

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Analyzing structure function relationships of serine proteases and their macromolecular inhibitors by combinatorial biochemistry

Permalink

<https://escholarship.org/uc/item/9x57x96n>

Author

Yang, Steve Qing

Publication Date

1997

Peer reviewed|Thesis/dissertation

Analyzing Structure Function Relationships
of Serine Proteases and Their Macromolecular Inhibitors
by Combinatorial Biochemistry

by

Steve Qing Yang

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

.
.
.
.
.

Copyright (1997)
by
Steve Qing Yang

For my parents, Mingzhong Yang and Hongwei Chen

For my brother, Kun Yang

For my wife

Iris Hong Chu

Preface

读万卷书行万里路

The education of a scholar is not complete until he has absorbed
ten thousand volumes and has traveled ten thousand miles.

----- Ancient Chinese Teaching

Since I was young, this ancient teaching has constantly reminded me that learning will be a life-long journey. Now at the end of graduate school, I have at least fulfilled the travel requirement. From my home town Wuhan, Hubei Province of China, to Shanghai, then to the United States (Michigan and California), with the aid of modern aviation technology, I have traveled more than 10,000 miles. Of course, the reading requirement will take a little bit longer.

I feel fortunate to live in a revolutionary era when molecular biology, together with other breakthroughs in the fields of biomedical science, has started to uncover fundamental rules of life. Being in UCSF and experiencing the thrill of discovery were exciting. It is even more fortunate for me to work under the guidance of my thesis advisor, Dr. Charles Craik. He has created a challenging environment to work on cool projects. Over the years Charly has spent tremendous amount of time and efforts to help me with my thinking and writing. I benefited from his input and suggestions. I thank him for his insights and high spirit. I also thank him for giving me the freedom and support to explore my own ideas. I know that he cares about my project and my career. I was really moved to hear Charly's cheers in my graduation ceremony.

I want to thank the other members of my thesis committee, Drs. Robert Fletterick and Jim Wells. Like Charly, they are my continuing sources of inspiration. Robert provides encouragement and guidance for me to work on extremely challenging problems such as the cold-adaptation of proteases and the relationship between conformational flexibility and activities of enzymes, which has been a wonderful learning experience. I want to thank Robert for his advice, support, and understanding of my career decisions. Dr. Jim Wells' research has a significant impact on my project. Over the years, Jim has offered me timely

advises and suggestions. I also appreciate the interactions with Lei Jin and Shane Atwell of the Wells lab.

The Craik Lab years has been truly memorable. I find that the intensity of this experience can be indirectly measured by how the former lab graduates and myself express our feelings in the preface of our Ph.D. thesis. According to an unpublished study (Yang, 1997), the average length of the doctoral dissertation preface by Craik Lab members is significantly longer than the UCSF average. These data are listed in Table 0.

Table 0 Length of Ph.D. Dissertation Preface Page

Lab Member by "label name"	by real name	Year of Graduation	Preface Length (page)
JRV	John Vasquez	1990	2.5
LE	Luke Evnin	1990	1.3
AE	Ann Eakin	1992	4.0
JRR	Jason Rosé	1995	2.1
CT	Christopher Tsu	1995	2.5
WSW	Scott Willett	1995	2.8
QY	Steve Qing Yang	1997	3.1
CSC	Craik Lab average	1990-1997	2.6 ± 0.9
Wild-Type	UCSF average	1990-1996	1.2 ± 0.5

Note: The average preface page length of the typical doctoral dissertation in UCSF between 1990-1996 was determined through random sampling of 40 dissertations at this period in the fields of pharmaceutical chemistry, biochemistry, biophysics, genetics, physiology, immunology, and neuroscience.

The significant difference between these two groups of data revealed the tremendous commitment and enthusiasm that all of the former Craik lab graduate students have devoted to their research. Having shared lab and bench spaces with most of them except for John and Luke, I had the opportunities to know each of them in person. They are talented, dynamic and friendly. I'd like to thank all of them for their help and friendship.

I came to Craik Lab before the early " generation" of postdocs left. During the last few years a new wave of postdocs have come in. I am probably the only one in the lab (except for Charly) that overlapped with these sixteen scientists at one time or the other. It was a rich experience to interact with these extraordinary colleagues and learn from their diverse

backgrounds. I still remembered my first postdoc mentor, Dr. Dave Corey, showed me how to run a western blot in my first week in the lab. Dave was an excellent teacher. I also want to thank Dr. Iain Murray for sharing my aspirations and frustrations on protein design and engineering. His constant encouragement has been crucial for my growth. I thank Dr. Cheng-I Wang for his many advises and help on experiments. I thank Dr. Martha Laboissière for her support and understanding to help me finish my projects. Martha has spent a lot of time reading and editing my manuscripts and thesis. I really appreciate her help. I wish the junior graduate students, Jeniffer Harris and Deborah Dauber, good luck with their projects. Finally, I thank Charly for attracting such a wonderful group of people.

I want to thank Sam Jackson for his friendship and encouragement. Sam has been a great friend, with whom I shared my aspirations. Sarah Gillmor of Fletterick Lab and Shane Atwell of Wells Lab have been my long time colleagues and close friends. I appreciate the exchange of ideas with them. I wish I had more discussions and sharing of experience with them.

It has been a long and winding path to climb to the mountain top of my graduate career. My heart is filled with the joy of discovery because I am enjoying a panoramic view that I always dreamed about. This may be the occasion that I can share Caesar's spirit of *Veni, vidi, vici*. However, I feel the opposite, "I came, I saw, I was humbled". In front of me is the rugged landscape of scientific knowledge where small islands and mountains of known facts were surrounded by the vast ocean of ignorance and mystery. Looking back, I see the deep valleys and dark caves that I have passed through and many crossroads and sidetracks that lead to the unknown territories. I am proud of where I stand as well as how far I have gone. The journey itself is what makes these years so rewarding. As I am leaving the campus, I see a much broader world that is waiting to be explored and many more new horizons and perspectives demanding the highest level of commitment and rigor from the curious mind. After all, the ancient Chinese teaching is true: life is learning.

It was my parents who has instilled the value of learning into my mind. My father, Mingzhong Yang, and my Mother, Hongwei Chen, have been my role models. They taught me persistence and hardworking. They paved the road for my education and growth. My gratitude for them is beyond writing.

Last, I thank my dear friend and wife, Hong, for ALWAYS being with me. Your patience and support are what make this thesis possible. Your love and hope are what make this thesis meaningful.

Table of Contents

Chapter One

An Introduction to Protein Engineering and Phage Display	
Introduction	2
Background	4
Filamentous Phage Display System	13

Chapter Two

Display Trypsin on Phage	
Introduction	21
Results	24
Design of the Trypsin Phage Display System	24
Preparation of Fusion Phage and Choice of a Phage Display System	27
Co-expression of an endogenous E. coli protease inhibitor, Ecotin	31
Characterization of Trypsin activity	34
Discussion	36
Materials and Methods	38
Materials	38
Construction of the Trypsin Phage Display Vector	38
Trypsin Phage Growth and Preparation	39
Immunoblotting Analysis of Trypsin Phage	40
Kinetic Characterization of Trypsin Phage	40

Chapter Three

Engineering Trypsin Variant Phage and Construction of a Trypsin Binding Pocket Library	
Introduction	43
Results	44
Trypsin Phage Variants D189S and D102N	44
Design of a Trypsin Phage Library Randomized at Three Positions in the Substrate Binding Pocket of Trypsin	46
Construction and Characterization of the Trypsin Phage Library	50
Ecotin Deleted E. coli Strains	54
Discussion	58
Materials and Methods	60
Materials	60
Activity Assay of Trypsin in Periplasmic Extract	60
Trypsin D189S and D102N Phage Activity Assay	61
Trypsin Phage Library Vector Construction	61
Trypsin Phage Library Preparation	63

Chapter Four

In Vitro Selection of Trypsin Phage	
Introduction	65
Results	67
Biopanning of Trypsin Phage	67
Biotin-Leupeptin Semicarbazone as a Ligand for in vitro Selection of Trypsin Phage	68

Design and Synthesis of an Indoxyl Chromogenic Protease Substrate	74
Biotinylated Inhibitors on Oxirane Beads for in vitro Selection of Active Trypsin Phage	79
Enrichment of Trypsin Phage over Background by Ecotin.....	83
Selective Enrichment of Active Trypsin Phage by Soybean Trypsin Inhibitor (STI) and Bovine Pancreatic Trypsin Inhibitor (BPTI).....	88
Discussion.....	92
Materials and Methods.....	95
Materials.....	95
Polystyrene Plate Panning and Blue / White Selection of Trypsin Phage	96
Immobilization of Protein Ligands on Oxirane Beads	96
Panning of Trypsin Phage with Oxirane Beads.....	97
PCR Amplification of Trypsin Phage.....	98
 Chapter Five	
Ecotin : A Serine Protease Inhibitor with Two Distinct and Interacting Binding Sites	
Introduction	101
Results.....	106
Ecotin Variants Preparation and Characterization.....	106
The Role of the Primary Binding Site.....	109
The Role of the C-terminal Dimer Interface	109
Region-Specific Mutagenesis at the Primary and Secondary Binding Sites.....	115
The Role of the 60s Loop and 100s Loop.....	120
Nonadditivity among Ecotin Variants.....	123
Discussion.....	127
Ecotin's Secondary Site as a Key Element in Macromolecular Recognition	127
Monomer Dimer Equilibrium	129
Protease Inhibitors with Multiple Binding Sites	129
Materials and Methods.....	133
Materials.....	133
Computer Modeling of Ecotin	134
Ecotin Mutagenesis.....	135
Expression and Purification of Recombinant Ecotin	136
Determination of Ecotin Ki's against Various Proteases.....	137
 Chapter Six	
Engineering Bidentate Macromolecular Inhibitors for Trypsin and Urokinase	
Introduction	141
Results.....	144
Identifying Key Residues at the 60s loop of the Secondary Binding Site	144
Designing and Constructing Ecotin Phage Display Libraries at the 60s loop	146
Panning of the ecotin 60X4 Library with Bovine and Rat Trypsin.....	149
Inhibition Kinetics of Ecotin Variants Y69F+D70P and Y69L+D70P.....	153
Liquid vs. Plate Amplification of Library M84R+60X4 with uPA.....	154
Panning of the Ecotin M84R+60X4 library with Rat Trypsin and uPA	157
Inhibition Kinetics of Ecotin M84R+D70R and Ecotin M84R+D70P	159
Discussion.....	161
Materials and Methods.....	168

Materials	168
Plasmid and Library Constructions	168
Ecotin Phage Preparation	170
Ecotin Library Panning	170
Ecotin Library Amplification	171
 Chapter Seven	
Future Directions	
Introduction	173
Discussion	174
Trypsin Phage Display	174
Overcome the Deletion Problem	174
Design Larger and More Diverse Trypsin Libraries	175
Display Other Trypsin-like Proteases on Phage	176
Ecotin Phage Display	176
The 100s Loop Libraries at the Secondary Site	177
The 50s and 80s Loop Libraries at the Primary Site	177
The Loop Combination Libraries	177
Ecotin Loop Trimming and Remodeling	178
Ecotin Minimization	178
Obligate Ecotin Monomer	179
Kinetic Characterizations of Monomer-Dimer Equilibria	179
Structural Characterization of Ecotin Variants	180
 References	 181
 Appendix One	
Restriction Map of Trypsin Phage Display Vector XbaI / Hind III	
Fragment of pMTsn	200
 Appendix Two	
Restriction Map of Trypsin Phage Display Vector XbaI / Hind III	
Fragment of pMTsn	206
 Appendix Three	
Restriction Map of Ecotin Phage Display Vector pBS eco-gIII	211

List of Figures

Figure 1.1 Recombinant Methods for Protein Modification.....	7
Figure 1.2 Diagram of Phage Display System	16
Figure 2.1 Trypsin Phage Display Vector	26
Figure 2.2 Trypsin Phage Tsn::III or Tsn::VIII Derived From a Phagemid/Helper Phage System	29
Figure 2.3 Trypsin Phage MTsn::III and MTsn::VIII.....	30
Figure 2.4 Trypsin Phage - Ecotin Complex.....	32
Figure 2.5 Immunoblot Analysis of Ecotin-Trypsin Phage Complex.....	33
Figure 3.1 Construction of Trypsin D189S Phage	45
Figure 3.2 The Primary Substrate Binding Pocket of Trypsin.....	49
Figure 3.3 Construction of Trypsin Phage Library Vector	51
Figure 3.4 Construction of Trypsin Phage Library.....	52
Figure 3.5 Expression of Trypsin in E. coli Strains X90 and IMΔecoX	55
Figure 4.1 Synthesis of the Biotin-Leupeptin Semicarbazone Conjugate.....	71
Figure 4.2 Inhibition of Trypsin Phage Activity by Biotin-Leupeptin- Semicarbazone	72
Figure 4.3 Design of a Chromogenic Trypsin Phage Substrate.....	77
Figure 4.4 Synthesis of the Chromogenic Substrate Suc-Leu-indoxyl	78
Figure 4.5 Enrichment Ratio of Trypsin WT Phage over M13 Phage	86
Figure 4.6 Enrichment Ratio of Trypsin D102N Phage over M13 Phage	86
Figure 5.1 Tetrameric Complex of Ecotin-Trypsin.....	104
Figure 5.2 SDS-PAGE Gel of Ecotin Primary Site and Dimer Interface Variants.....	108
Figure 5.3 SDS-PAGE Gel of Ecotin Secondary Site Variants.....	108
Figure 5.4 K _i Plot of Ecotin Truncation Variants.....	113
Figure 5.5 Proteolytic Digestion of Ecotin Multiple Alanine Substitution Variants.....	119
Figure 5.6 K _i Data of Ecotin 60s Loop Variants.....	121
Figure 5.7 K _i Data of Ecotin 100s Loop Variants	121
Figure 5.8 K _i Data of Ecotin 60s and 100s Loops Variants.....	121
Figure 5.9 Nonadditivity of Ecotin Mutations	126
Figure 6.1 Comparison of Liquid and Plate Amplification	156
Figure 6.2 Modeled Electrostatic Interaction at the 60s Loop of Ecotin.....	163

List of Tables

Table 1.1	Methods for Analyzing Macromolecule Repertoires.....	9
Table 1.2	Four Classes of Phage Display Vectors.....	17
Table 2.1	Kinetic Constants of MTsn.....	34
Table 2.2	Characterization of Trypsin Phage.....	35
Table 2.3	Trypsin Phage Display Vectors.....	36
Table 3.1	Kinetic Characterization of Trypsin Variant on Phage	46
Table 3.2	Individual Clones in the Trypsin Binding Pocket Phage Library	53
Table 3.3	Sample Nucleotide Distribution of Trypsin Phage Library	53
Table 3.4	Characterization of Ecotin Deletion Strains IM Δ ecoJ and IM Δ ecoX	56
Table 4.1	Ligands and Matrixes Tested in Panning of Trypsin Phage	68
Table 4.2	Inhibition of Trypsin Phage Activity by Biotin-Leupeptin-Semicarbazone	73
Table 4.3	Leupeptin and Chloromethylketone as Ligands for Trypsin Phage Panning	74
Table 4.4	Development of Biotinylated Inhibitors as Ligands for Trypsin Phage Panning	81
Table 4.5	Enrichment Ratio of Trypsin WT Phage over Trypsin D189S or D102N Phage against Biotin-RQR-CH ₂ Cl on Oxirane Beads.....	83
Table 4.6	Enrichment Ratio of Trypsin WT and D189S Phage against Ecotin Variants on Oxirane Beads	87
Table 4.7	Capturing Trypsin Activity by BPTI and STI Oxirane Beads.....	89
Table 4.8	Enrichment Ratio of Trypsin WT, D102N and D189S Phage against STI and BPTI	90
Table 4.9	Enrichment Ratio of Trypsin WT Phage over Trypsin D189S Phage against STI and BPTI	91
Table 5.1	Definition of Ecotin Variants.....	107
Table 5.2	K _i of Ecotin P1 Variants at Position 84 against Rat Trypsin	109
Table 5.3	K _d of Ecotin C-terminal Truncation Variants	111
Table 5.4	K _i of Ecotin C-terminal Truncation Variants	112
Table 5.5	Identification of Ecotin Secondary Site Residues.....	116
Table 5.6	K _i of Ecotin Multiple Alanine Substitution Variants.....	117
Table 5.7	K _i of Ecotin Combinatorial Variants	120
Table 5.8	Relative Change in Interaction Energy among Ecotin's Primary, Secondary Sites and Dimer Interface.....	125
Table 6.1	K _i of Ecotin Combinatorial Variants	146
Table 6.2	The Consensus Sequence from 60X4 Library Panning against Rat Trypsin.....	152
Table 6.3	K _i of Ecotin Variants from Library 60X4	153
Table 6.4	The Consensus Sequence from M84R+60X4 Library Panning	

against uPA.....	158
Table 6.5 Designing Specificity & Potency of Ecotin.....	160
Table 6.6 Ecotin Variants $\Delta\Delta G$	165
Table 6.7 Summary of Ecotin Phage Library Panning Results	165

