& MOLECULAR BIOLOGY

References:

Agami, R., Blandino, G., Oren, M. and Shaul, Y. (1999) Interaction of c-Abl and p73alpha and their collaboration to induce apoptosis. *Nature*, **399**, 809-813.

Andrade, F., Roy, S., Nicholson, D., Thornberry, N., Rosen, A. and Casciola-Rosen, L. (1998) Granzyme B directly and efficiently cleaves several downstream caspase substrates: implications for CTL-induced apoptosis. *Immunity*, **8**, 451-460.

Arnold, F.H. and Moore, J.C. (1997) Optimizing industrial enzymes by directed evolution. *Advances in Biochemical Engineering/Biotechnology*, **58**, 1-14.

Backes, B.J. and Ellman, J.A. (1999) An alkanesulfonamide "safety-catch" linker for solid-phase synthesis. *Journal of Organic Chemistry*, **64**, 2322-2330.

Ballinger, M.D., Tom, J. and Wells, J.A. (1995) Designing subtilisin BPN' to cleave substrates containing dibasic residues. *Biochemistry*, **34**, 13312-13319.

Ballinger, M.D., Tom, J. and Wells, J.A. (1996) Furilisin: a variant of subtilisin BPN' engineered for cleaving tribasic substrates. *Biochemistry*, **35**, 13579-13585.

Berman, J., Green, M., Sugg, E., Anderegg, R., Millington, D.S., Norwood, D.L., McGeehan, J. and Wiseman, J. (1992) Rapid optimization of enzyme substrates using defined substrate mixtures. *Journal of Biological Chemistry*, **267**, 1434-1437.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer Jr, E.E., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, **112**, 535.

Betz, S.F., Liebman, P.A. and DeGrado, W.F. (1997) De novo design of native proteins: characterization of proteins intended to fold into antiparallel, rop-like, four-helix bundles. *Biochemistry*, **36**, 2450-2458.

Bevan, A., Brenner, C. and Fuller, R.S. (1998) Quantitative assessment of enzyme specificity in vivo: P2 recognition by Kex2 protease defined in a genetic system. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 10384-10389.

Birkett, A.J., Soler, D.F., Wolz, R.L., Bond, J.S., Wiseman, J., Berman, J. and Harris, R.B. (1991) Determination of enzyme specificity in a complex mixture of peptide substrates by N-terminal sequence analysis. *Analytical Biochemistry*, **196**, 137-143.

Black, M.E., Newcomb, T.G., Wilson, H.M. and Loeb, L.A. (1996) Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 3525-3529.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. and Hofsteenge, J. (1989) The refined 1.9 Å crystal structure of human α-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *The EMBO Journal*, **8**, 3467-3475.

Bode, W., Turk, D. and Karshikov, A. (1992) The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Science*, **1**, 426-471.

Bonagura, C.A., Sundaramoorthy, M., Pappa, H.S., Patterson, W.R. and Poulos, T.L. (1996) An engineered cation site in cytochrome c peroxidase alters the reactivity of the redox active tryptophan. *Biochemistry*, **35**, 6107-6115.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**, 248-254.

Braha, O., Walker, B., Cheley, S., Kasianowicz, J.J., Song, L., Gouaux, J.E. and Bayley, H. (1997) Designed protein pores as components for biosensors. *Chemistry and Biology*, **4**, 497-505.

Brake, A.J., Merryweather, J.P., Coit, D.G., Heberlein, U.A., Masiarz, F.R., Mullenbach, G.T., Urdea, M.S., Valenzuela, P. and Barr, P.J. (1984) Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **81**, 4642-4646.

Brinen, L.S., Willett, W.S., Craik, C.S. and Fletterick, R.J. (1996) X-ray structures of a designed binding site in trypsin show metal-dependent geometry. *Biochemistry*, **35**, 5999-6009.

Bunin, B.A. (1998) *The combinatorial index.* Academic Press, San Diego.

Burley, S.K. and Petsko, G.A. (1986) Amino-aromatic interactions in proteins. *Febs Letters*, **203**, 139-143.

Campbell, P.G. and Andress, D.L. (1997) Plasmin degradation of insulin-like growth factor-binding protein-5 (IGFBP-5): regulation by IGFBP-5-(201-218). *American Journal of Physiology*, **273**, E996-1004.

Cantu, C., III, Huang, W. and Palzkill, T. (1996) Selection and characterization of amino acid substitutions at residues 237-240 of TEM-1 beta-lactamase with altered substrate specificity for aztreonam and ceftazidime. *Journal of Biological Chemistry*, **271**, 22538-22545.

Caputo, A., James, M.N., Powers, J.C., Hudig, D. and Bleackley, R.C. (1994) Conversion of the substrate specificity of mouse proteinase granzyme B. *Nat Struct Biol*, **1**, 364-367.

Carpino, L.A., Ionescu, D. and Elfaham, A. (1996) Peptide Coupling in the Presence of Highly Hindered Tertiary Amines. *Journal of Organic Chemistry*, **61**, 2460-2465.

Caughey, G.H. (1994) Serine proteinases of mast cell and leukocyte granules. A league of their own. *Am J Respir Crit Care Med*, **150**, S138-142.