List of Abbreviations

AMC, 7-amino-4-methylcoumarin
APPI, amyloid β -protein precursor inhibitor protein
boc, t-butyloxycarbonyl
BPTI, bovine pancreatic trypsin inhibitor
BSA, bovine serum albumin
IPTG, isopropylthio- β -D-galactoside
LMuPA, low molecular weight uPA
MUGB, 4-Methylumbelliferyl *p*-guanidinobenzoate
PAGE, polyacrylamide gel electrophoresis
PBST, 137 mM NaCl/2.7 mM KCl/10 mM Na₂HPO₄/1.8 mM K₂HPO₄ pH 7.5, 0.5% Tween-20 buffer
PCR, polymerase chain reaction
pNa, *p*-nitroanilide
R.M.S., root mean square
SDS, sodium dodecyl sulphate
STI, soybean trypsin inhibitor
TE, 10 mM Tris, 1mM EDTA buffer
Tn100, 50 mM NaCl/50 mM Tris/10 mM CaCl₂, pH 8.0 buffer
uPA, urokinase-type plasminogen activator (uPA)
X-Gal, 5-bromo-4-chloro-3-indoxyl- β -D-galactoside
Z, carbobenzoxy
Z-GPR-AMC, N α -benzyloxy-carbonyl-L-glyylproylarginine-7-amino-4-methylcoumarin
Z-GPR-pNa, N α -benzyloxy-carbonyl-L-glyylproylarginine-*p*-nitroanilide

Analyzing Structure Function Relationships of Serine Proteases and Their Macromolecular Inhibitors by Combinatorial Biochemistry

Steve Qing Yang

Bacteriophage display that genetically fuses the macromolecules of interest to the surface of bacteriophage is an innovative tool to investigate the combinatorial interactions among the functionally important residues of serine proteases and their macromolecular inhibitors. Rat anionic trypsin was displayed on the surface of filamentous phage and achieved normal catalytic activity. A complete trypsin binding pocket library, in which residues 216, 225, and 226 were randomized, was constructed and characterized. Macromolecular and low molecular weight protease inhibitors were tested to establish an effective *in vitro* selection system for trypsin phage. Soybean trypsin inhibitor coupled to oxirane beads selectively enriched active trypsin phage. Ecotin, a dimeric serine protease inhibitor from *Escherichia coli*, is a novel platform for inhibitor design. Ecotin binds to proteases through the distinct combinations of primary and secondary site surface loops. The interactions between ecotin and two target proteases have been systematically dissected to understand the structural basis of ecotin's broad inhibitory specificity and the role of the secondary binding site. Potent bidentate ecotin inhibitors were created for trypsin and human urokinase-type plasminogen activator (uPA). A limited set of key residues at ecotin's secondary binding site play essential roles in binding to the target protease. The ecotin 60s loop which binds to the target protease at a region distal from the enzyme active site was redesigned. Two ecotin phage libraries randomized at the 60s loop were constructed to exploit the binding interactions at the secondary site. Panning these libraries with trypsin and uPA generated two distinct consensus sequences which encoded ecotin variants with improved potency towards the respective target proteases. One ecotin variant identified was a 50 pM inhibitor of uPA. Furthermore, this ecotin variant achieved 13,680-fold improvement of specificity towards uPA vs. rat trypsin. The step-wise optimization of ecotin for protease inhibition was effective in combining two powerful approaches, structure-based mutagenesis and *in vitro* selection. The combinatorial approach of phage display has been shown to have great potential for designing of potent and specific proteases and inhibitors.



Chapter One : An Introduction to Protein Engineering and Phage Display

ABSTRACT

Protein engineering involves the creation of novel proteins through either *de novo* design of new proteins or redesigning preexisting ones for altered functions. The rational modification of a protein requires the identification of the critical determinants of protein activity and the detailed study of their structural and functional contributions. In the absence of sufficient preestablished information regarding structure-activity relationships in a protein, one must confront the challenge of locating key residues that are important for altering function. This challenge has stimulated innovative development of screens or selections to effectively sort through the large macromolecular repertoires. Bacteriophage display that genetically fuses the macromolecules of interest to the surface of bacteriophage is the most widely used procedure to date for screening repertoires of protein macromolecules. The phage display technique is an innovative and powerful tool for engineering enzymes and their macromolecular inhibitors. It offers new opportunities to investigate the combinatorial interactions among the functionally important residues of a target macromolecule. This work has been published (Yang *et al.*, 1995).

Introduction

Protein engineering involves the creation of novel proteins. One approach for making new proteins is redesigning preexisting ones for altered functions. Knowledge of structure-function relationships of the protein of interest serves as a powerful starting point for engineering that protein for a particular purpose. Although any amino acid can play a significant role in maintaining the final stable structure, only a relatively small set of amino acid residues is directly responsible for function. The identification of these critical determinants of activity and the detailed study of their structural and functional contributions provide an essential database for the rational modification of a protein.

Biophysical analysis can provide critical information regarding potential roles of particular side chains in a given protein. In those cases where a significant amount of information already exists for the target protein, a directed mutagenesis approach involving the substitution of a limited number of functional residues can be applied to alter protein activity. In the absence of sufficient preestablished information regarding structure-activity relationships in a protein, one must confront the challenge of locating key residues that are important for altering function. One approach is to adapt a random mutagenesis strategy to generate large and diverse molecular repertoires. These repertoires can be searched for individual proteins with desired properties. Selected variants can then be analyzed to elucidate the structural basis for the particular function. An attractive method to aid screening of the diverse macromolecular repertoires for specific functional properties is bacteriophage display.

Phage display is a method that genetically fuses the macromolecules of interest to the surface of bacteriophage and is the most widely used procedure to date for

screening repertoires of protein macromolecules. Phage display has been used to analyze peptides (Cwirla *et al.*, 1990; Scott & Smith, 1990), antibodies (Barbas *et al.*, 1991; Clackson *et al.*, 1991; McCafferty *et al.*, 1990), growth hormone variants (Bass *et al.*, 1990; Lowman *et al.*, 1991), DNA binding proteins (Jamieson *et al.*, 1994; Rebar & Pabo, 1994), enzymes (Corey *et al.*, 1993; McCafferty *et al.*, 1991; Soumillion *et al.*, 1994), protease inhibitors (Pannekoek *et al.*, 1993; Roberts *et al.*, 1992a; Roberts *et al.*, 1992b; Wang *et al.*, 1995), cytokines (Buchli *et al.*, 1997; Cabibbo *et al.*, 1995; Saggio *et al.*, 1995), and protein structural domains such as the IgG binding domain of peptostreptococcal protein L (Gu *et al.*, 1995), the four-helix bundle protein cytochrome b562 (Ku & Schultz, 1995), an alpha-helical domain derived from a bacterial receptor (Nord *et al.*, 1995), the B1 immunoglobulin-binding domain from streptococcal protein G (O'Neil *et al.*, 1995), the "minibody" (Martin *et al.*, 1996), and the "peptabody" (Terskikh *et al.*, 1997). Phage display systems have been extensively reviewed (Clackson & Wells, 1994; Dunn, 1996; Gallop *et al.*, 1994; Gordon *et al.*, 1994; Hoogenboom, 1997; McGregor, 1996; Smith & Scott, 1993; Winter *et al.*, 1994).

Background

An ultimate goal of protein engineering efforts is the *de novo* design of proteins with prescribed functions. However, until the fundamental rules for protein design become established, the *redesign* of preexisting proteins is commonly used to obtain proteins with altered functions. The field of protein engineering has witnessed continuous progress regarding protein modification and analysis techniques, which has provided greater insight on the structure-function relationship of these macromolecules. The generation of diverse libraries of protein variants and the isolation of interesting and potentially useful proteins from these libraries are two major themes for the development of new methodologies.

Protein chemistry provides a host of reagents that can be used to chemically modify one specific residue or a class of residues depending on the reactivities of the functional groups (for background, see (Lundblad, 1991; Lundblad, 1995)). However, certain reagents such as affinity labels (Shaw *et al.*, 1965) are frequently too specific for general use while less specific reagents can only be applied to highly reactive side chains and solvent accessible sites. Furthermore, the extent of chemical modification is seldom complete leaving a heterogeneous population of modified protein. Finally, although the chemical reagents can aid in the characterization of a protein, they are rarely useful for altering the function of the target protein.

In the last decade, the development of recombinant DNA technology has offered the protein chemist new, powerful methods for modifying proteins. Site-directed mutagenesis makes it feasible to precisely and completely substitute an amino acid with any other amino acid. Heterologous expression of the mutant DNA sequence permits the production of reagent quantities of the variant protein. Coupled with X-ray

crystallography or NMR spectroscopy, this new approach allows the effect of substituting a specific, single amino acid residue to be addressed in atomic detail. Many early mutagenesis studies were single amino acid substitution experiments at the active site of well-studied model enzymes (For reviews, see Knowles, 1987; Perona & Craik, 1995; Shaw, 1987). Clearly defined questions regarding the role of specific residues were addressed with the assumption that the contributions of these catalytically important amino acids were fairly predominant and independent. Though extremely powerful, the iterative process of mutagenize-analyze-mutagenize can become time consuming and labor intensive if more than a limited number of mutants are analyzed simultaneously. Furthermore, the simplicity and the power of the mutagenesis technology rapidly outstrips the theoretical basis for structure-function relationships for a particular protein. Without extensive *a priori* knowledge of the protein, it is seldom clear what amino acids are appropriate targets for mutagenesis.

The sophistication of contemporary protein modification techniques has provided unprecedented opportunities to create altered or even completely novel functions of proteins, especially enzymes. Although enzymes are highly efficient and specialized, it is sometimes possible to alter or actually improve them for specific reactions. An example of this type of redesign can be seen in naturally occurring enzymes with similar sequences and three-dimensional structures that possess distinct functions. Eventually it should be possible to create synthetic proteins with novel catalytic activities from preexisting structures with sufficient mutations. However, it is not obvious how to do this in most cases. Although there are examples where a few amino acid substitutions have resulted in drastic changes in enzyme function while maintaining sufficient catalytic activity (Higaki *et al.*, 1990; Sloane *et al.*, 1991; Wells *et al.*, 1987; Wilks *et al.*, 1988), it is often found that many substitutions are required to successfully alter the properties of an enzyme (Hedstrom *et al.*, 1992; Wells & Estell,

1988). Therefore, the protein engineer faces the formidable task of locating and identifying the amino acids necessary for altering the function of a protein.

This challenge has stimulated two innovative advances in the field:

- (1) The generation of molecular repertoires with sufficient diversity for altered function to be represented;
- (2) The development of screens or selections to effectively sort through the large repertoires.

It is now common practice to generate large diverse DNA libraries that express variant proteins due to simultaneous and complementary advances in: 1) automated solid phase oligonucleotide synthesis; 2) DNA cloning; and 3) heterologous expression of foreign DNA in bacteria. Figure 1.1 gives an overview of various mutagenesis strategies for altering a protein. In one method referred to as "cassette mutagenesis" (Botstein & Shortle, 1985; Wells *et al.*, 1985), one or a few codons in the gene encoding the target protein are replaced with nucleotide mixtures encoding all twenty amino acids. This creates a library of variant proteins with random mutations in defined positions. Other methods, including mutator strains (McCoy & Khorana, 1983), chemical mutagenesis of DNA (Kadonaga & Knowles, 1985), enzymatic misincorporation of nucleotides (Shortle & Lin, 1985), doped DNA synthesis (Hermes *et al.*, 1989), error prone PCR (Leung *et al.*, 1989), or DNA shuffling (Stemmer, 1994) produce mutations randomly distributed along the target gene (see Figure 1 for comparison). The diversity of these libraries is only limited by the transformation efficiency of *E. coli*, usually on the order of 10^8 to 10^9 colonies per microgram of DNA. The absence of *a priori* knowledge for altering the function of a protein can be overcome by increasing the size and diversity of the variant library and efficiently searching the library for the variant that exhibits the desired activity.

Figure 1.1 Recombinant Methods for Protein Modification

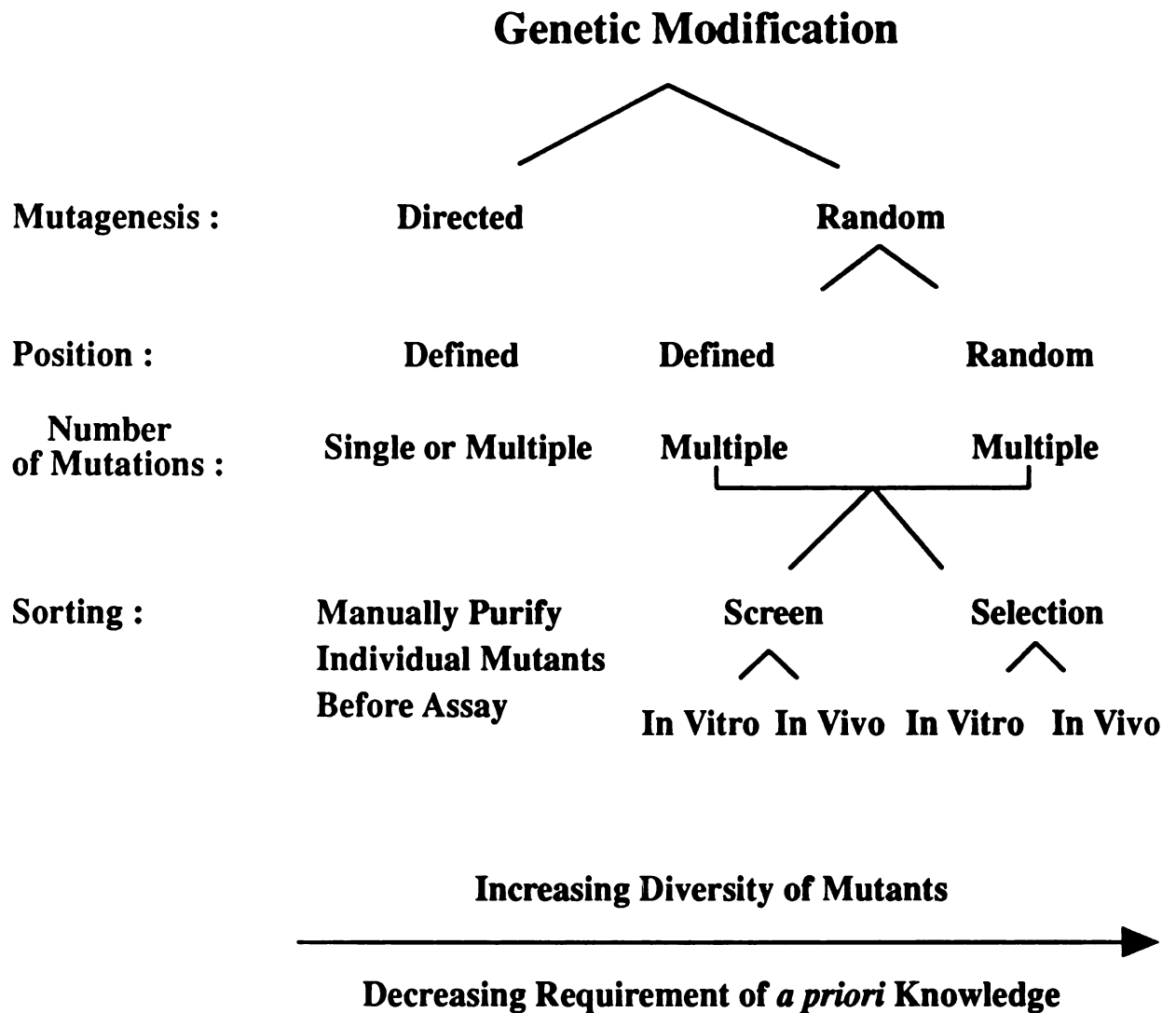


Figure 1.1 Recombinant Methods for Protein Modification

Different methods to analyze protein mutations are classified based on the nature of the mutations, the number of mutations, and the analysis procedures.

The approaches that have been established for sorting through combinatorial molecular repertoires can be grouped into two broad categories: selections and screens. A screen will provide relative scores for a spectrum of mutants under particular assay conditions. It is then up to the researcher to decide which mutant is worthy of further study. A selection procedure, on the other hand, will enrich a subset of mutants that meet the criteria of the experimental design while the rest of the population is lost or discarded. Depending upon the sensitivity and the dynamic range of the selection, the activity of the variant protein determines which mutants survive and undergo further study. Screens and selections can also be defined based on the location of the interaction, either *in vivo* or *in vitro*. Representative examples of each system are listed in Table 1.1.