Chain, D., Kreizman, T., Shapira, H. and Shaltiel, S. (1991) Plasmin cleavage of vitronectin. Identification of the site and consequent attenuation in binding plasminogen activator inhibitor-1. *Febs Letters*, **285**, 251-256.

Chang, T.K., Jackson, D.Y., Burnier, J.P. and Wells, J.A. (1994) Subtiligase: a tool for semisynthesis of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 12544-12548.

Chasan, R. and Anderson, K.V. (1989) The role of easter, an apparent serine protease, in organizing the dorsal-ventral pattern of the *Drosophila* embryo. *Cell*, **56**, 391-400.

Chen, R., Greer, A. and Dean, A.M. (1996) Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 12171-12176.

Chinnaiyan, A.M., Hanna, W.L., Orth, K., Duan, H., Poirier, G.G., Froelich, C.J. and Dixit, V.M. (1996) Cytotoxic T-cell-derived granzyme B activates the apoptotic protease ICE-LAP3. *Curr Biol*, **6**, 897-899.

Christians, F.C., Dawson, B.J., Coates, M.M. and Loeb, L.A. (1997) Creation of human alkyltransferases resistant to O6-benzylguanine. *Cancer Research*, **57**, 2007-2012.

Coombs, G.S., Dang, A.T., Madison, E.L. and Corey, D.R. (1996) Distinct mechanisms contribute to stringent substrate specificity of tissue-type plasminogen activator. *Journal of Biological Chemistry*, **271**, 4461-4467.

Coussens, L.M., Raymond, W.W., Bergers, G., Laig-Webster, M., Behrendtsen, O., Werb, Z., Caughey, G.H. and Hanahan, D. (1999) Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes and Development*, **13**, 1382-1397.

Crameri, A., Dawes, G., Rodriguez, E., Jr., Silver, S. and Stemmer, W.P. (1997) Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nature Biotechnology*, **15**, 436-438.

Crameri, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology*, **14**, 315-319.

Cubitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. and Tsien, R.Y. (1995) Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences*, **20**, 448-455.

Dalal, S., Balasubramanian, S. and Regan, L. (1997) Protein alchemy: changing beta-sheet into alpha-helix. *Nature Structural Biology*, **4**, 548-552.

Daniel, R., Chung, S., Chen, H. and Wong, P.M. (1998) Retroviral transfer of antisense sequences results in reduction of C-Abl and induction of apoptosis in hemopoietic cells. *Journal of Biomedical Science*, **5**, 383-394.

Daniel, R., Wong, P.M. and Chung, S.W. (1996) Isoform-specific functions of c-abl: type I is necessary for differentiation, and type IV is inhibitory to apoptosis. *Cell Growth and Differentiation*, **7**, 1141-1148.

Darmon, A.J., Ehrman, N., Caputo, A., Fujinaga, J. and Bleackley, R.C. (1994) The cytotoxic T cell proteinase granzyme B does not activate interleukin-1 beta-converting enzyme. *J Biol Chem*, **269**, 32043-32046.

Darmon, A.J., Nicholson, D.W. and Bleackley, R.C. (1995) Activation of the apoptotic protease CPP32 by cytotoxic T-cell-derived granzyme B. *Nature*, **377**, 446-448.

Davie, E.W., Fujikawa, K. and Kisiel, W. (1991) The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry*, **30**, 10363-10370.

Del Nery, E., Juliano, M.A., Meldal, M., Svendsen, I., Scharfstein, J., Walmsley, A. and Juliano, L. (1997) Characterization of the substrate specificity of the major cysteine protease (cruzipain) from Trypanosoma cruzi using a portion-mixing combinatorial library and fluorogenic peptides. *Biochemical Journal*, **323**, 427-433.

Ding, L., Coombs, G.S., Strandberg, L., Navre, M., Corey, D.R. and Madison, E.L. (1995) Origins of the specificity of tissue-type plasminogen activator. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 7627-7631.

Dix, D.J., Allen, J.W., Collins, B.W., Mori, C., Nakamura, N., Poorman-Allen, P., Goulding, E.H. and Eddy, E.M. (1996) Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 3264-3268.

Dooley, C.T. and Houghten, R.A. (1998) Synthesis and screening of positional scanning combinatorial libraries. *Methods in Molecular Biology*, **87**, 13-24.

Duan, H., Chinnaiyan, A.M., Hudson, P.L., Wing, J.P., He, W.W. and Dixit, V.M. (1996) ICE-LAP3, a novel mammalian homologue of the Caenorhabditis elegans cell death protein Ced-3 is activated during Fas- and tumor necrosis factor-induced apoptosis. *J Biol Chem*, **271**, 1621-1625.

Dufour, E., Storer, A.C. and Maenard, R. (1995) Engineering nitrile hydratase activity into a cysteine protease by a single mutation. *Biochemistry*, **34**, 16382-16388.

Eakin, A.E., Mills, A.A., Harth, G., McKerrow, J.H. and Craik, C.S. (1992) The sequence, organization, and expression of the major cysteine protease (cruzain) from Trypanosoma cruzi. *Journal of Biological Chemistry*, **267**, 7411-7420.

Ehrlich, H.J., Grinnell, B.W., Jaskunas, S.R., Esmon, C.T., Yan, S.B. and Bang, N.U. (1990) Recombinant human protein C derivatives: altered response to calcium resulting in enhanced activation by thrombin. *Embo Journal*, **9**, 2367-2373.

el Hawrani, A.S., Sessions, R.B., Moreton, K.M. and Holbrook, J.J. (1996) Guided evolution of enzymes with new substrate specificities. *Journal of Molecular Biology*, **264**, 97-110.

Ewoldt, G.R., Smyth, M.J., Darcy, P.K., Harris, J.L., Craik, C.S., Horowitz, B., Woodard, S.L., Powers, J.C. and Hudig, D. (1997) P-4 and RNKP-7, new granzyme-like serine proteases expressed in activated rat lymphocytes. *J Immunol*, **158**, 4574-4583.

Fath, M.A., Wu, X., Hileman, R.E., Linhardt, R.J., Kashem, M.A., Nelson, R.M., Wright, C.D. and Abraham, W.M. (1998) Interaction of secretory leukocyte protease inhibitor with heparin inhibits proteases involved in asthma. *Journal of Biological Chemistry*, **273**, 13563-13569.

Fenton, J.W.d., Fasco, M.J. and Stackrow, A.B. (1977) Human thrombins. Production, evaluation, and properties of alpha-thrombin. *Journal of Biological Chemistry*, **252**, 3587-3598.

Fernandes-Alnemri, T., Armstrong, R.C., Krebs, J., Srinivasula, S.M., Wang, L., Bullrich, F., Fritz, L.C., Trapani, J.A., Tomaselli, K.J., Litwack, G. and Alnemri, E.S. (1996) In vitro activation of CPP32 and Mch3 by Mch4, a novel human apoptotic cysteine protease containing two FADD-like domains. *Proc Natl Acad Sci U S A*, **93**, 7464-7469.

Froelich, C.J., Hanna, W.L., Poirier, G.G., Duriez, P.J., D'Amours, D., Salvesen, G.S., Alnemri, E.S., Earnshaw, W.C. and Shah, G.M. (1996a) Granzyme B/perforin-mediated apoptosis of Jurkat cells results in cleavage of poly(ADP-ribose) polymerase to the 89-kDa apoptotic fragment and less abundant 64-kDa fragment. *Biochem Biophys Res Commun*, **227**, 658-665.

Froelich, C.J., Orth, K., Turbov, J., Seth, P., Gottlieb, R., Babior, B., Shah, G.M., Bleackley, R.C., Dixit, V.M. and Hanna, W. (1996b) New paradigm for lymphocyte granule-mediated cytotoxicity. Target cells bind and internalize granzyme B, but an endosomolytic agent is necessary for cytosolic delivery and subsequent apoptosis. *J Biol Chem*, **271**, 29073-29079.

Gibbs, C.S., Coutrae, S.E., Tsiang, M., Li, W.X., Jain, A.K., Dunn, K.E., Law, V.S., Mao, C.T., Matsumura, S.Y., Mejza, S.J. and et al. (1995) Conversion of thrombin into an anticoagulant by protein engineering. *Nature*, **378**, 413-416.

Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data [published erratum appears in Anal Biochem 1990 Sep;189(2):283]. *Anal Biochem*, **182**, 319-326.

Goldberg, A.L. (1995) Functions of the proteasome: the lysis at the end of the tunnel. *Science*, **268**, 522-523.

Gundersen, D., Traan-Thang, C., Sordat, B., Mourali, F. and Reuegg, C. (1997) Plasmin-induced proteolysis of tenascin-C: modulation by T lymphocyte-derived urokinase-type plasminogen activator and effect on T lymphocyte adhesion, activation, and cell clustering. *Journal of Immunology*, **158**, 1051-1060.

Halfon, S.a.C.S.C. (1996) Regulation of proteolytic activity by engineered tridentate metal binding loops. *Journal of the American Chemical Society*, **118**, 1227-1228.

Hanes, J. and Pleuckthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 4937-4942.

Harris, J.L., Backes, B.J., Leonetti, F., Craik, C.S. and Ellman, J.A. (1999) Synthesis of Positional-Scanning Libraries of Fluorogenic Peptide Substrates that Incorporate Diverse P1 Substituents: Extended Specificity Determination of Plasmin and Thrombin(Submitted).

Harris, J.L. and Craik, C.S. (1998) Engineering enzyme specificity. *Current Opinion in Chemical Biology*, **2**, 127-132.

Harris, J.L., Peterson, E.P., Hudig, D., Thornberry, N.A. and Craik, C.S. (1998) Definition and redesign of the extended substrate specificity of granzyme B. *J Biol Chem*, **273**, 27364-27373.

Hedstrom, L., Szilagyi, L. and Rutter, W.J. (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science*, **255**, 1249-1253.

Heim, R. and Tsien, R.Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology*, **6**, 178-182.

Heusel, J.W., Wesselschmidt, R.L., Shresta, S., Russell, J.H. and Ley, T.J. (1994) Cytotoxic lymphocytes require granzyme B for the rapid induction of DNA fragmentation and apoptosis in allogeneic target cells. *Cell*, **76**, 977-987.

Hof, P., Mayr, I., Huber, R., Korzus, E., Potempa, J., Travis, J., Powers, J.C. and Bode, W. (1996) The 1.8 A crystal structure of human cathepsin G in complex with Suc-Val-Pro-PheP-(OPh)2: a Janus-faced proteinase with two opposite specificities. *Embo Journal*, **15**, 5481-5491.