Table 1.1 Methods for Analyzing Macromolecule Repertoires

		Site of Interaction	
		<i>In Vivo</i>	<i>In Vitro</i>
Method	Screen	• Yeast two-hybrid system [1]	• cDNA cloning by primer hybridization or antibody recognition [4] • Monoclonal antibody [5] • Peptides-on-Pins [6] • "Tea-bag" [7]
	Selection	• cDNA cloning by complementation [2] • Activity selection [3]	• SELEX [8] • Aptamer [9] • Peptides-on-Plasmids [10] • Phage display [11]

Note: [1] (Chien *et al.*, 1991)[2] (Chang *et al.*, 1978)[3] (Evnin *et al.*, 1990)[4] (Sambrook *et al.*, 1989)[5] (Kohler & Milstein, 1975)[6] (Geysen *et al.*, 1984)[7] (Houghten, 1985)[8] (Tuerk & Gold, 1990)[9] (Bock *et al.*, 1992)[10] (Cull *et al.*, 1992)[11] (Smith, 1985)

There are a number of efficient methods for isolating a particular member from a library of molecules by *in vivo* (two-hybrid system) or *in vitro* (antibody screen or oligonucleotide hybridization) screening. However, a selection is inherently more stringent and can analyze more mutants compared with a screening procedure. Although *in vivo* selections have been used widely by geneticists, there are only a few cases in which a library of variants of a particular enzyme or other protein has been

subjected to *in vivo* selection. The target proteins include β -lactamase, lambda repressor, triosephosphate isomerase (TIM), trypsin, thymidine kinase, thymidylate synthase and catalytic antibodies.

A mutant library saturated at position 71 of β -lactamase has been characterized by antibiotic resistance in *E. coli*. The selection revealed that this position was important for protein stability but was not essential for catalysis (Schultz & Richards, 1986). Recently, *in vitro* homologous recombination (DNA shuffling) was used to generate a library of β -lactamase mutants from which a mutant was selected with 32,000-fold increase of minimum inhibitory concentration against the antibiotic cefotaxime (Stemmer, 1994). The selection of functional variants of lambda repressor from a library of dimer interface mutants showed that the number and type of substitutions allowed at each position were extremely variable (Reidhaar & Sauer, 1988). Knowles and coworkers have made a library of mutant triosephosphate isomerases incorporating random mutations that spanned over the entire gene. Variant enzymes were isolated that incorporated mutations that could improve the catalytic efficiency of a sluggish mutant (Hermes *et al.*, 1990). An *E. coli* auxotroph has been used to select trypsin variants that prefer arginine or lysine substrates from a binding pocket library (Evnin *et al.*, 1990; Perona *et al.*, 1993). A partially randomized library involving eleven codons of thymidine kinase has been subjected to genetic selection. Conserved amino acids that may be important for thymidine kinase catalysis were identified (Munir *et al.*, 1992; Munir *et al.*, 1993). Saturated mutagenesis has also been applied to conserved residues in thymidylate synthase (Climie *et al.*, 1990). Catalytically active mutants were isolated by complementation using a Thy^- strain of *E. coli*. The results point to a high degree of flexibility of thymidylate synthase in accommodating function with structural change. Recently, genetic selections have also been used to characterize libraries of potential catalytic antibodies (Lesley *et al.*, 1993; Smiley & Benkovic, 1994).

These encouraging examples illustrate the great potential of using *in vivo* selections to sort through libraries of variant proteins. However, the particular activity under selective pressure has to be related to a vital metabolic step in the host organism. This restriction greatly reduces the number of enzymes or other proteins that may be the target of *in vivo* selection. Besides, our limited understanding of the overall metabolism of the host organism also overshadows the efforts to design and construct such experimental systems. Therefore, the development of *in vitro* selection systems provides a promising alternative to overcome the inherent difficulties in developing a practical genetic selection.

The *in vitro* selection systems for sorting through large nucleic acid libraries (Bartel & Szostak, 1993; Lorsch & Szostak, 1994; Szostak, 1992; Tuerk & Gold, 1990) take advantage of the polymerase chain reaction (PCR) to directly amplify selected individual variants *in vitro*. Since no equivalent of the polymerase chain reaction for amplifying nucleic acids exists for directly amplifying proteins, methods that link the function (phenotype) and identity (genotype) of the protein of interest are required for protein repertoires.

Techniques have been developed that couple chemical tags to peptide sequences (Brenner & Lerner, 1992) (Needels *et al.*, 1993). These "encoded libraries" facilitate rapid identification of a particular member in a library. However, peptides alone cannot be amplified *in vivo*. Therefore, a more versatile and efficient system for screening protein libraries needs to incorporate both the identification and amplification of individual members. One way to achieve this is to link the macromolecule of interest to a biological carrier, which can be uniquely identified and reproduced.

One of these methods is called "peptides-on-plasmids" (Cull *et al.*, 1992; Schatz, 1993). A LacI-peptide fusion protein is specifically bound to the Lac operator sequence

in the plasmid that encodes the peptide sequence. The protein-DNA complex can be isolated and passed through an immobilized ligand matrix to retain those peptides that bind the ligand. After elution from the matrix, the plasmid DNA can be amplified and sequenced. This system allows intracellular expression of peptides on the C-terminus of the *Lac* repressor. The substrate specificity of an *E.coli* biotinylation enzyme has been mapped with peptide libraries based on this system (Schatz, 1993). Other systems use the surface protein on bacteria or virus as the expression platform (Little *et al.*, 1993). The intact bacteria or virus particle becomes the vehicle to display foreign proteins or peptides of interest. *E. coli* surface protein LamB (Charbit *et al.*, 1988) and OmpA (Francisco *et al.*, 1992) have been tested for this purpose. Although each of these methods offer unique insights into the particular protein of interest, the approach that has gained wide acceptance for sorting through libraries of peptide or protein variants is filamentous phage display.

Filamentous Phage Display System

The structure and life cycle of filamentous phage M13 have been the subject of extensive investigation as a model system for virology. The detailed understanding of M13 filamentous phage and its widespread use in recombinant DNA technology permitted the rapid development of it as a display vehicle for foreign proteins and peptides (Smith, 1988). M13 phage can be rapidly propagated through the infection of *E. coli*, resulting in massive amplification of phage clones. The phage particles are stable in a wide range of pH and salt concentrations, allowing flexible conditions for washing and elution procedures. The phage particle displaying the protein of interest can bind to an immobilized ligand matrix through the foreign protein domain and be specifically enriched. The process of phage binding, washing and elution on a solid support is given the name "panning" (Parmley & Smith, 1988)(Figure 1.2). The ability to carry out multiple panning cycles is a definite advantage for the phage display system. Even if the enrichment of a single round of panning is not very significant, strong binders may still be identified through repetitive rounds of panning. This productive loop greatly increases the sensitivity of the *in vitro* selection and enables rare individual clones with target binding properties to be isolated from large, random libraries with up to 10^{8-9} individual clones.

The ease to generate and reamplify large and diverse pools of fusion phage variants, the versatility towards binding conditions and ligands, plus the unique physical linkage between recognition and replication are the unique advantages of the phage display technique. In the phage display system, the selective force is the binding interaction between the phage and ligand in solution. This allows one to manipulate various conditions to enrich the desired phage populations. Parameters like pH, ionic strength, temperature, duration of binding incubation, number of wash steps and concentration

of competing ligands can affect the stringency of the panning experiment. For example, variation of the incubation period with competing ligands can be used to select for phage with slow k_{off} rates (Hawkins *et al.*, 1992). The flexibility of the phage display system to different selection conditions makes it easier to isolate different variant populations from a single pool under adjustable stringencies. Compared with the "life or death" outcome of the genetic selection, this approach can often provide more quantitative information on the nature of the binding interaction. Another distinct advantage of the phage display system is that it poses few constraints on the physical and chemical nature of the ligand and matrix. In a genetic selection, the desired scheme has to fit within the metabolic constraints of the host organism. This limitation greatly reduces the number of possible ligands with unnatural chemical moieties.

George Smith's pioneering work on filamentous phage virology built the foundation for developing the techniques of phage display and making this technology a versatile system for *in vitro* selection of combinatorial libraries. pIII of filamentous phage was identified as a key element for phage infection (Nelson *et al.*, 1981). Further study revealed that the C-terminal domain of pIII is responsible for phage morphogenesis (Crissman & Smith, 1984). In 1985, in a landmark paper (Smith, 1985), Smith demonstrated that foreign DNA fragments can be inserted into filamentous phage gene III to create a fusion protein with the foreign sequence in the middle. The fusion protein is incorporated into the virion, which retains infectivity and displays the foreign amino acids in immunologically accessible form. These "fusion phage" can be enriched more than 1000-fold over ordinary phage by affinity for antibody directed against the foreign sequence. In 1990, Scott and Smith established a phage display system displaying short peptides (Scott & Smith, 1990) as an "epitope library". Since then, phage display has been widely used to search peptide libraries for tight binding peptides to antibodies, receptors or other binding proteins with great

success. An excellent example of the application of peptide phage display is the identification of small peptides as potent mimetics of the protein hormone erythropoietin (EPO) (Wrighton *et al.*, 1996). The crystal structure at 2.8 Å resolution of a complex of this agonist peptide with the extracellular domain of EPO receptor reveals that a peptide dimer induces an almost perfect twofold dimerization of the receptor (Livnah *et al.*, 1996). Today peptide phage display has become a powerful tool for the drug discovery efforts of biotech and pharmaceutical companies.

In basic research, phage display finds its most widespread use in studying the interactions between antibodies and antigens (Winter *et al.*, 1994). Numerous experimental systems have also been established to display enzymes and macromolecular inhibitors on phage to try to explore the potential of this technology. The enzymes that have been displayed include alkaline phosphatase (McCafferty *et al.*, 1991), trypsin (Chapters Two to Four) (Corey *et al.*, 1993), β -lactamase (Soumillion *et al.*, 1994), prostate specific antigen (PSA) (Eerola *et al.*, 1994), staphylococcal nuclease (Light & Lerner, 1995), glutathione transferase (Widersten & Mannervik, 1995), hen egg-white lysozyme (Maenaka *et al.*, 1996), human plasminogen (Lasters *et al.*, 1997). The macromolecular protease inhibitors being displayed are BPTI (Roberts *et al.*, 1992a), PAI-1 (Pannekoek *et al.*, 1993, van meijer, 1996), Alzheimer's amyloid beta-protein precursor inhibitor (APPI) Kunitz domain (Dennis & Lazarus, 1994a; Dennis & Lazarus, 1994b), ecotin (Chapters Five and Six) (Wang *et al.*, 1995; Yang & Craik, 1997), chicken cystatin (Tanaka *et al.*, 1995), first Kunitz domain of human lipoprotein-associated coagulation inhibitor (LACI-D1) (Markland *et al.*, 1996a; Markland *et al.*, 1996b).

Figure 1.2 Diagram of Phage Display System

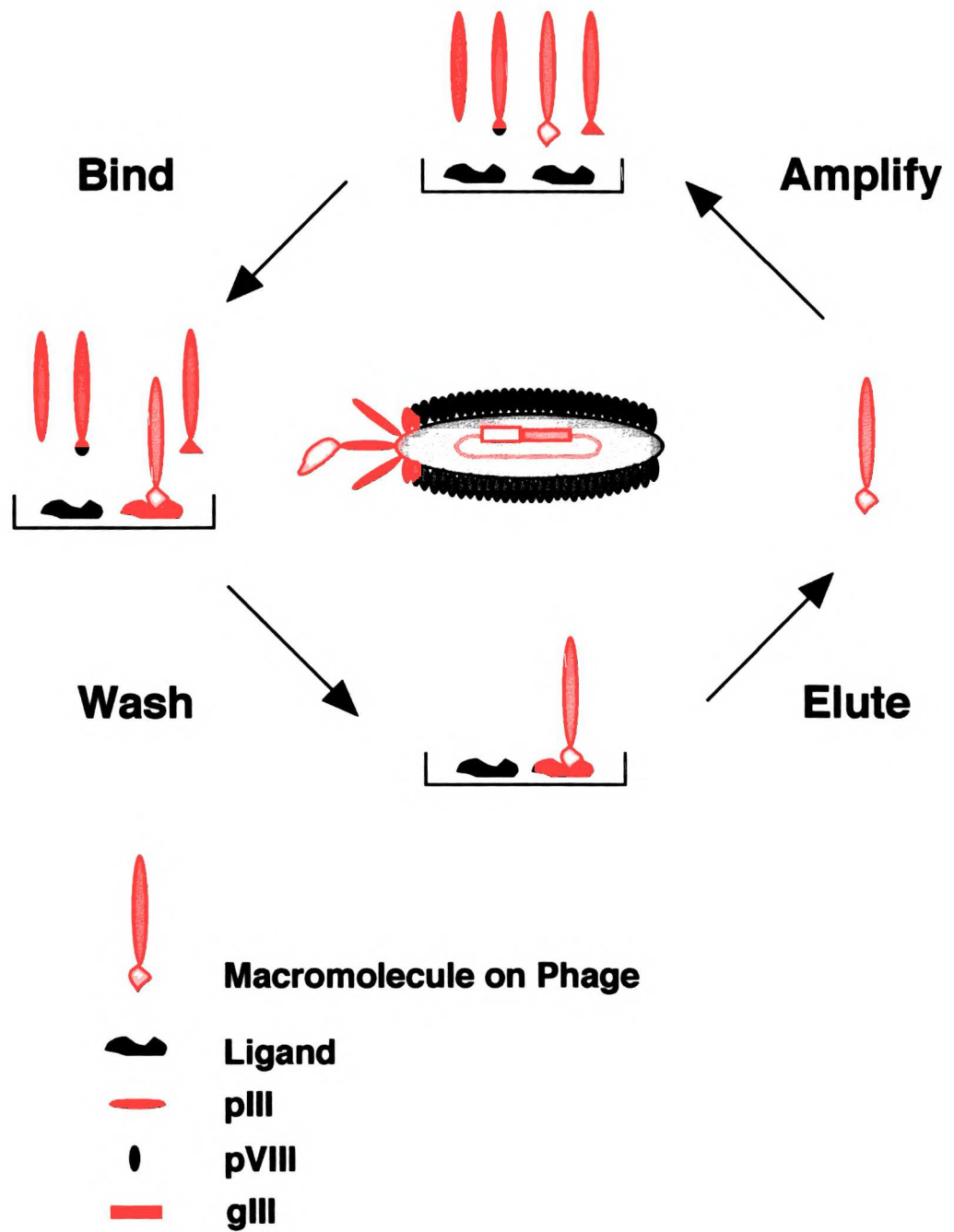


Figure 1.2 Diagram of Phage Display System

Foreign macromolecule (shown in yellow) is displayed on phage through a fusion with minor coat protein pIII (shown in red). The macromolecular that selectively binds to its ligand is captured on sold support while the background phage are washed away. The selected macromolecule on phage is dissociated from the phage by low pH buffer elution. This sub population of phage is then amplified by infecting *E. coli* cells. The amplified phage population will go through additional rounds of panning process until a dominant species emerges.

Phage display systems can be divided into four classes, based on the filamentous coat protein used for display and the requirement of helper phage. These four classes of phage display vectors are denoted as 33, 3+3, 88, and 8+8 systems and are listed in Table 1.2. Among them, 33 and 88 systems are derived from full-length M13 phage vectors, in which the foreign DNA gIII (or gVIII) fusion was inserted. Two copies of gIII (or gVIII) are present in the 33 and 88 systems. By contrast, 3+3 and 8+8 systems are derived from phagemid. Since wild-type pIII coat protein is required for filamentous phage infection, helper phage is used for the assembly of phage displaying foreign peptides or proteins. Only one copy of gIII (or gVIII) is present in the 3+3 and 8+8 phage display vectors.

Table 1.2 Four Classes of Phage Display Vectors

Name	Helper Phage Requirement	Coat Protein for display	Copy of gIII	Copy of gVIII	Examples
33	NO	pIII	2	1	pMTsn::III (Chapter Two)
3+3	YES	pIII	1	1	pBS eco-gIII (Chapter Five)
88	NO	pVIII	1	2	pMTsn::VIII (Chapter Two)
8+8	YES	pVIII	1	1	pTsn::VIII (Chapter Two)

There are distinct advantages and disadvantages for the phage and phagemid systems. The phage system is independent of helper phage thus reduces certain

experimental complexity. In addition, for proteins or peptides that can't be displayed under the phagemid context, the phage system offers an alternative approach and vice versa. The disadvantage of the phage system is the size of the vector. The full-length vector is prone to spontaneous deletion during phage growth and amplification, affecting the stability of the foreign DNA being displayed (Zinder & Boeke, 1982). Phagemid vector is smaller in size and can overcome some of the inherent instabilities of the phage system. In addition, the helper phage offers another level of control on the number of foreign peptides or proteins displayed on phage. By modulating the ratio of helper phage to phagemid, monovalent phage display can be achieved to avoid the avidity problem during the *in vitro* binding and selection process (Lowman et al., 1991). In this study, both the phage and phagemid approaches were taken to display trypsin (33 and 88 systems, Chapter Two) and ecotin (3+3 system, Chapter Five). The potential and limitation of the two different systems are investigated. In summary, there is no general criteria of choosing between the two systems. Experimenting with both approaches are necessary to identify the optimal phage display approach for each macromolecule of interest.