Huang, W., Petrosino, J., Hirsch, M., Shenkin, P.S. and Palzkill, T. (1996) Amino acid sequence determinants of beta-lactamase structure and activity. *Journal of Molecular Biology*, **258**, 688-703.

Hudig, D., Gregg, N.J., Kam, C.M. and Powers, J.C. (1987) Lymphocyte granule-mediated cytolysis requires serine protease activity. *Biochem Biophys Res Commun*, **149**, 882-888.

Hurley, J.H., Chen, R. and Dean, A.M. (1996) Determinants of cofactor specificity in isocitrate dehydrogenase: structure of an engineered NADP+ --> NAD+ specificity-reversal mutant. *Biochemistry*, **35**, 5670-5678.

Jameson, G., Roberts, DV, Adams, RW, Kyle, WS, Elmore, DT. (1973a) Determination of the operational molarity of solutions of bovine alpha-chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration. *Biochemical Journal*, **131**, 107-117.

Jameson, G.W., Roberts, D.V., Adams, R.W., Kyle,W.S.A. and Elmore, D.T. (1973b) Determination of the operational molarity of solutions of α–chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration. *Biochemical Journal*, **131**, 107-117.

Jans, D.A., Jans, P., Briggs, L.J., Sutton, V. and Trapani, J.A. (1996) Nuclear transport of granzyme B (fragmentin-2). Dependence of perforin in vivo and cytosolic factors in vitro. *J Biol Chem*, **271**, 30781-30789.

Jehn, B.M., Bielke, W., Pear, W.S. and Osborne, B.A. (1999) Protective effects of notch-1 on TCR-induced apoptosis. *Journal of Immunology*, **162**, 635-638.

Kanda, S., Tomasini-Johansson, B., Klint, P., Dixelius, J., Rubin, K. and Claesson-Welsh, L. (1999) Signaling via fibroblast growth factor receptor-1 is dependent on extracellular matrix in capillary endothelial cell differentiation. *Experimental Cell Research*, **248**, 203-213.

Kast, P., Asif-Ullah, M., Jiang, N. and Hilvert, D. (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: the importance of electrostatic catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 5043-5048.

Kawabata, S., Miura, T., Morita, T., Kato, H., Fujikawa, K., Iwanaga, S., Takada, K., Kimura, T. and Sakakibara, S. (1988) Highly sensitive peptide-4-methylcoumaryl-7-amide substrates for blood-clotting proteases and trypsin. *European Journal of Biochemistry*, **172**, 17-25.

Ke, S.H., Coombs, G.S., Tachias, K., Corey, D.R. and Madison, E.L. (1997a) Optimal subsite occupancy and design of a selective inhibitor of urokinase. *Journal of Biological Chemistry*, **272**, 20456-20462.

Ke, S.H., Coombs, G.S., Tachias, K., Navre, M., Corey, D.R. and Madison, E.L. (1997b) Distinguishing the specificities of closely related proteases. Role of P3 in substrate and inhibitor discrimination between tissue-type plasminogen activator and urokinase. *J Biol Chem*, **272**, 16603-16609.

Ke, S.H., Coombs, G.S., Tachias, K., Navre, M., Corey, D.R. and Madison, E.L. (1997c) Distinguishing the specificities of closely related proteases. Role of P3 in substrate and inhibitor discrimination between tissue-type plasminogen activator and urokinase. *Journal of Biological Chemistry*, **272**, 16603-16609.

Keller, F.G., Ortel, T.L., Quinn-Allen, M.A. and Kane, W.H. (1995) Thrombin-catalyzed activation of recombinant human factor V. *Biochemistry*, **34**, 4118-4124.

Klemba, M. and Regan, L. (1995) Characterization of metal binding by a designed protein: single ligand substitutions at a tetrahedral Cys2His2 site. *Biochemistry*, **34**, 10094-10100.

Kost, C., Benner, K., Stockmann, A., Linder, D. and Preissner, K.T. (1996) Limited plasmin proteolysis of vitronectin. Characterization of the adhesion protein as morpho-regulatory and angiostatin-binding factor. *European Journal of Biochemistry*, **236**, 682-688.

Kothakota, S., Azuma, T., Reinhard, C., Klippel, A., Tang, J., Chu, K., McGarry, T.J., Kirschner, M.W., Koths, K., Kwiatkowski, D.J. and Williams, L.T. (1997) Caspase-3-generated fragment of gelsolin: effector of morphological change in apoptosis. *Science*, **278**, 294-298.

Kuida, K., Haydar, T.F., Kuan, C.Y., Gu, Y., Taya, C., Karasuyama, H., Su, M.S., Rakic, P. and Flavell, R.A. (1998) Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell*, **94**, 325-337.

Kuliopulos, A., Covic, L., Seeley, S.K., Sheridan, P.J., Helin, J. and Costello, C.E. (1999) Plasmin desensitization of the PAR1 thrombin receptor: kinetics, sites of truncation, and implications for thrombolytic therapy. *Biochemistry*, **38**, 4572-4585.

Kunkel, T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A*, **82**, 488-492.

Kurth, T., Ullmann, D., Jakubke, H.D. and Hedstrom, L. (1997) Converting trypsin to chymotrypsin: structural determinants of S1' specificity. *Biochemistry*, **36**, 10098-10104.