The phage display technique is an innovative tool for engineering enzymes and their macromolecular inhibitors. It offers new opportunities to investigate the combinatorial interactions among the functionally important residues of a target macromolecule. It provides an efficient means to sort through diverse libraries of mutants based on the *in vitro* binding properties of the variant protein. The molecular diversity that the current phage display technique can generate and process permits *in vitro* applied molecular evolution (Kauffman, 1992) experiments, by which novel catalytic or inhibitory functions can be searched within large phage repertoires. The above examples of enzymes and protease inhibitors successfully displayed on phage suggest that this technique may be applied to other experimental systems as well. The continued

development of *in vitro* selection techniques that are highly discriminating and that incorporate activity-based selection criteria will have a profound impact on the field of protein engineering.

Chapter Two: Display Trypsin on Phage

ABSTRACT

The gene encoding mature rat anionic trypsin was fused via a di-glycine linker to the C-terminal domain of geneIII or the entire geneVIII of bacteriophage M13 coat protein. The fusion protein was assembled onto the surface of the filamentous phage. The trypsin on phage formed a complex with ecotin, a protease inhibitor in the periplasm of *E. coli*. Active trypsin phage were purified from ecotin after extensive dialysis at pH 2.2. The trypsin phage could be stained with anti-trypsin antibody after electrophoresis and immunoblotting, indicating that trypsin was expressed on phage. The trypsin phage was active against small synthetic substrates such as benzyloxycarbonyl-Gly-Pro-Arg-aminomethylcoumarin (Z-GPR-AMC) and was inhibited by the tight binding trypsin inhibitor bovine pancreatic trypsin inhibitor (BPTI). These results showed that trypsin on phage was successfully secreted to the periplasm, could be correctly folded and achieved normal catalytic activity while fused to geneIII or geneVIII proteins (pIII or pVIII). The expression level of trypsin pVIII fusion was significantly lower than that of trypsin pIII fusion. The fusion phage possessed similar steady-state kinetic parameters (k_{cat} and K_m) to that of wild type trypsin. A part of this work has been published (Corey *et al.*, 1993).

Introduction

Based on the known three-dimensional crystal structures, the serine protease family of enzymes can be divided into five structural classes. Representative members of each of these classes are trypsin, subtilisin, wheat serine carboxypeptidase (Liao *et al.*, 1992), CMV protease (Qiu *et al.*, 1996; Shieh *et al.*, 1996; Tong *et al.*, 1996) and the Clp protease (Chung *et al.*, 1996; Shin *et al.*, 1996). The subfamily that trypsin represents includes some of the most extensively studied proteases at the structure-function level such as trypsin, chymotrypsin and elastase. These enzymes share the same dual anti-parallel β -barrel scaffold with a well-developed S1 pocket that accommodates the substrate P1 side chain (Perona & Craik, 1995). The exceptional understanding of these proteases stimulates fundamental questions regarding the evolution of their highly precise and sophisticated substrate recognition properties and catalytic machinery. In addition, the prevalence of serine proteases in vital biological pathways and their widespread use as essential biochemical reagents are also strong driving forces for the extensive research to modify and alter the properties of these enzymes.

The accumulated knowledge on the family of serine proteases encourages protein engineering efforts to introduce novel substrate specificity into the protease. One way to accomplish this goal is through the careful examination of enzyme sequence and structure and the subsequent site-directed substitution of amino acids. This powerful, simple, and direct approach has been employed to successfully alter the non-polar substrate-binding pockets of α -lytic protease (Bone *et al.*, 1991), subtilisin (Wells & Estell, 1988), and 15-lipoxygenase (Sloane *et al.*, 1991). However, our knowledge of many other enzymes, particularly those with more polar substrate-binding pockets, is not adequate to allow us to effectively choose substitutions that will yield mutated enzymes with desired altered

specificities. The combination of electrostatic and steric variables currently precludes sufficiently accurate analysis.

Attempts to convert trypsin, a serine protease which is specific for arginine or lysine at the P1 position of substrates, into chymotrypsin, which is specific for large hydrophobic residues, have only been partially successful, despite extensive mutagenesis and high homology between the enzymes (Hedstrom *et al.*, 1992; Perona & Craik, 1995). Experiments to alter the Arg/Lys specificity (Craik *et al.*, 1985) or to introduce specificity for acidic side-chains (Graf *et al.*, 1987), though informative regarding structure/function relationships, have not yielded enzymes possessing altered specificities while retaining catalytic efficiency.

This lack of a detailed understanding of substrate binding can be overcome in the design process by the coupling of random mutagenesis with selections or screens. A genetic selection strategy for novel protease activity using auxotrophic bacteria has been developed (Evnin *et al.*, 1990; Perona *et al.*, 1993). Each selection, however, depends on the enzymatic conversion of a non-nutritive source of an amino acid to its natural form in a bacterial strain that is auxotrophic for that amino acid. Construction of such strains is not readily accomplished for certain amino acids and may not be possible for ligands that are not utilized metabolically. An alternative strategy is phage display (Smith, 1985). This affords a complement to *in vivo* selection strategies because it relies on *in vitro* binding rather than catalysis. Since it is possible to couple a wide range of chemical compounds to a matrix for screening purposes, the phage display technology can be readily extended to assay the binding of unnatural ligands.

To investigate phage display as a tool for redesigning proteases, wild-type trypsin was fused to either gene III or gene VIII coat protein. The stability, extent of trypsin display, and catalytic properties of the resulting chimeric phage were determined. The co-

Results

Design of the Trypsin Phage Display System

The mechanism of zymogen activation of trypsinogen to trypsin requires a unique cleavage event near the N-terminus of the protein at the Lys15-Ile16 peptide bond (Fehlhammer *et al.*, 1977; Huber & Bode, 1978). The cleavage event releases an N-terminal fragment, the activation peptide, and the new N-terminus then forms a specific interaction with Asp194 of the enzyme. This interaction at Ile16 restructures the "oxyanion hole", a critical component of the active site (Kossiakoff *et al.*, 1977). By deleting the sequences encoding the activation peptide and juxtaposing the signal peptide to the enzyme sequences beginning with Ile16, mature, active trypsin could be produced in the periplasm of *E. coli*.

The three-dimensional structure of trypsin reveals a helical peptide at the C-terminus of the enzyme (Kossiakoff *et al.*, 1977). Although the C-terminal residue makes specific interactions with the rest of the molecule, it appeared that a Gly-Gly spacer could be used to link trypsin to the C-terminal domain of the gene III or the entire gene VIII protein of M13 bacteriophage (Figure 2.1). Rat anionic trypsin was expressed on the surface of M13 bacteriophage as a fusion protein with coat protein pIII or pVIII. The expression level was transcriptionally controlled by a tac promoter, allowing induction with isopropylthio- β -D-galactoside (IPTG). The fusion protein was expressed and exported into the periplasm of *E. coli* by a hisJ signal sequence. Processing of the signal sequence by bacterial signal peptidase resulted in active trypsin-pIII or -pVIII being assembled into the newly formed bacteriophage.

Rat anionic trypsin can be expressed in the periplasm of *E. coli* utilizing a tac promoter and a His J signal sequence in the phagemid pT3 (Corey & Craik, 1992; Higaki *et al.*, 1989). An MluI restriction endonuclease site was introduced at the C-terminus of the

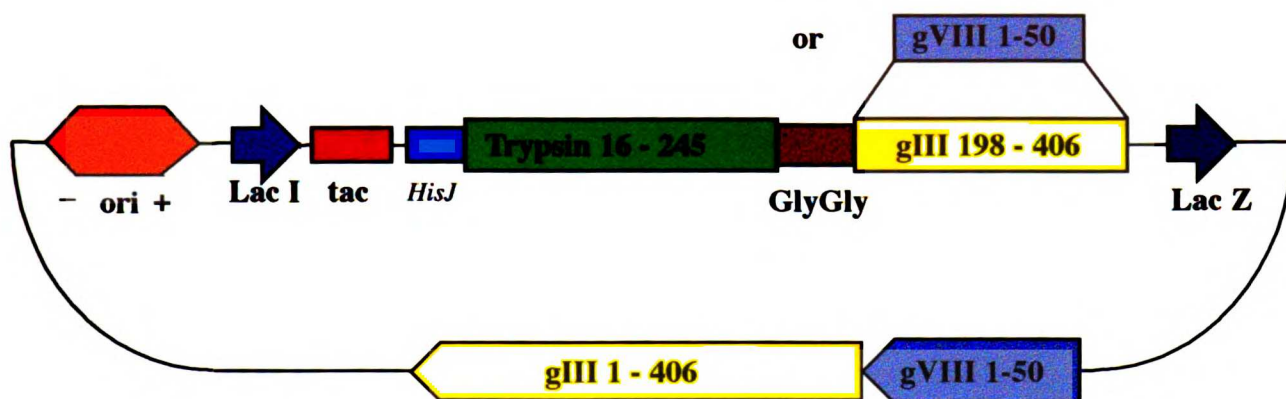
trypsin gene in pT₃ via site-directed mutagenesis (Kunkel, 1985). The genes coding for either residues 198-406 of the gene III protein or the entire gene VIII protein were amplified from M13mp18 using the polymerase chain reaction with flanking primers containing MluI and HindIII restriction endonuclease sites. These genes were inserted into the pT₃-derived phagemid bearing the introduced MluI site. Subsequent mutagenesis was performed to remove the MluI site and introduce a Gly-Gly spacer between trypsin and the coat proteins. The fusions were initially expressed in phagemid vectors pTsn::III and pTsn::VIII.

Further manipulations were performed to introduce the gene fusions directly into M13mp18. A unique restriction endonuclease site, XbaI, was introduced 5' of the tac promoter of pTsn::III and pTsn::VIII. XbaI/HindIII digestion allowed the trypsin-Gene III and trypsin-Gene VIII fusions to be excised and cloned directly into the polylinker region of M13mp18. This yielded two constructions containing the genes coding for either the trypsin-Gene III fusion and the wild-type Gene III coat protein (pMTsn::III), or the trypsin-Gene VIII fusion and the wild-type Gene VIII coat protein (pMTsn::VIII) (Figure 2.1). Similarly, the XbaI/SalI DNA fragments from pTsn::III or pTsn::VIII were used to insert the gene fusions into Bluescript SK⁻ (Stratagene) to yield alternate phagemid vectors pBSTsn::III and pBSTsn::VIII.

Figure 2.1 Trypsin Phage Display Vector

Trypsin phage display vector pMTsn::III or pMTsn::VIII is shown. The same trypsin gIII or gVIII fusion cassette was tested in three different vectors, pT₃, M13mp18 and pBluescript, resulting in three sets of constructions pTsn::III (VIII), pMTsn::III (VIII), and pBS Tsn::III (VIII), respectively.

Figure 2.1 Trypsin Phage Display Vector



pMTsn::III 8.7 kb or pMTsn::VIII 8.4 kb

Preparation of Fusion Phage and Choice of a Phage Display System

Trypsin phage were obtained from the phagemid constructs pTsn::III and pTsn::VIII using VCSM13 as a helper phage. Bacteriophage were purified via PEG-precipitation followed by dialysis using a 300,000 Dalton molecular weight cut-off Microsep micro concentrator. Addition of the Tsn::III and Tsn::VIII trypsin-phage to a solution of the fluorogenic substrate Z-Gly-Pro-Arg-aminomethyl coumarin (Z-GPR-AMC) indicated that the bacteriophage possessed trypsin-like hydrolytic activity. Immunoblot analysis of the isolated helper phage and phagemid with anti-trypsin antibody confirmed that trypsin was being displayed on the phage coat (Figure 2.2). Trypsin was displayed on both the helper phage and the phagemid. Unfortunately, in spite of repeated attempts, every combination of experimental conditions resulted in a large excess (10-100 fold) of helper phage relative to phagemid. A preponderance of helper phage would hinder future selection experiments by increasing non-specific binding to the affinity matrix. In addition, the fraction of helper phage which express trypsin on their surface would compete with the trypsin-phagemid for binding to immobilized ligand.

Since the ratio of phagemid/helper phage production with the parent trypsin expression phagemid pT₃ was high, it seemed likely that the attachment of the coat protein genes to the gene encoding trypsin was disrupting some aspect of phagemid replication. It was unlikely that the incorporation of trypsin on the phage accounted for the low yields, since the helper phage, which also displayed trypsin, were well expressed. To improve phagemid yield the coat protein-trypsin gene fusions were introduced into pBluescript SK⁻ to generate pBSTsn::III and pBSTsn::VIII. The expectation was that utilization of the higher copy number pBluescript vector might yield more phagemid. Phage preparations with these constructions did afford an excess of phagemid relative to helper phage, confirming that the

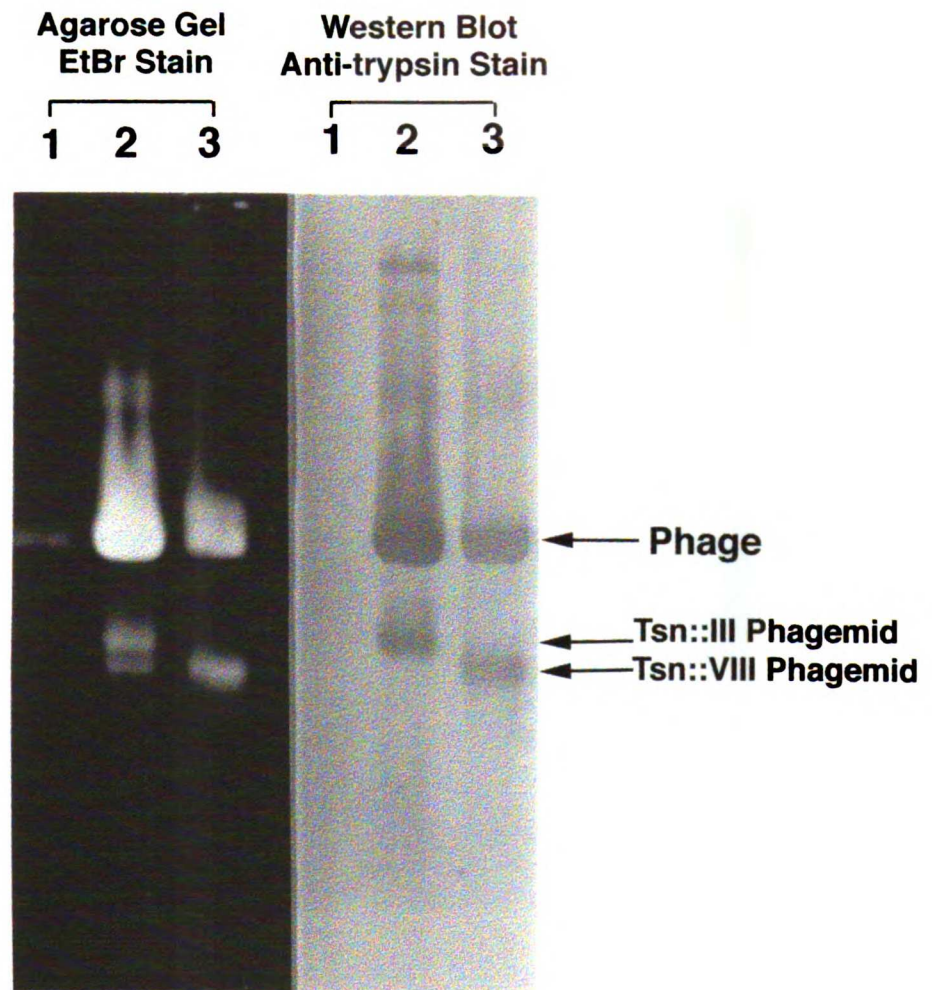
low levels of phagemid synthesis were due to the genetic context of the original phagemid system and not to a systematic problem resulting from the expression of the protease-coat protein fusion.

The requirement for helper phage was avoided altogether by directly inserting the genes encoding the fusion proteins into M13mp18 to yield pMTsn::III (gIII-trypsin fusion) and pMTsn::VIII (gVIII-trypsin fusion). The isolated phage were hydrolytically active and immunoblot analysis indicated that bacteriophage containing the fusion genes displayed trypsin on their surface (Figure 2.3). The specific covalent nature of the attachment of trypsin to phage was established by comparing immunostaining of a free trypsin/wild-type bacteriophage M13mp19 mixture with immunostaining of the trypsin phage (Lanes 1, 3, 4). The free trypsin mixed with bacteriophage appeared as a diffuse smear when stained with anti-trypsin antibodies. In contrast, the bacteriophage designed to covalently express the fusion proteins (Lanes 3 and 4) stained as a compact band. The small difference in mobility of the MTsn::III and MTsn::VIII fusion phage most likely reflected the different sizes of the gene III DNA and gene VIII DNA. The direct expression of the fusion protein within an M13 vector circumvented the technical complexities inherent in helper phage/phagemid systems. This strategy may be the most simple and rapid way to evaluate the expression of proteins on the surface of bacteriophage.

Figure 2.2 Trypsin Phage Tsn::III or Tsn::VIII Derived From a Phagemid/Helper Phage System

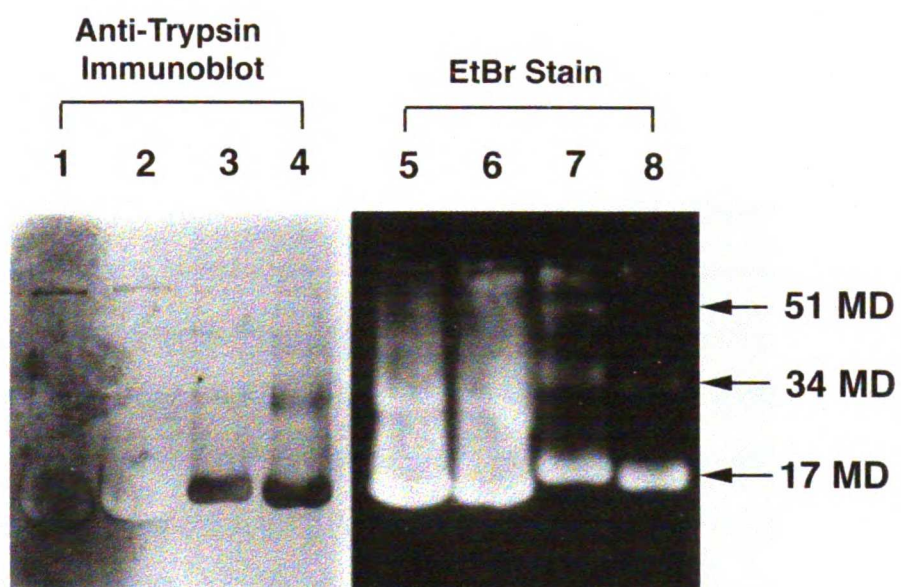
Tsn::III and Tsn::VIII Phage were grown using VCSM13 helper phage. The expression of the trypsin-coat protein fusions was induced 1 hour after the addition of helper phage by the addition of IPTG to a final concentration of 200 μ M. Phage was harvested as described in Methods. An aliquot of 10 μ l of phage was loaded in each lane. The majority of the trypsin was incorporated onto full length phage particles instead of phagemid.

**Figure 2.2 Trypsin Phage Tsn::III and Tsn::VIII
Derived From a Phagemid System**



Lane 1 M13 Phage
Lane 2 Tsn::III Phage
Lane 3 Tsn::VIII Phage

Figure 2.3 Trypsin Phage MTsn::III and MTsn::VIII



Lanes 1,5 M13mp19 plus free trypsin
Lanes 2,6 M13mp18
Lanes 3,7 MTsn::III phage
Lanes 4,8 MTsn::VIII phage

Figure 2.3 Trypsin Phage MTsn::III and MTsn::VIII

Bands at 17, 34 and 51 MDa are indicative of monomeric, dimeric and trimeric polyphage, respectively. An aliquot of 10 µl of phage sample was loaded in each lane.

Co-expression of an endogenous *E. coli* protease inhibitor, Ecotin

E. coli produces an endogenous serine protease inhibitor, ecotin (Chung *et al.*, 1983; McGrath *et al.*, 1991), which can bind to most of the trypsin produced by our heterologous expression system in bacteria. Immunoblot analysis of bacteriophage stored in 10 mM MES pH 6.0 buffer with anti-ecotin antibodies indicated that ecotin was co-purified with the trypsin-phage (Figure 2.4). To accurately gauge the enzymatic activity of the fusion phage or to conduct affinity binding experiments it was necessary to dissociate ecotin. This was accomplished by resuspending the MTsn::III or MTsn::VIII trypsin-phage in 0.1 N HCl/Glycine, pH 2.2, containing 1 mg/ml BSA. The BSA was required for bacteriophage solubility at the low pH. The catalytic activity of the trypsin phage stored at low pH increased ten-fold relative to those stored at pH 6.0, confirming that most of the trypsin had been bound to ecotin at the higher pH. Agarose gel electrophoresis of intact phage that had been stored at low pH followed by immunoblot analysis with anti-ecotin antibodies indicated that ecotin was no longer present (Figure 2.5). The infectivity of phage resuspended at pH 2.2 remained constant over several weeks.

Figure 2.4 Trypsin Phage - Ecotin Complex

Trypsin phage forms a non-covalent complex with free ecotin in the periplasm of *E. coli*. Ecotin-trypsin phage complex is stable during the phage purification procedure. Trypsin phage is inactive in the presence of ecotin. Also illustrated are the 33 vector pMTsn::III and 88 vector pMTsn::VIII.

Figure 2.4 Trypsin Phage - Ecotin Complex

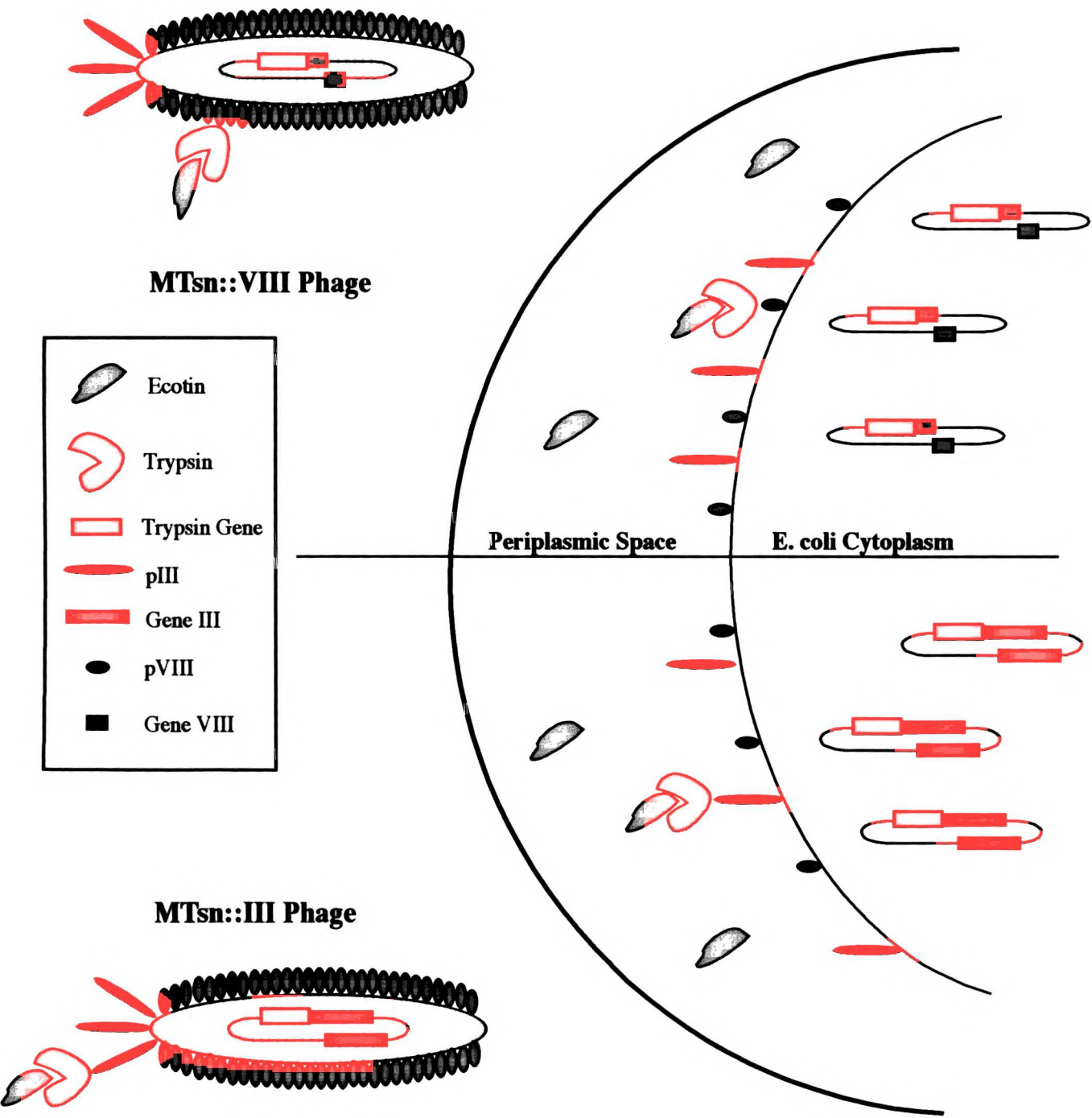


Figure 2.5 Immunoblot Analysis of Ecotin-Trypsin Phage Complex

anti-trypsin					+	+	+	+
anti-ecotin	+	+	+	+				
pMTsn::III		+		+		+		+
pMTsn::VIII	+		+		+		+	
pH 6.0	+	+			+	+		
pH 2.2			+	+			+	+
	1	2	3	4	5	6	7	8



Figure 2.5 Immunoblot Analysis of Ecotin-Trypsin Phage Complex

Lanes 1-4 show the result of treatment with anti-ecotin antibodies. Lanes 5-8 show the result of treatment with anti-trypsin antibodies. Lanes 1,5; MTsn::VIII phage, stored at pH 6.0. Lanes 2,6; MTsn::III phage, stored at pH 6.0. Lanes 3,7 MTsn::VIII phage, stored at pH 2.2. Lanes 4,8; MTsn::III phage, stored at pH 2.2.

Characterization of Trypsin activity

The trypsin phage were isolated via PEG precipitation and resuspended in 0.1 N HCl-glycine, pH 2.5, containing 1 mg/ml BSA. The solution ($\sim 10^{12}$ phage/ml) was then subjected to dialysis against the same low pH buffer using a 300,000 MW cutoff microconcentrator. This dialysis was continued until no trypsin activity could be detected in the flow-through. The concentration of trypsin in the dialyzed phage stock was determined by titration with bovine pancreatic trypsin inhibitor (BPTI) and the kinetic constants k_{cat} and K_m were determined using a fluorogenic substrate, Z-GPR-AMC (Table 2.1).

Table 2.1 Kinetic Constants of MTsn::III and MTsn::VIII Phage

	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m ($\text{min}^{-1}\mu\text{M}^{-1}$)
Trypsin	3050 ± 200	8.8 ± 1	350 ± 20
MTsn::III Phage	2500 ± 400	42 ± 12	75 ± 34
MTsn::VIII Phage	350 ± 9	16 ± 1	23 ± 1

The K_m 's for both the MTsn::III and the MTsn::VIII phage were increased slightly relative to wild-type trypsin. The turnover number, k_{cat} , for the MTsn::III-trypsin phage was similar to that of trypsin (2500 vs. 3050 min^{-1}), while k_{cat} for the MTsn::VIII-phage was less (350 min^{-1}) than trypsin. The similarity of the kinetics of the trypsin phage to

those of trypsin indicates that the six disulfide bonds within trypsin had formed correctly and that the protease was properly folded despite its fusion to the N-terminus of the bacteriophage coat proteins. This similarity also suggests that attachment of an enzyme to the phage surface does not block catalysis.

Determination of the concentration of infective phage (pfu) and total phage revealed that a relatively small fraction of both the MTsn::III and the MTsn::VIII trypsin phage were infective when compared to M13mp18 (Table 2.2). The low ratio of infective trypsin phage relative to total phage may be an indication that the protease is having a significant effect on infectivity. The number of trypsin molecules per phage was calculated to be 0.14 (MTsn::III trypsin phage) or 0.17 (MTsn::VIII trypsin phage). There were about one to two trypsin molecules displayed per ten phage particles. Therefore, it was most likely that there was no more than one trypsin per phage, i.e. the trypsin phage was monovalent. The similarity in these numbers suggests that the relatively low level of trypsin expression (<1 mg/L media) may be limiting the incorporation of the coat protein-trypsin fusions into the phage coat. A higher level of expression may yield a higher valency. Proteolysis of the trypsin-coat protein fusions prior to insertion into the bacteriophage coat may also limit the pool of fusion protein available for display.

Table 2.2 Characterization of Trypsin Phage

Phage	pfu (nM)	phage (nM)	Trypsin-phage (nM)	Trypsin/phage
MTsn::III	2.5 ± 0.5	29 ± 3	4 ± 1	0.14 ± 0.3
MTsn::VIII	3.4 ± 0.5	29 ± 3	5 ± 1	0.17 ± 0.3
M13mp19	35 ± 3	72 ± 10	---	---

Discussion

The construction of trypsin phage display vector pMTsn::III and pMTsn::VIII demonstrates that there is no easy and simple rational for designing an effective phage display system. The series of trypsin phage display vectors are compared and summarized in Table 2.3.

Table 2.3 Trypsin Phage Display Vectors

Name	Class	Phage/ Phagemid	Helper Phage	Performance
pTsn::III	3+3	Phagemid	Yes	Excess of phage vs. phagemid Results irreproducible
pTsn::VIII	8+8	Phagemid	Yes	Excess of phage vs. phagemid Results irreproducible
pBS Tsn::III	3+3	Phagemid	Yes	Results irreproducible
pBS Tsn::VIII	8+8	Phagemid	Yes	Results irreproducible
pMTsn::III	33	Phage	No	Active trypsin phage displayed
pMTsn::VIII	88	Phage	No	Active trypsin phage displayed

As the fusion protein partner, the phage coat protein pIII or pVIII does not have a significant impact on the performance of the phage display vector. However, the choice between a phage-based vs. a phagemid-based system directly affects whether active trypsin phage can be reproducibly obtained. Based on the above experimental results, the full-length phage vector is the only available platform for displaying trypsin. Kinetic characterization shows that there is less than one trypsin molecule per phage. This result is especially significant in a pVIII-based system, in which there is less than one trypsin-pVIII

fusion protein for every 2,800 of wild-type pVIII major coat proteins expressed on the surface of the filamentous phage. The exceptionally low level of trypsin displayed on phage correlates with the low level expression of wild-type trypsin in *E. coli*. The toxicity of a fully active protease in *E. coli* could be a critical selective pressure for keeping trypsin expression at an insignificant level. Although both pMTsn::III and pMTsn::VIII vectors generate fully active trypsin phage with comparable yield, the catalytic efficiency (k_{cat}/K_m) of MTsn::III is three-fold higher than that of MTsn::VIII. Therefore, the pIII based pMTsn::III vector is used for subsequent library construction and trypsin variant phage preparation.

Materials and Methods

Materials

Enzymes and reagents for molecular cloning were purchased from New England Biolabs (Beverly, MA) and were used following the manufacturer's instructions. The *E. coli* strain JM101, and the VCSM13 helper phage were from Stratagene (La Jolla, CA). Rat trypsin was expressed in *E. coli* using the expression vector pZ3 and purified as described. The chromogenic substrate Na-benzyloxy-carbonyl-L-glyylproylarginine-7-amino-4-methylcoumarin (Z-GPR-AMC) was purchased from Bachem Biosciences, Inc (King of Prussia, PA). 4-Methylumbelliferyl *p*-guanidinobenzoate was from Sigma (St. Louis, MO). Bovine serum albumin (BSA) was purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN). α -³⁵S-dATP was from DuPont NEN (Boston, MA). Sequenase Version 2.0 sequencing kit was from U.S. Biochemical Corp (Cleveland, OH). GeneClean® was from Bio101, Inc (La Jolla, CA). Filtron 300 KD columns was from Filtron (Beverly, MA). Oligonucleotides were synthesized with an Applied Biosystems 391 DNA synthesizer (Foster City, CA).

Construction of the Trypsin Phage Display Vector

The plasmid pT₃ is a pBR322 based plasmid for expression of mature rat trypsin in the periplasm of *E. coli*. An MluI restriction endonuclease site was introduced into the parent expression phagemid pT₃ (Higaki et al., 1989) at the terminus of the gene coding for trypsin. The pT₃-MluI DNA was cut with MluI and HindIII to remove a 100 bp fragment and MluI/HindIII PCR fragments coding for either residues 1-50 of the Gene VIII protein or residues 249-406 of Gene III protein were then inserted. The forward and reverse PCR primers for Gene III were 5'-CCG-GAC-GCG-TGT-CCA-TTC-GTT-TGT-GAA-TAT-CAA-GGC-CAA-TCG-3' and 5'-TTG-GAA-GCT-TTA-AGA-CTC-CTT-ATT-ACG-CAG-TAT-GTT-AGC-3' respectively (design based on sequence data from van

Wezenbeek et al., 1980). The forward and reverse PCR primers for Gene VIII were 5'-G-GAC-GCG-TGT-GCT-CAG-GGT-GAG-GGT-GAC-GAT-CCC-GCA-AAA-GCG-3' and 5'-G-GAA-GCT-TTA-GCT-TCG-TTT-CGA-GGT-GAA-TTT-CTT-AAA-3'. The MluI restriction endonuclease site, which encodes a Cys-Ala linkage between trypsin and either coat protein was mutated to code for a GlyGly linkage. The phagemid trypsin display vector was named pTsn::III and pTsn::VIII, respectively. To construct the phage trypsin display vector, a unique XbaI site was introduced at the 5' of the Tac promoter in pTsn::III and pTsn::VIII. An XbaI/HindIII fragment that included a tac promoter, the hisJ signal peptide, mature trypsin, a Gly-Gly linker and part of gIII or the entire gVIII was removed from the parent plasmid pTsn::III or pTsn::VIII and cloned into the polylinker region of M13mp18 to produce trypsin phage display vectors pMTsn::III and pMTsn::VIII.