Lam, K.S. and Lebl, M. (1998) Synthesis of a one-bead one-compound combinatorial peptide library. *Methods in Molecular Biology*, **87**, 1-6.

Lawson, S.L., Wakarchuk, W.W. and Withers, S.G. (1997) Positioning the acid/base catalyst in a glycosidase: studies with Bacillus circulans xylanase. *Biochemistry*, **36**, 2257-2265.

Lazebnik, Y.A., Kaufmann, S.H., Desnoyers, S., Poirier, G.G. and Earnshaw, W.C. (1994) Cleavage of poly(ADP-ribose) polymerase by a proteinase with properties like ICE. *Nature*, **371**, 346-347.

Le Bonniec, B.F., Myles, T., Johnson, T., Knight, C.G., Tapparelli, C. and Stone, S.R. (1996) Characterization of the P2' and P3' specificities of thrombin using fluorescence-quenched substrates and mapping of the subsites by mutagenesis. *Biochemistry*, **35**, 7114-7122.

Legeai-Mallet, L., Benoist-Lasselin, C., Delezoide, A.L., Munnich, A. and Bonaventure, J. (1998) Fibroblast growth factor receptor 3 mutations promote apoptosis but do not alter chondrocyte proliferation in thanatophoric dysplasia [published erratum appears in J Biol Chem 1998 Jul 24;273(30):19358]. *Journal of Biological Chemistry*, **273**, 13007-13014.

Liu, Y., Shah, K., Yang, F., Witucki, L. and Shokat, K.M. (1998) Engineering Src family protein kinases with unnatural nucleotide specificity. *Chemistry and Biology*, **5**, 91-101.

Lottenberg, R., Hall, J.A., Blinder, M., Binder, E.P. and Jackson, C.M. (1983) The action of thrombin on peptide p-nitroanilide substrates. Substrate selectivity and examination of hydrolysis under different reaction conditions. *Biochimica et Biophysica Acta*, **742**, 539-557.

Lustig, K.D., Stukenberg, P.T., McGarry, T.J., King, R.W., Cryns, V.L., Mead, P.E., Zon, L.I., Yuan, J. and Kirschner, M.W. (1997) Small pool expression screening: identification of genes involved in cell cycle control, apoptosis, and early development. *Methods in Enzymology*, **283**, 83-99.

Mace, J.E., Wilk, B.J. and Agard, D.A. (1995) Functional linkage between the active site of alpha-lytic protease and distant regions of structure: scanning alanine mutagenesis of a surface loop affects activity and substrate specificity. *Journal of Molecular Biology*, **251**, 116-134.

Mathews, II, Padmanabhan, K.P., Ganesh, V., Tulinsky, A., Ishii, M., Chen, J., Turck, C.W., Coughlin, S.R. and Fenton, J.W., 2nd. (1994) Crystallographic structures of thrombin complexed with thrombin receptor peptides: existence of expected and novel binding modes. *Biochemistry*, **33**, 3266-3279.

Matthews, D.J., Goodman, L.J., Gorman, C.M. and Wells, J.A. (1994) A survey of furin substrate specificity using substrate phage display. *Protein Sci*, **3**, 1197-1205.

Matthews, D.J. and Wells, J.A. (1993) Substrate phage: selection of protease substrates by monovalent phage display. *Science*, **260**, 1113-1117.

McGeehan, G.M., Bickett, D.M., Green, M., Kassel, D., Wiseman, J.S. and Berman, J. (1994) Characterization of the peptide substrate specificities of interstitial collagenase and 92-kDa gelatinase. Implications for substrate optimization. *Journal of Biological Chemistry*, **269**, 32814-32820.

McGrath, M.E., Erpel, T., Bystroff, C. and Fletterick, R.J. (1994) Macromolecular chelation as an improved mechanism of protease inhibition: structure of the ecotin-trypsin complex. *Embo J*, **13**, 1502-1507.

McKee, P.A., Andersen, J.C. and Switzer, M.E. (1975) Molecular structural studies of human factor VIII. *Annals of the New York Academy of Sciences*, **240**, 8-33.

Meldal, M. (1998) Intramolecular fluorescence-quenched substrate libraries. *Methods in Molecular Biology*, **87**, 65-74.

Meldal, M., Svendsen, I., Breddam, K. and Auzanneau, F.I. (1994) Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 3314-3318.

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

Mosser, D.D., Caron, A.W., Bourget, L., Denis-Larose, C. and Massie, B. (1997) Role of the human heat shock protein hsp70 in protection against stress-induced apoptosis. *Molecular and Cellular Biology*, **17**, 5317-5327.

Muzio, M., Chinnaiyan, A.M., Kischkel, F.C., O'Rourke, K., Shevchenko, A., Ni, J., Scaffidi, C., Bretz, J.D., Zhang, M., Gentz, R., Mann, M., Krammer, P.H., Peter, M.E. and Dixit, V.M. (1996) FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death--inducing signaling complex. *Cell*, **85**, 817-827.

Neamati, N., Fernandez, A., Wright, S., Kiefer, J. and McConkey, D.J. (1995) Degradation of lamin B1 precedes oligonucleosomal DNA fragmentation in apoptotic thymocytes and isolated thymocyte nuclei. *J Immunol*, **154**, 3788-3795.