Trypsin Phage Growth and Preparation

Trypsin phage display vector pMTsn::III or pMTsn::VIII DNA was transformed into a male strain (F') of JM101 using standard heat shock procedures. The transformed cells were plated on LB plates with JM101 lawn cells and top agar. The plate was incubated at 37 °C overnight. Trypsin phage appeared as clear plaques against the background of lawn cells. Single phage plaques were picked with sterile micro pipette tips and transferred into 3 ml 2YT medium with 60 µl of fresh JM101 F' lawn cells from an overnight culture. The phage was grown at 37 °C with vigorous shaking for one hour, then transferred into 100 ml 2YT liquid medium and grown for six hours with shaking. The phage particles were harvested by centrifugation of the cell pellet and precipitation of the supernatant at 4 °C for 4 hr to overnight with 20% PEG 8000/2.5 M NaCl (1/4 of final volume). The trypsin phage was collected by centrifugation at 14,000 g for 20 minutes and resuspended in 0.1 N HCl/Glycine buffer (pH 2.2) with 1 mg/ml bovine serum albumin (BSA). The phage sample was then dialyzed extensively in the same low pH buffer with a 300 KDa cut-off Filtron micro concentrator to remove bound ecotin, an endogenous protease inhibitor in the

periplasm of *E. coli*. After dialysis, the trypsin phage was stored in siliconized microfuge tubes at 4°C for further characterization or manipulation in biopanning experiments. The typical yield was about 10^{13} pfu in total and the concentration was approximately 10^{10} pfu/ μ l.

Immunoblotting Analysis of Trypsin Phage

A trypsin phage sample of approximately 5×10^{10} cfu was loaded in duplicate onto a single 1% agarose gel with 25 mM HCl/250 mM Glycine Tris pH 8.6 as the running buffer. The gel was electrophoresed at 6 mA constant current for 16 hours in the same buffer. The gel was cut into two duplicate sets. One set of samples was transblotted onto a nitrocellulose filter and allowed to react with rabbit anti-trypsin antibodies followed by reaction with horseradish peroxidase-conjugated goat anti-rabbit IgG antibodies. The filter was stained with hydrogen peroxide and 4-chloro-1-naphthol in 20% methanol aqueous solution. The other set of samples was denatured by soaking the agarose gel in 0.5 N NaOH for 4 hours, washed by soaking in water for 4-8 hours, stained with ethidium bromide (EtBr) for visualization of the DNA bands with UV illumination. An immunoblot of the agarose gel using nitrocellulose paper allowed the localization of trypsin phage with anti-trypsin polyclonal antibodies. The existence of an ecotin-trypsin phage complex was observed if anti-ecotin antibodies were used to probe this immunoblot before the low pH buffer dialysis step described above. Ecotin was dissociated from trypsin phage during the dialysis step as shown by subsequent immunoblot analysis.

Klnetic Characterization of Trypsin Phage

The catalytic activity of trypsin phage was measured by a fluorescence assay using the amide substrate Z-GPR-AMC (benzyloxycarbonyl-Gly-Pro-Arg-aminomethyl-coumarin). The assays were performed in 50 mM Tris/HCl, pH 8.0, 20 mM CaCl₂, 100 mM NaCl at 25°C with 10 μ l of a 5 mM Z-GPR-AMC solution. The total reaction volume was 1000 μ l.

Ten to fifty μl of purified trypsin phage stock were neutralized with 10 μl of 1 M Tris/HCl, pH 8.0 solution, then added to the cuvette to start the reaction. The data acquisition time took between 10-30 minutes depending on the concentration of the phage preparation. The slope of the fluorescence data was converted to the turnover rate of the reaction. The kinetic parameters k_{cat} and K_m were calculated from these rates. The number of trypsin molecules on phage was determined by active site titration of trypsin activity with a tight-binding inhibitor, bovine pancreas trypsin inhibitor (BPTI). Trypsin phage was incubated with a serial dilution of BPTI for 15 minutes. The remaining catalytic activity of trypsin phage was determined by measuring the substrate turnover rate of the sample. Assuming that the dissociation of BPTI from the trypsin-BPTI complex was negligible, the data were extrapolated to zero activity, yielding the concentration of active trypsin molecules. The total number of phage particles (virions) was quantified by ethidium bromide staining of the phage genome after agarose gel electrophoresis (Parmley & Smith, 1988). The fluorescence intensity of the phage band was compared with that of a series of DNA standards of known concentrations. The plaque forming units (pfu) were measured by titrating the phage stock with JM101 (F') lawn cells.

Chapter Three : Engineering Trypsin Variant Phage and Construction of a Trypsin Binding Pocket Library

ABSTRACT

Trypsin active site variants D189S and D102N were independently engineered on phage to facilitate the development of *in vitro* selection procedures. Based on the three-dimensional structure of trypsin, a trypsin binding pocket library was designed. This library, in which residues 216, 225, and 226 were randomized, was constructed on the trypsin phage display vector pMTsn::III which contains trypsin fused to gIII. A complete trypsin phage library with over 6.7×10^5 individual clones was generated and characterized. The generation of a phage library at the binding pocket of trypsin is a critical step towards the *in vitro* selection of trypsin variant with altered substrate specificities. To increase the yield and efficiency of trypsin phage purification, two *E.coli* strains with deletions of the ecotin (an endogenous protease inhibitor) gene were constructed and characterized.

Introduction

Rat trypsin is a well-established model system used in protein engineering efforts using the techniques of DNA cloning (Craik *et al.*, 1984), heterologous expression (Craik *et al.*, 1985) and site-directed mutagenesis (Craik *et al.*, 1985). Rat trypsin active site variants D102N (Craik *et al.*, 1987) and D189S (Graf *et al.*, 1987) have provided new insights on the roles of these amino acid in catalysis and substrate binding. Introducing these two important substitutions onto trypsin phage not only creates an opportunity to examine the impact of these active site mutations in a new context, but also offers a useful tool to establish and improve the *in vitro* selection protocol for trypsin phage.

Phage display can be used to isolate variant macromolecules with novel properties but that maintaining pre-existing protein scaffold. To do this, the library generation using site-directed mutagenesis must be combined with the *in vitro* selection of phage display. The success of such a combinatorial approach depends on two essential prerequisites. The first requirement is the availability of a macromolecular repertoire that has sufficient diversity to include all the potential variants with new properties. The second requirement is a sensitive and effective *in vitro* selection system that can capture and enrich the target variants through limited iterative amplification. A trypsin phage display system has been previously established (Chapter Two). Trypsin displayed on phage was fully active and possessed normal catalytic parameters. Taking advantage of these preliminary results, substantial effort was made to fulfill the above prerequisites for the trypsin phage display system. In this chapter, the construction of a complete trypsin phage library is described. Chapter 4 discusses the development of bio-panning procedures for the trypsin phage.

Results

Trypsin Phage Variants D189S and D102N

Since trypsin variants can be made using the vector pZ3 through Kunkel mutagenesis, a direct cloning method was used to introduce the active site mutations into trypsin phage (Figure 3.1). A BstXI/StyI fragment of the pZ₃ vector (which contains Trypsin D102N or D189S genes) was inserted into pMTsn::III vector to yield pMTsn D102N::III and pMTsn D189S::III. The D189S variant also carried a new SacI site at residue 209 of the trypsin gene. The codon of residue 209 (Leu) was changed from CTG to CTC. This restriction site not only provides a unique marker for the D189S substitution, but also will be used as one of the two unique restriction sites to introduce the random library primer cassette, described in detail in the next section. For Kunkel mutagenesis (Kunkel, 1985), uracil-laden single-stranded DNA template of trypsin phage display vector pMTsn::III was made. An oligonucleotide primer encoding the D102N mutation was used to mutagenize the parent vector. However, repeated attempts of mutagenesis were unsuccessful. Thus, the restriction digestion and religation strategy was used for subsequent library construction experiments.

The substitutions were confirmed by single strand DNA sequencing of the mutation region. pMTsn D102N::III and pMTsn D189S::III were transformed into JM101 F'. Trypsin phage MTsn D102N::III and MTsn D189S::III were harvested and purified as described. The proteolytic activity of trypsin active site variant phage was measured by substrate hydrolysis of Z-GPR-AMC. The results are shown in Table 3.1.

Figure 3.1 Construction of Trypsin D189S Phage

Trypsin phage vector pMTsn D102N::III was prepared by the same method.
Trypsin variants fused with gVIII were also constructed based on the same method.

Figure 3.1 Construction of Trypsin D189S Phage

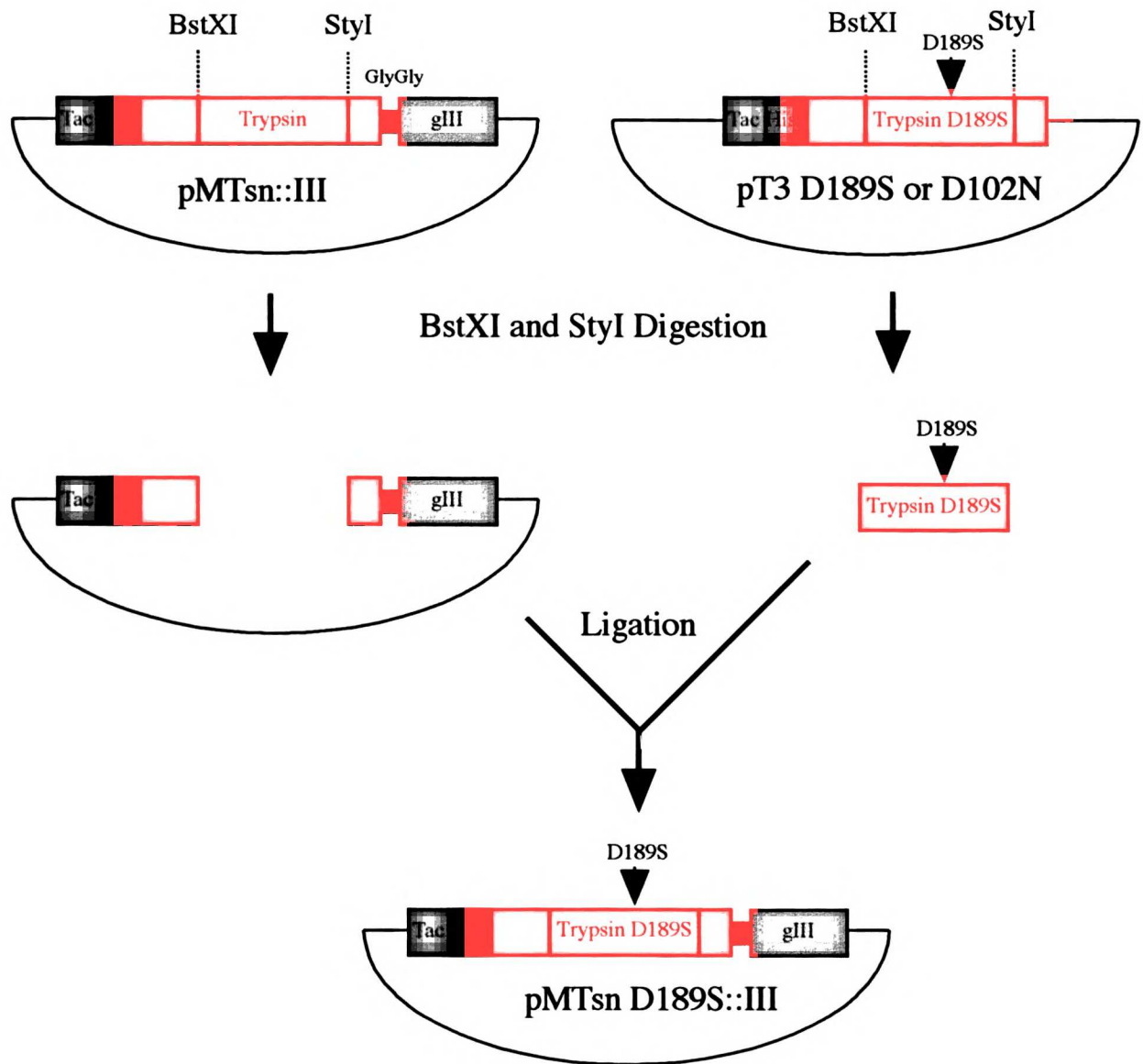


Table 3.1 Kinetic Characterization of Trypsin Variant on Phage

Trypsin Phage	Rate	Linear Coefficient	Titer (pfu)	Normalized Rate
WT	8.854	1.000	2.7×10^8	32.80
D189S	0.054	0.167	1.2×10^9	0.05
D102N	0.012	0.008	1.6×10^9	0.01

Note: The rate was normalized by assuming a standard phage titer of 1.0×10^9 .

The kinetic rate differences on the synthetic substrates among the three trypsin phage samples reflected the dramatic reduction of catalytic efficiency of trypsin D189S phage and trypsin D102N phage compared to that of wild-type(WT) trypsin phage. Both the linear correlation coefficients and the reaction rates of trypsin D189S phage and trypsin D102N phage indicated that their catalytic machinery have been severely perturbed. The loss of activity is similar to that of previously described trypsin (Craik et al., 1987; Graf et al., 1987). Trypsin D102N phage had a dysfunctional catalytic triad but maintained an intact substrate binding pocket (Sprang *et al.*, 1987). The crystal structural of trypsin D102N showed that this binding pocket had the same conformation as that of WT trypsin. On the other hand, the negative charge at the bottom of the substrate binding pocket was removed in D189S trypsin phage, resulting in the lost of selectivity for positively-charged amino acid side chains (Arg or Lys) at the P1 position of the substrate. The kinetic characterization of D102N and D189S trypsin phage demonstrated that trypsin on phage behaves similarly in function as free trypsin. These two variant phage were subsequently used for the characterization and optimization of *in vitro* selection protocols (Chapter Four).

Design of a Trypsin Phage Library Randomized at Three Positions in the Substrate Binding Pocket of Trypsin

Both primary and tertiary structural information was used to identify the dominant determinants of substrate specificity in the binding pocket of trypsin. The establishment of different substrate specificities in homologous enzymes is the result of long term differentiation. It is naive to presume that very few amino acid changes can easily replace one specificity with another. However, by comparing primary sequences of different trypsin molecules at this region, certain consensus patterns emerge. An alignment of 25 different trypsin sequences of different specificities showed that positions 216 (Gly), 225 (Pro) and 226 (Gly) are invariable in all the trypsins (The trypsin sequences used are: cod trypsin, salmon trypsins [I, Ia, Ib, II and III], plaice trypsin, dogfish trypsin, two frog trypsins, human trypsins [I,II, and III], rat trypsins [cationic, anionic I, anionic II, IV, V a-form, and V b-form], mouse trypsin, bovine trypsins [anionic and cationic], dog trypsins [anionic and cationic], and pig trypsin). A large scale alignment which included many other serine proteases of the chymotrypsin family indicated that each of these three positions could accommodate other amino acids, for example, Val at 216, Tyr and Phe at 225, Thr, Ser, Ala, Glu, Gln, Asp, Val at 226. This conservation of Gly, Pro, Gly in trypsins suggested these residues played fundamental roles in providing the optimal conformation and electrostatic environment at the primary binding pocket of these enzymes.

The importance of positions 216, 225 and 226 was further strengthened by examining the three-dimensional structure of WT trypsin complexed with bovine pancreatic trypsin inhibitor (BPTI) and of rat trypsin D189G/G226D complexed with amyloid β -protein precursor inhibitor protein (APPI) and BPTI (Perona *et al.*, 1993). All three residues occupied key positions of the binding pocket, making close contacts with the side chain of the P1 residue (Figure 3.2). Considering the unusual conformational propensities of proline and glycine, these three residues could play significant roles in maintaining the specific local conformation of the binding pocket surface loops. A less constrained binding pocket would be an important step towards the redesign of a binding pocket with novel

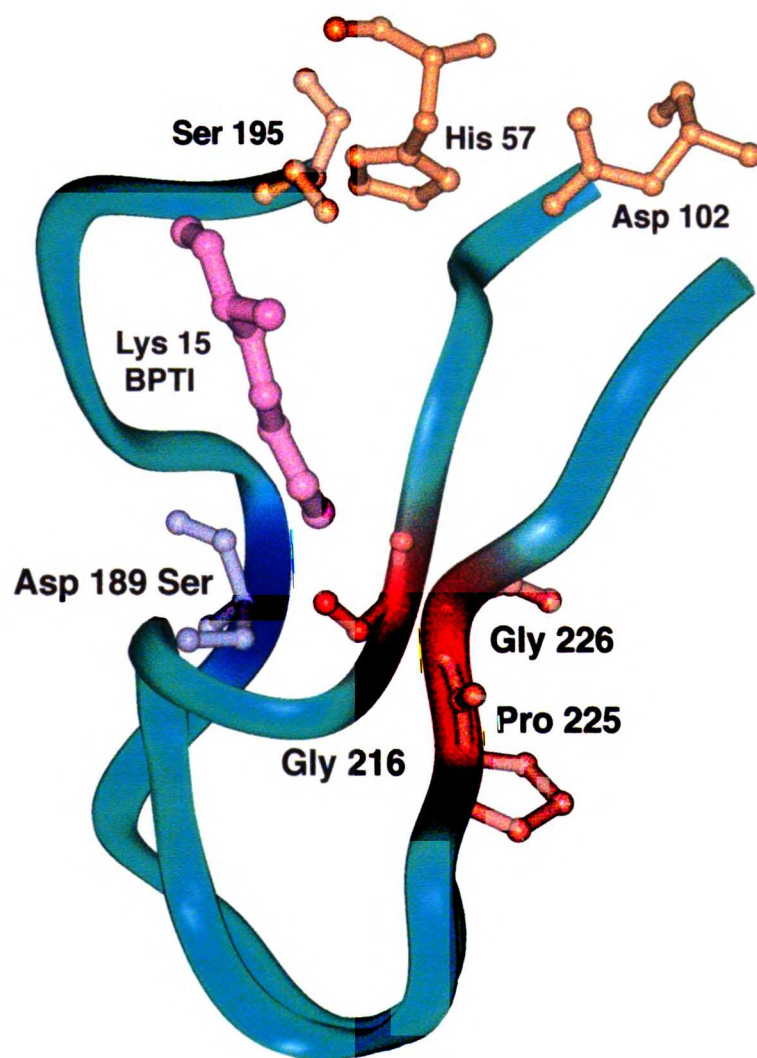
specificities. The structural modeling of the trypsin's binding pocket further enforced the suggestion that residues 216, 225 and 226 were ideal candidates for a random library in this region.

Asp 189 is one of the most crucial elements in substrate recognition by trypsin. The negative charge of Asp 189 provides the favorable electrostatic interactions with the positively-charged side chain fitting into the binding pocket. To expand the specificity of trypsin beyond Arg and Lys, D 189 was replaced with the neutral residue Ser. The combined substitutions at positions 189, 216, 225 and 226 established a new framework through which alternate substrate specificity would be searched.

Figure 3.2 The Primary Substrate Binding Pocket of Trypsin

Trypsin primary substrate binding pocket is shown based on a trypsin-BPTI complex structure. The P1 side chain of BPTI, Lys 15, is shown in magenta. The catalytic triad of trypsin, Ser195, His57, and Asp102 are shown in yellow. Residues Gly216, Pro225, and Gly226 in the binding pocket are shown in red. Asp189 (shown in blue) at the bottom of the binding pocket is replaced by a serine to accommodate non-charged amino acid side chains in the pocket. The trypsin binding pocket library is designed to randomize the three positions 216, 225, and 226 and generate conformational diversity for novel or altered substrate specificities.

Figure 3.2 Primary Substrate Binding Pocket of Trypsin



Construction and Characterization of the Trypsin Phage Library

The trypsin binding pocket library was made by randomizing residues Gly 216, Pro 225, Gly 226 of the trypsin-gIII fusion protein with all twenty amino acids. As described, the fusion protein also carried a D189S substitution. The trypsin phage library vector pMTsnLib::III was constructed from pMTsn D189S::III SacI(+) by the insertion of a library cassette between the SacI and StyI sites of the vector. The process of vector construction is described in detail in Materials and Methods and is shown diagrammatically in Figures 3.3 and 3.4. Pilot studies showed that the double digestion protocol with intermediate GeneClean steps effectively removed all background WT trypsin phage. The final ligation product of the library was electroporated into the high competency *E. coli* strain TOP10 F'. Titration of the transformation mixture indicated that there were 6.7×10^5 individual clones in the library. The completeness of the library was calculated using the equation.

$$N = \ln(1-p)/\ln[1-(1/n)] \quad (3.1)$$

where N is the number of total individual clones in the library, n is the number of possible combinations, and p is the probability that any clone can be found in the library given a library size of N (Sambrook *et al.*, 1989). Since n was equal to 32768 (32^3), the library was calculated to be over 99.99 % ($p > 0.99999$) complete in representing all the possible three amino acid sequence combinations at positions 216, 225 and 226. There was an average of 20 copies of each unique codon combination in the library. After amplification in liquid growth medium, the trypsin phage library was dialyzed extensively in low pH buffer to dissociate ecotin. The titer of the final library stock was $2.3 \times 10^8/\mu\text{l}$.

Figure 3.3 Construction of Trypsin Phage Library Vector

Trypsin phage library vector pMTsn D189S::VIII SacI(+) was constructed by the same method.

Figure 3.3 Construction of Trypsin Phage Library Vector

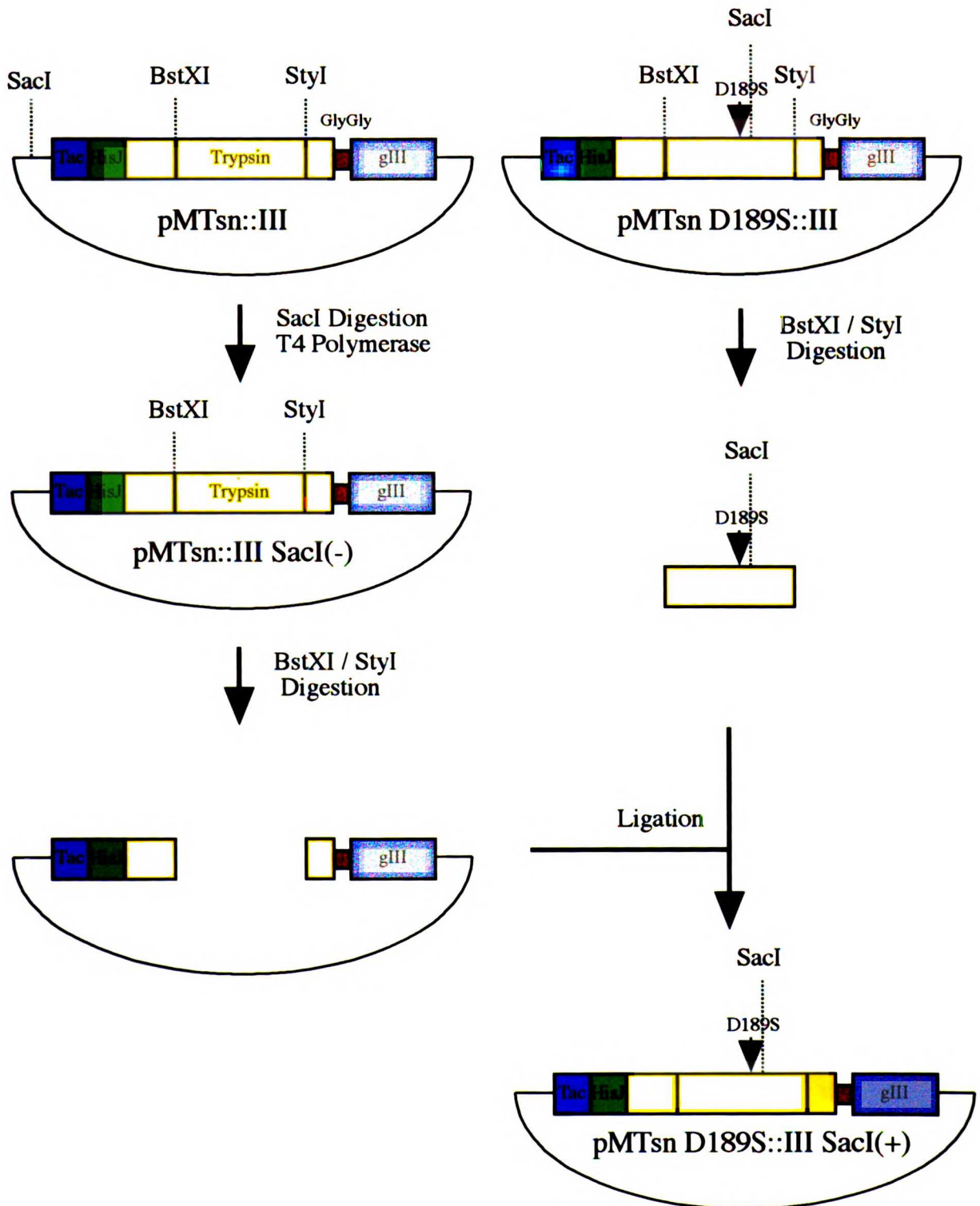
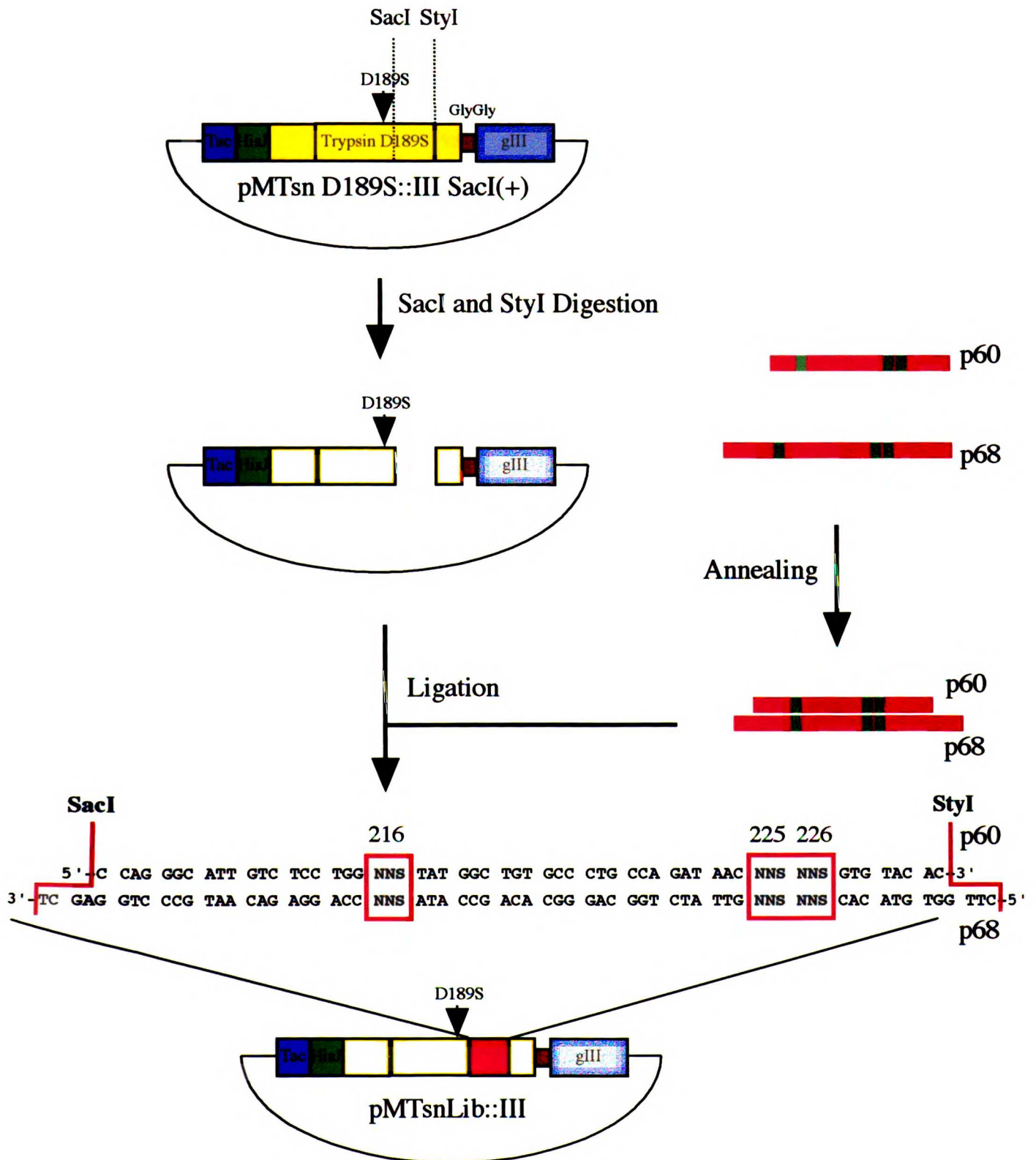


Figure 3.4 Construction of Trypsin Phage Library



The trypsin phage library was characterized by nucleotide sequencing of 12 random clones. The results are shown in Tables 3.2 and 3.3. Of the 12 clones sequenced, none had the wide-type sequence GPG. Fifteen of the 20 amino acids appeared in the sample sequences. The analysis of nucleotide frequencies among this limited set of samples indicated that the distributions of A,C, G,T at the first and second codon position, and of G and C at the third codon position were not significantly skewed or biased.

Table 3.2 Individual Clones in the Trypsin Binding Pocket Phage Library

Positions	216	225	226
	TTC Phe	AAC Asn	GAC Asp
	CTG Leu	GAG Glu	TCG Ser
	CGC Arg	ACC Thr	CAC His
	GTC Val	CGC Arg	TTG Leu
	TAC Tyr	ATC Ile	ACC Thr
	CCG Pro	CCC Pro	ACC Thr
	CCG Pro	CAC His	ACC Thr
	TTG Leu	TCG Ser	ACG Thr
	ACG Thr	GTC Val	CTC Leu
	ACC Thr	TTG Leu	TGC Cys
	TCC Ser	GAG Glu	TGG Trp
	AGC Ser	ACG Thr	AAC Asn

Table 3.3 Sample Nucleotide Distribution of Trypsin Phage Library

Nucleotide	First/Second Position	Third Position
A	20	-
C	23	22
G	9	14
T	20	-
Total	72	36

Ecotin Deleted *E. coli* Strains

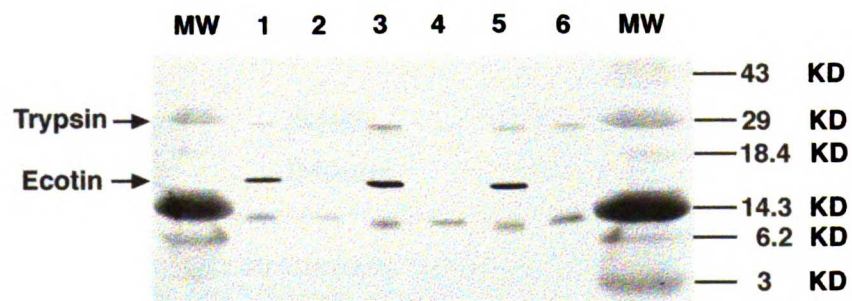
Ecotin is an endogenous serine protease inhibitor found in the periplasm of *E. coli*. (Chung *et al.*, 1983) Ecotin forms a tightly bound non-covalent complex with trypsin phage (Chapter Two). Therefore the majority of the trypsin phage initially purified from bacteria were catalytically inactive. Extensive dialysis at low pH could dissociate ecotin from trypsin phage. However, the dialysis procedure was time consuming and laborious. It also greatly reduced the final yield of active trypsin phage due to nonspecific binding of phage to the surface and membrane of the concentrator. Creating an *E. coli* strain where the ecotin gene had been deleted offered a solution to this problem.

Two ecotin deletion strains were constructed by Dr. Iain Murray (Murray and Craik, unpublished results). These two strains were derived from strains X90 and JM101 and named IM Δ ecoJ and IM Δ ecoJ, respectively. To characterize the level of ecotin and trypsin expression, rat trypsin was expressed in X90 and IM Δ ecoX. The crude periplasmic fractions containing trypsin were prepared. The sample was probed by immunoblotting with anti-ecotin and anti-trypsin polyclonal antibodies (Figure 3.5). The immunoblot clearly showed that there were no detectable levels of ecotin expression in IM Δ ecoX, suggesting that the ecotin deletion was successful. On the other hand, the trypsin expression levels were comparable in these two strains determined by the intensity of the trypsin band.

Figure 3.5 Expression of Trypsin in *E. coli* Strains X90 and IM Δ ecoX

This is an SDS-PAGE gel stained with polyclonal anti-trypsin and anti-ecotin antibodies. Lanes 1, 3, 5 are periplasmic fractions from three individual clones of X90 cells transformed with trypsin expression vector pT3. Lanes 2, 4, 6 are periplasmic fractions from three individual clones of ecotin deletion strain IM Δ ecoX cells transformed with trypsin expression vector pT3. An aliquot of 5 μ l sample was loaded in each lane.

Figure 3.5 Expression of Trypsin in E. coli Strains X90 and IM Δ ecoX



Lanes 1,3,5 X90 Cells

Lanes 2,4,6 IM Δ ecoX Cells

To quantitatively characterize the activity of free trypsin and trypsin phage obtained from the ecotin-deleted strains, trypsin activity was measured by substrate hydrolysis of Z-GPR-pNa monitored at 405 nM. In addition, trypsin phage from the normal and ecotin deletion strains were also prepared without acid dialysis. The activity of trypsin phage was quantified through substrate hydrolysis of Z-GPR-AMC. The results of the kinetic analysis are shown in Table 3.4

Table 3.4 Characterization of Ecotin Deletion Strains IM Δ ecoJ and IM Δ ecoX

Sample	Strain	Activity	Number of Samples
Rat Trypsin	X90	95.3 $\pm$ 9.9	2
Rat Trypsin	IM Δ ecoX	156.4 $\pm$ 29.7	5
Trypsin Phage	JM101	0.09 $\pm$ 0.04	2
Trypsin Phage	IM Δ ecoJ	6.90 $\pm$ 3.00	6

Phage activity was measured by incubating 50 μ l phage sample with 945 μ l Tn100 buffer(50 mM NaCl/50 mM Tris/10 mM CaCl₂, pH 8.0), substrate 5 μ l of Z-GPR-AMC 5 mM for 10 min at 25 ° C.