Nicholson, D.W. and Thornberry, N.A. (1997) Caspases: killer proteases. *Trends Biochem Sci*, **22**, 299-306.

Niles, A.L., Maffitt, M., Haak-Frendscho, M., Wheeless, C.J. and Johnson, D.A. (1998) Recombinant human mast cell tryptase beta: stable expression in Pichia pastoris and purification of fully active enzyme. *Biotechnology and Applied Biochemistry*, **28**, 125-131.

Novak, J.F., Hayes, J.D. and Nishimoto, S.K. (1997) Plasmin-mediated proteolysis of osteocalcin. *Journal of Bone and Mineral Research*, **12**, 1035-1042.

O'Neil, K.T. and Hoess, R.H. (1995) Phage display: protein engineering by directed evolution. *Current Opinion in Structural Biology*, **5**, 443-449.

Odake, S., Kam, C.M., Narasimhan, L., Poe, M., Blake, J.T., Krahenbuhl, O., Tschopp, J. and Powers, J.C. (1991) Human and murine cytotoxic T lymphocyte serine proteases: subsite mapping with peptide thioester substrates and inhibition of enzyme activity and cytolysis by isocoumarins. *Biochemistry*, **30**, 2217-2227.

Oei, S.L., Griesenbeck, J. and Schweiger, M. (1997) The role of poly(ADP-ribosyl)ation. *Rev Physiol Biochem Pharmacol*, **131**, 127-173.

Omar, M.N. and Mann, K.G. (1987) Inactivation of factor Va by plasmin. *Journal of Biological Chemistry*, **262**, 9750-9755.

Orth, K., Chinnaiyan, A.M., Garg, M., Froelich, C.J. and Dixit, V.M. (1996) The CED-3/ICE-like protease Mch2 is activated during apoptosis and cleaves the death substrate lamin A. *J Biol Chem*, **271**, 16443-16446.

Ostresh, J.M., Winkle, J.H., Hamashin, V.T. and Houghten, R.A. (1994) Peptide libraries: determination of relative reaction rates of protected amino acids in competitive couplings. *Biopolymers*, **34**, 1681-1689.

Park, Y., Luo, J., Schultz, P.G. and Kirsch, J.F. (1997) Noncoded amino acid replacement probes of the aspartate aminotransferase mechanism. *Biochemistry*, **36**, 10517-10525.

Perona, J.J., Hedstrom, L., Rutter, W.J. and Fletterick, R.J. (1995) Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry*, **34**, 1489-1499.

Perona, J.J., Tsu, C.A., Craik, C.S. and Fletterick, R.J. (1997) Crystal structure of an ecotin-collagenase complex suggests a model for recognition and cleavage of the collagen triple helix. *Biochemistry*, **36**, 5381-5392.

Petithory, J.R., Masiarz, F.R., Kirsch, J.F., Santi, D.V. and Malcolm, B.A. (1991) A rapid method for determination of endoproteinase substrate specificity: specificity of the 3C proteinase from hepatitis A virus. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 11510-11514.

Pham, C.T.N., Thomas, D.A., Mercer, J.D., Ley, T.J. (1998) Production of fully active recombinant murine granzyme B in yeast. *Journal of Biological Chemistry*, **273**, 1629-1633.

Pinilla, C., Appel, J.R., Blanc, P. and Houghten, R.A. (1992) Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries. *Biotechniques*, **13**, 901-905.

Pinto, A.L., Hellinga, H.W. and Caradonna, J.P. (1997) Construction of a catalytically active iron superoxide dismutase by rational protein design. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 5562-5567.

Pittman, D.D., Tomkinson, K.N., Michnick, D., Selighsohn, U. and Kaufman, R.J. (1994) Posttranslational sulfation of factor V is required for efficient thrombin cleavage and activation and for full procoagulant activity. *Biochemistry*, **33**, 6952-6959.

Podack, E.R. (1995) Execution and suicide: cytotoxic lymphocytes enforce Draconian laws through separate molecular pathways. *Curr Opin Immunol*, **7**, 11-16.

Poe, M., Blake, J.T., Boulton, D.A., Gammon, M., Sigal, N.H., Wu, J.K. and Zweerink, H.J. (1991) Human cytotoxic lymphocyte granzyme B. Its purification from granules and the characterization of substrate and inhibitor specificity. *J Biol Chem*, **266**, 98-103.

Pozsgay, M., Szabao, G., Bajusz, S., Simonsson, R., Gaaspaar, R. and Eleodi, P. (1981) Study of the specificity of thrombin with tripeptidyl-p-nitroanilide substrates. *European Journal of Biochemistry*, **115**, 491-495.

Pryzdial, E.L., Lavigne, N., Dupuis, N. and Kessler, G.E. (1999) Plasmin converts factor X from coagulation zymogen to fibrinolysis cofactor. *Journal of Biological Chemistry*, **274**, 8500-8505.

Quaemaeneur, E., Moutiez, M., Charbonnier, J.B. and Maenez, A. (1998) Engineering cyclophilin into a proline-specific endopeptidase [published erratum appears in Nature 1998 Jul 16;394(6690):302]. *Nature*, **391**, 301-304.

Rano, T.A., Timkey, T., Peterson, E.P., Rotonda, J., Nicholson, D.W., Becker, J.W., Chapman, K.T. and Thornberry, N.A. (1997) A combinatorial approach for determining protease specificities: application to interleukin-1beta converting enzyme (ICE). *Chem Biol*, **4**, 149-155.

Rao, L., Perez, D. and White, E. (1996) Lamin proteolysis facilitates nuclear events during apoptosis. *J Cell Biol*, **135**, 1441-1455.

Remington, S.J., Woodbury, R.G., Reynolds, R.A., Matthews, B.W. and Neurath, H. (1988) The structure of rat mast cell protease II at 1.9-A resolution. *Biochemistry*, **27**, 8097-8105.

Robbins, K.C., Summaria, L. and Wohl, R.C. (1981) Human plasmin. *Methods in Enzymology*, **80 Pt C**, 379-387.

Sarin, A., Williams, M.S., Alexander-Miller, M.A., Berzofsky, J.A., Zacharchuk, C.M. and Henkart, P.A. (1997) Target cell lysis by CTL granule exocytosis is independent of ICE/Ced-3 family proteases. *Immunity*, **6**, 209-215.

Schagger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*, **166**, 368-379.

Schechter, I., Berger, A. (1968) On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Chemistry Communications*, **27**, 157-162.

Schellenberger, V., Turck, C.W., Hedstrom, L. and Rutter, W.J. (1993) Mapping the S' subsites of serine proteases using acyl transfer to mixtures of peptide nucleophiles. *Biochemistry*, **32**, 4349-4353.

Severin, K., Lee, D.H., Kennan, A.J. and Ghadiri, M.R. (1997) A synthetic peptide ligase. *Nature*, **389**, 706-709.

Shah, K., Liu, Y., Deirmengian, C. and Shokat, K.M. (1997) Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 3565-3570.

Shelly, L.L., Fuchs, C. and Miele, L. (1999) Notch-1 inhibits apoptosis in murine erythroleukemia cells and is necessary for differentiation induced by hybrid polar compounds. *Journal of Cellular Biochemistry*, **73**, 164-175.

Shi, L., Mai, S., Israels, S., Browne, K., Trapani, J.A. and Greenberg, A.H. (1997) Granzyme B (GraB) autonomously crosses the cell membrane and perforin initiates apoptosis and GraB nuclear localization. *Journal of Experimental Medicine*, **185**, 855-866.

Shresta, S., Graubert, T.A., Thomas, D.A., Raptis, S.Z. and Ley, T.J. (1999) Granzyme A initiates an alternative pathway for granule-mediated apoptosis. *Immunity*, **10**, 595-605.

Smith, G.P. (1994) Applied evolution. The progeny of sexual PCR [news; comment]. *Nature*, **370**, 324-325.

Smyth, M.J., O'Connor, M.D. and Trapani, J.A. (1996) Granzymes: a variety of serine protease specificities encoded by genetically distinct subfamilies. *J Leukoc Biol*, **60**, 555-562.

St. Hilaire, P.M., Willert, M., Juliano, M.A., Juliano, L. and Meldal, M. (1999) Fluorescence-quesnched solid phase combinatorial libraries in the characterization of cysteine protease substrate specificity. *Journal of Combinatorial Chemistry*, (in press).

Stemmer, W.P. (1994a) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 10747-10751.

Stemmer, W.P. (1994b) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389-391.

Still, W.C., Kahn, M. and Mitra, A. (1978) Rapid chromatographic technique for preparative separations with moderate resolution. *Journal of Organic Chemistry*, **43**, 2923-2925.

Stubbs, M.T. and Bode, W. (1993) A model for the specificity of fibrinogen cleavage by thrombin. *Seminars in Thrombosis and Hemostasis*, **19**, 344-351.

Sun, W.Y., Witte, D.P., Degen, J.L., Colbert, M.C., Burkart, M.C., Holmbeack, K., Xiao, Q., Bugge, T.H. and Degen, S.J. (1998) Prothrombin deficiency results in embryonic and neonatal lethality in mice. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 7597-7602.

Thomas, K.A., Smith, G.M., Thomas, T.B. and Feldmann, R.J. (1982) Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. *Proceedings of the National Academy of Sciences of the United States of America*, **79**, 4843-4847.

Thornberry, N.A., Rano, T.A., Peterson, E.P., Rasper, D.M., Timkey, T., Garcia-Calvo, M., Houtzager, V.M., Nordstrom, P.A., Roy, S., Vaillancourt, J.P., Chapman, K.T. and Nicholson, D.W. (1997) A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J Biol Chem*, **272**, 17907-17911.

Trapani, J.A., Jans, D.A., Jans, P.J., Smyth, M.J., Browne, K.A. and Sutton, V.R. (1998) Efficient nuclear targeting of granzyme B and the nuclear consequences of apoptosis induced by granzyme B and perforin are caspase-dependent, but cell death is caspase-independent. *J Biol Chem*, **273**, 27934-27938.

Tsiang, M., Paborsky, L.R., Li, W.X., Jain, A.K., Mao, C.T., Dunn, K.E., Lee, D.W., Matsumura, S.Y., Matteucci, M.D., Coutrae, S.E., Leung, L.L. and Gibbs, C.S. (1996)

Protein engineering thrombin for optimal specificity and potency of anticoagulant activity in vivo. *Biochemistry*, **35**, 16449-16457.

Tsirka, S.E., Bugge, T.H., Degen, J.L. and Strickland, S. (1997) Neuronal death in the central nervous system demonstrates a non-fibrin substrate for plasmin [published erratum appears in Proc Natl Acad Sci U S A 1997 Dec 23;94(26):14976]. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9779-9781.

Tsu, C.A., Perona, J.J., Schellenberger, V., Turck, C.W. and Craik, C.S. (1994) The substrate specificity of Uca pugilator collagenolytic serine protease 1 correlates with the bovine type I collagen cleavage sites. *Journal of Biological Chemistry*, **269**, 19565-19572.

Turner, D.L. and Weintraub, H. (1994) Expression of achaete-scute homolog 3 in Xenopus embryos converts ectodermal cells to a neural fate. *Genes and Development*, **8**, 1434-1447.

Van de Craen, M., Van den Brande, I., Declercq, W., Irmler, M., Beyaert, R., Tschopp, J., Fiers, W. and Vandenabeele, P. (1997) Cleavage of caspase family members by granzyme B: a comparative study in vitro. *Eur J Immunol*, **27**, 1296-1299.

van der Wal, A.C. and Becker, A.E. (1999) Atherosclerotic plaque rupture--pathologic basis of plaque stability and instability. *Cardiovascular Research*, **41**, 334-344.

Vindigni, A., Dang, Q.D. and Di Cera, E. (1997) Site-specific dissection of substrate recognition by thrombin. *Nature Biotechnology*, **15**, 891-895.

Vu, T.K., Hung, D.T., Wheaton, V.I. and Coughlin, S.R. (1991) Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell*, **64**, 1057-1068.

Wade, H. and Scanlan, T.S. (1997) The structural and functional basis of antibody catalysis. *Annual Review of Biophysics and Biomolecular Structure*, **26**, 461-493.

Wang, C.I., Yang, Q. and Craik, C.S. (1995) Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin. *J Biol Chem*, **270**, 12250-12256.

Wang, X., Lin, X., Loy, J.A., Tang, J. and Zhang, X.C. (1998) Crystal structure of the catalytic domain of human plasmin complexed with streptokinase. *Science*, **281**, 1662-1665.

Welle, M. (1997) Development, significance, and heterogeneity of mast cells with particular regard to the mast cell-specific proteases chymase and tryptase. *Journal of Leukocyte Biology*, **61**, 233-245.

Wentworth, P., Datta, A., Blakey, D., Boyle, T., Partridge, L.J. and Blackburn, G.M. (1996) Toward antibody-directed "abzyme" prodrug therapy, ADAPT: carbamate prodrug activation by a catalytic antibody and its in vitro application to human tumor cell killing. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 799-803.

Wenzel, T.J., Migliazza, A., Steensma, H.Y. and van den Berg, J.A. (1992) Efficient selection of phleomycin-resistant Saccharomyces cerevisiae transformants. *Yeast*, **8**, 667-668.

Werb, Z. (1997) ECM and cell surface proteolysis: regulating cellular ecology. *Cell*, **91**, 439-442.

Willett, W.S., Brinen, L.S., Fletterick, R.J. and Craik, C.S. (1996) Delocalizing trypsin specificity with metal activation. *Biochemistry*, **35**, 5992-5998.

Williams, M.S. and Henkart, P.A. (1994) Apoptotic cell death induced by intracellular proteolysis. *J Immunol*, **153**, 4247-4255.

Yang, S.Q., Wang, C.I., Gillmor, S.A., Fletterick, R.J. and Craik, C.S. (1998) Ecotin: a serine protease inhibitor with two distinct and interacting binding sites. *Journal of Molecular Biology*, **279**, 945-957.

You, L. and Arnold, F.H. (1996) Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide [published erratum appears in Protein Eng 1996 Aug;9(8):719]. *Protein Engineering*, **9**, 77-83.

Zerner, B., Richard P. M. Bond and Myron L. Bender. (1964) Kinetic evidence for the formation of the acyl-enzyme intermediates in the α-chymotrypsin-catalyzed hydrolyses of specific substrates. *Journal of the American Chemical Society*, **86**, 3674-3679.

Zhang, J.H., Dawes, G. and Stemmer, W.P. (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 4504-4509.

Zhao, H. and Arnold, F.H. (1997a) Combinatorial protein design: strategies for screening protein libraries. *Current Opinion in Structural Biology*, **7**, 480-485.

Zhao, H. and Arnold, F.H. (1997b) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Research*, **25**, 1307-1308.

Zhao, H., Li, Y. and Arnold, F.H. (1996) Strategy for the directed evolution of a peptide ligase. *Annals of the New York Academy of Sciences*, **799**, 1-5.

Zimmerman, M., Ashe, B., Yurewicz, E. and Patel, G. (1977) Sensitive assay for trypsin, elastase, and chymotrypsin using fluorogenic substrates. *Analytical Biochemistry*, **78**, 47-51.

Zunino, S.J., Bleackley, R.C., Martinez, J. and Hudig, D. (1990) RNKP-1, a novel natural killer-associated serine protease gene cloned from RNK-16 cytotoxic lymphocytes. *J Immunol*, **144**, 2001-2009.

UCSF LIBRARY

For reference
Not to be taken from the room.