The data showed that trypsin obtained from the periplasmic extract of IM Δ ecoX was 50% more active than that trypsin obtained from the control strain X90. Furthermore, the trypsin phage obtained from IM Δ ecoJ strain was 75 fold more active than the phage obtained from JM101. In addition, the trypsin phage grown in IM Δ ecoJ was fully active without dialysis in low pH buffer.

Competent cells from both ecotin deletion strains were prepared. Their transformation efficiencies for electroporation and heat shock transformation were measured and were comparable to those of X90 and JM101. Deletion of the ecotin gene did not result in any observable phenotype. However, these cells did grow slower, partially due to the fact that

Discussion

The generation of a phage library randomized at specific sites in the binding pocket of trypsin employed careful considerations and thorough experimentation. A compelling reason to choose the three residues 216, 225 and 226 for the trypsin library was the delicate balance between the size limitation and potential diversity of a phage library. The maximum number of random residues in a phage library is usually less than five since a five residue library would have 3.2×10^6 (20^5) possible amino acid residue combinations. Thus to obtain a reasonably complete five member library with 99% confidence level, at least 1.5×10^8 individual clones are required. However, it would be very difficult to generate such a large number of individual clones through conventional small scale transformation. The transformation efficiencies of most *E. coli* strains are less than 10^8 . Therefore, this intrinsic limitation of transformation efficiency is the major obstacle to achieve a library size over 10^8 . A three residue library is well within the limits of current technologies, and in this case it still would provide substantial diversity to explore the alternative conformations of the trypsin binding pocket. If each amino acid is replaced by all the other 19 amino acids, there would be 8,000 (20^3) different side chain combinations encoded by 32,768 (32^3) distinct nucleotide sequences.

To achieve maximum size and diversity of the trypsin phage library, exhaustive trials were conducted to optimize the key steps in the experimental procedures. Kunkel mutagenesis and library cassette ligation (Wells *et al.*, 1985) were two major approaches used for library construction. The uracil-laden template mutagenesis however did not generate positive results. The two copies of gIII on the same vector may have led to intermolecular or intramolecular recombination events that destroyed the vector. In addition, the large size of the vector may also be responsible for the rapid deletion of a portion of the vector during the growth process. Thus, library cassette ligation became the method of choice.

The demand for a complete library has required a rigorous reexamination of the standard protocols and practices of site-directed mutagenesis. Several critical improvements and modification of the conventional protocols have emerged. For example, to maximize the religation yield of restriction digestion fragments, GeneClean Kits are used to purify the digestion products. Substantial evidence suggests that the agarose gel electrophoresis damages the cohesive end of the DNA fragment and significantly decreases the religation yield of the fragments obtained from restriction digest. The GeneClean kit was also very effective in purifying the ligation product since the glass beads do not bind to DNA fragments less than 500 bp. This extra purification step removed the majority of the unligated duplex library cassette primers and greatly improved the transformation efficiency of the ligation products. It was found that a large excess of short DNA primers could reduce transformation efficiency between five and fifteen fold. TOP10 F' cells were chosen due to their extremely high electroporation transformation efficiency (up to 10^9 per μg of DNA, according to the manufacturer). The ecotin deletion strain IM Δ ecoJ had the potential to minimize the loss of phage yield due to low pH dialysis. Although the transformation efficiency of this strain was still relatively low (10^5 per μg of DNA), it could be used to transform the double strand library DNA stock that had already been amplified through TOP10 F' strains. In summary, the construction of a complete trypsin phage library was built upon the optimization of many technical details.

Materials and Methods

Materials

The *E. coli* strain JM101 was from Stratagene (La Jolla, CA). The high competency *E. coli* strain TOP10 F' strain was from Invitrogen (Carlsbad, CA). The *E. coli* ecotin gene deletion strains IM Δ ecoJ and IM Δ ecoX were derived from JM101 and were the generous gifts of Dr. Iain Murray (Murray and Craik, unpublished results). Enzymes and reagents for the manipulation of DNA were purchased from Promega (Madison, WI) or New England Biolabs (Beverly, MA) and were used following the manufacturer's instructions. The chromogenic substrates N α -benzyloxy-carbonyl-L-glyylproylarginine-7-amino-4-methylcoumarin (Z-GPR-AMC) and N α -benzyloxy-carbonyl-L-glyylproylarginine-paranitroanilide (Z-GPR-pNa) were purchased from Bachem Biosciences, Inc (King of Prussia, PA). Bovine serum albumin (BSA) was purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN). DNA sequencing was performed with the Sequenase 2.0 kit from U.S. Biochemical Corp. (Cleveland, OH). α -³⁵S-dATP was from DuPont NEN (Boston, MA). GeneClean[®] was from Bio101, Inc (La Jolla, CA). Filtron concentrators were from Filtron (Beverly, MA). All other chemicals were of reagent grade or better and were used without further purification.

Activity Assay of Trypsin in Periplasmic Extract

Rat trypsin was expressed in *E. coli* strains X90 and IM Δ ecoX in 3 ml cultures. Three milliliter cultures was induced with a final concentration of 0.5 mM IPTG. Crude periplasmic extract containing trypsin was prepared by osmotic shock. Rat Trypsin activity was measured by incubating 10 μ l of crude periplasmic extract with 188 μ l of Tn100 (50 mM NaCl/50 mM Tris/10 mM CaCl₂, pH 8.0) buffer. An aliquot of 2 μ l of 10 mM Z-GPR-pNa was added to initiate the reaction. Substrate hydrolysis was measured at 405 nm for 2 min using a microtiter plate reader.

Trypsin D189S and D102N Phage Activity Assay

Rat trypsin phage activity assays were described in Chapter Two. Briefly, trypsin phage 100 μ l in 0.1 N HCl/Glycine pH 2.2 buffer was neutralized with 30 μ l of 1 M Tris pH 8.0 and added to 865 μ l of Tn100 buffer. Following a 10-minute equilibration at room temperature, 5 μ l of 5 mM Z-GPR-AMC substrate was added and the rate of substrate hydrolysis measured by monitoring the change of emission at 460 nm during a 5 minute period at 25 °C.

Trypsin Phage Library Vector Construction

Vectors pMTsn::III and pMTsn::VIII both had a single SacI site in the multiple cloning region. There was no SacI site in wild-type trypsin gene. These two vectors were linearized through digestion with SacI, treated with T4 DNA polymerase at 12 °C for 15 min to generate blunt-ends and religated to yield vectors pMTsn::III SacI(-) and pMTsn::VIII SacI(-). Trypsin phage display vectors pMTsnD189S::III and pMTsnD189S::VIII carried two SacI sites, one at the multiple cloning site, the other at residue 209 of the trypsin gene. The BstX1/StyI fragment of pMTsnD189S::III and pMTsnD189S::VIII (which contained the trypsin gene) was inserted into vectors pMTsn::III SacI(-) and pMTsn::VIII SacI(-) that had been digested with BstXI and StyI to yield pMTsn::III SacI(+) and pMTsn::VIII SacI(+), respectively. They both carried a single SacI site at residue 209 of the trypsin gene. This SacI site was a convenient marker of D189S mutation and was used for screening purpose. These two vectors were the templates for the trypsin phage library preparation.

To construct the trypsin phage library at the binding pocket of trypsin, 12.5 μ g of vector pMTsn::III SacI(+) were digested with 10 μ l of SstI (an isozyme of SacI) for 1 hour at 37 °C in a total volume of 200 μ l, and the enzyme was heat inactivated at 65 °C for 15 min. The reaction mixture was purified using the GeneClean and then resuspended in

water. The total sample was digested for 1 hour at 37 °C with 10 µl of StyI in a total volume of 200 µl. Restriction enzymes were inactivated by incubation at 65 °C for 15 min. The double-digested sample was purified using the GeneClean again and resuspended in 50 µl of TE (10 mM Tris, 1 mM EDTA, pH 8.0) buffer. This two-step protocol was developed to accommodate the different optimal buffer conditions of restriction endonucleases SstI and StyI. The SacI isozyme SstI was used instead of SacI since it gave higher religation results.

Two oligo-nucleotide primers, p68 and p60, 68 and 60 nucleotides long respectively, that were used in the library construction were the generous gifts of Dr. Sergio Pichuanes of Chiron Corporation. These two primers had the following sequences (the randomized codons are italicized, where N=A/C/G/T, S=C/G),

p68: 5'-CTT-GGT-GTA-CAC-SNN-SNN-GTT-ATC-TGG-CAG-GGC-ACA-GCC-ATA-SNN-CCA-GGA-GAC-AAT-GCC-CTG-GAG-CT-3'

p60: 5'-C-CAG-GGC-ATT-GTC-TCC-TGG-NNS-TAT-GGC-TGT-GCC-CTG-CCA-GAT-AAC-NNS-NNS-GTG-TAC-AC-3'

One nanomole each of p68 and p60 were annealed slowly overnight to form the duplex library cassette that carried both a 5' and a 3' overhangs of the longer strand. The 5' and 3' sequences of the duplex DNA corresponded to SacI and StyI restriction sites, respectively. The total annealing mix volume was 40 µl. The final concentration of the library cassette was 25 pmol/µl. To generate the library vector pMTsnLib::III, a 0.5 µl aliquot of the annealed library cassette was ligated to 35 µl of the linearized vector in a total volume of 200 µl. The insert to vector ratio was 25 to 1. The ligation reaction proceeded for 20 hours at 16 °C, followed by inactivation of the ligase at 65 °C for 15 min. The reaction mixture was purified using the GeneClean kit. The final sample was resuspended in 20 µl of water and used directly for transformation of *E. coli* strain TOP 10 F. A similar

procedure was used to construct a trypsin phage library on pVIII based on vector pMTsnLib::VIII Sac(+).

Trypsin Phage Library Preparation

Twenty μ l of ligation mix of pMTsnLib::III DNA were transformed into *E. coli* strain TOP10 F' using electroporation. An aliquot of 400 μ l of TOP10 F' cells were used in two separate electroporation reactions. An aliquot of 960 μ l of S.O.C. media (20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 0.186 g KCl, 5 g MgSO₄·7H₂O, 20 mM glucose per liter medium, supplied by Invitrogen) were immediately added to the transformed cells. To determine the size of the final library, one microliter of the transformed cells was added to fresh lawn cells and plated on LB plates. The rest of the library mixture was added to 4 ml of S.O.C. media and incubated at 37 °C with gentle shaking for 1 hour. All the cell culture was transferred into 300 ml of 2YT liquid medium (16 g of tryptone, 10 g of yeast extract, and 5 g of NaCl/liter) with 1000 μ l of 25 mg/ml of tetracycline. After 2 hours of growth, A stock solution of 0.5 M IPTG was added to a final concentration of 0.5 mM. The culture was continuously grown for another 6 hours.

The cells were separated by centrifugation. The double stranded DNA stock of the trypsin phage library pMTsnLib::III vector was purified from the cell pellet using Promega Wizard kit. The trypsin phage was harvested from the supernatant using PEG precipitation with one fourth final volume of 20% PEG8000/2.5 M NaCl. The trypsin phage pellet was resuspended in 0.1 N HCl/Glycine pH 2.2 buffer with 1 mg/ml of bovine serum albumin (BSA). The phage sample was dialyzed against a 300 KD membrane for 2 days. The final trypsin phage sample was stored in the HCl/Glycine buffer for subsequent *in vitro* selection experiments.

Chapter Four : In Vitro Selection of Trypsin Phage

ABSTRACT

Various macromolecular or low molecular weight protease inhibitors were tested to establish an effective *in vitro* selection system for the trypsin phage panning experiment. Ligands tested included Biotin-Leupeptin-semicarbazone and biotinylated chloromethylketone for selection and a chromogenic indoxyl substrate for screening. Ecotin coupled to oxirane beads enriched trypsin phage from background M13 phage. The enrichment ratio ranged from 2,040 to 15,700, independent of the identity of the amino acid residues at the P1 position of ecotin. Soybean trypsin inhibitor (STI) coupled to oxirane beads was another effective ligand/matrix combination for the enrichment of trypsin phage. Trypsin D189S phage and trypsin D102N phage have been prepared and independently panned against STI oxirane beads along with wild-type trypsin phage. STI oxirane beads selectively enriched fully active trypsin phage over phage displaying less active trypsin variants.

Introduction

Successful application of a combinatorial approach to protein design requires the isolation of particular individual variants with useful properties from diverse populations of mutagenized proteins. Two prerequisites for effectively applying phage display and subsequent panning to the combinatorial approach are the construction of a diverse and large random phage library and the development of sensitive *in vitro* selection procedures with optimal ligand and solid support. Although there are several different methodologies for phage library construction, such as oligonucleotide mediated mutagenesis (Kunkel mutagenesis) (Kunkel, 1985), library cassette ligation (Wells *et al.*, 1985) or error-prone PCR (Leung *et al.*, 1989), the experimental procedures are based on the same standard recombinant DNA principles. Phage library construction is less constrained by the nature of macromolecules displayed than by the intrinsic limitation of current molecular biology techniques. In contrast, the choice of a proper ligand and solid support combination requires not only a thorough understanding of the biochemical mechanism of the molecular recognition process being studied, but also the deployment of an array of tools to design, synthesize, and attach the ligands of choice to the solid support matrices.

The serine protease trypsin is an ideal system for such a study. Trypsin has been a favorite platform for enzymologists for over half a century (Craik *et al.*, 1987; Neurath, 1985). Whenever a new and powerful technique emerges, trypsin is among the first proteins to be tested with the novel approach. The examples of the techniques that have been used to provide deeper insight into the structure and mechanism of enzymes include amino acid sequencing (Hermodson *et al.*, 1973), X-ray crystallography (Fehlhammer *et al.*, 1977; Kossiakoff *et al.*, 1977), and molecular biology techniques (Craik *et al.*, 1984; Craik *et al.*, 1985). The application of these techniques to trypsin and other serine proteases has resulted in their being perhaps the best studied class of enzymes regarding

enzyme substrate specificity (Perona & Craik, 1995) and catalytic mechanism. Powerful affinity labels, mechanism-based inhibitors, and transition state analogs have been instrumental in understanding the reaction mechanism of the serine protease and developing highly specific inhibitors for this class of enzymes. These reagents provide rich resources to search for ligands for the *in vitro* selection of trypsin phage.

Results

Biopanning of Trypsin Phage

Commercially available trypsin inhibitors were tested in combination with different solid supports for their potential to capture and enrich trypsin phage from a trypsin phage/mixture. A blue/white selection assay was established to evaluate the efficacy of the protocol. A mixture of trypsin phage and M13mp18 control phage was incubated with a ligand coupled to a solid matrix through a non-covalent linkage. After overnight binding, the solid supports (beads or plates) were washed extensively to remove the non-specific binding phage. Low pH elution buffer was used to release the phage that still attached to the ligand. The phage samples in both the input fraction and the eluate were titrated by infecting of JM101 F⁺ cells. Trypsin phage appeared on the X-Gal/IPTG LB plates as clear(white) plaques, while background M18mp18 phage plaques stained blue. By monitoring the ratio of the number of blue plaque versus white plaques in both the input and elution phage population, an enrichment ratio that reflected the discriminating power of the *in vitro* selection protocols could be calculated. Table 4.1 shows a list of matrices and ligands that were tested using this method. None of the ligand and matrix combinations gave a significant enrichment of trypsin phage over the background. In all cases, there was substantial non-specific binding of M13mp18 phage to either the plates or the beads which resulted in low enrichment ratios. The number of trypsin phage captured by the immobilized ligands was small compared to the background level. Control experiments showed that active rat trypsin could be inhibited by these immobilized ligands listed in table 4.1. The inhibitory activity of the ligands were definitely sustained after immobilization. Therefore, the lack of selectivity likely reflected the surface nature of these solid supports under the experimental conditions.

Table 4.1 Ligands and Matrixes Tested in Panning of Trypsin Phage

Matrix	Ligand	MW (Da)	P1 Side chain	Enrichment
Agarose Beads	Bovine Pancreatic Trypsin Inhibitor (BPTI)	6,000	-- Lys --	< 2
	Lima Bean Trypsin Inhibitor	9,000	-- Lys --	< 2
	Soybean Trypsin Inhibitor	20,000	-- Arg --	< 2
Polystyrene Plate	Turkey Ovomucoid	28,000	-- Lys --	< 2
	Chicken Ovomucoid	20,000	-- Arg --	< 2
	Ecotin	16,100	-- Met --	< 2
Acrylic Beads	Aprotinin (APPI)	6,000	-- Lys --	< 2

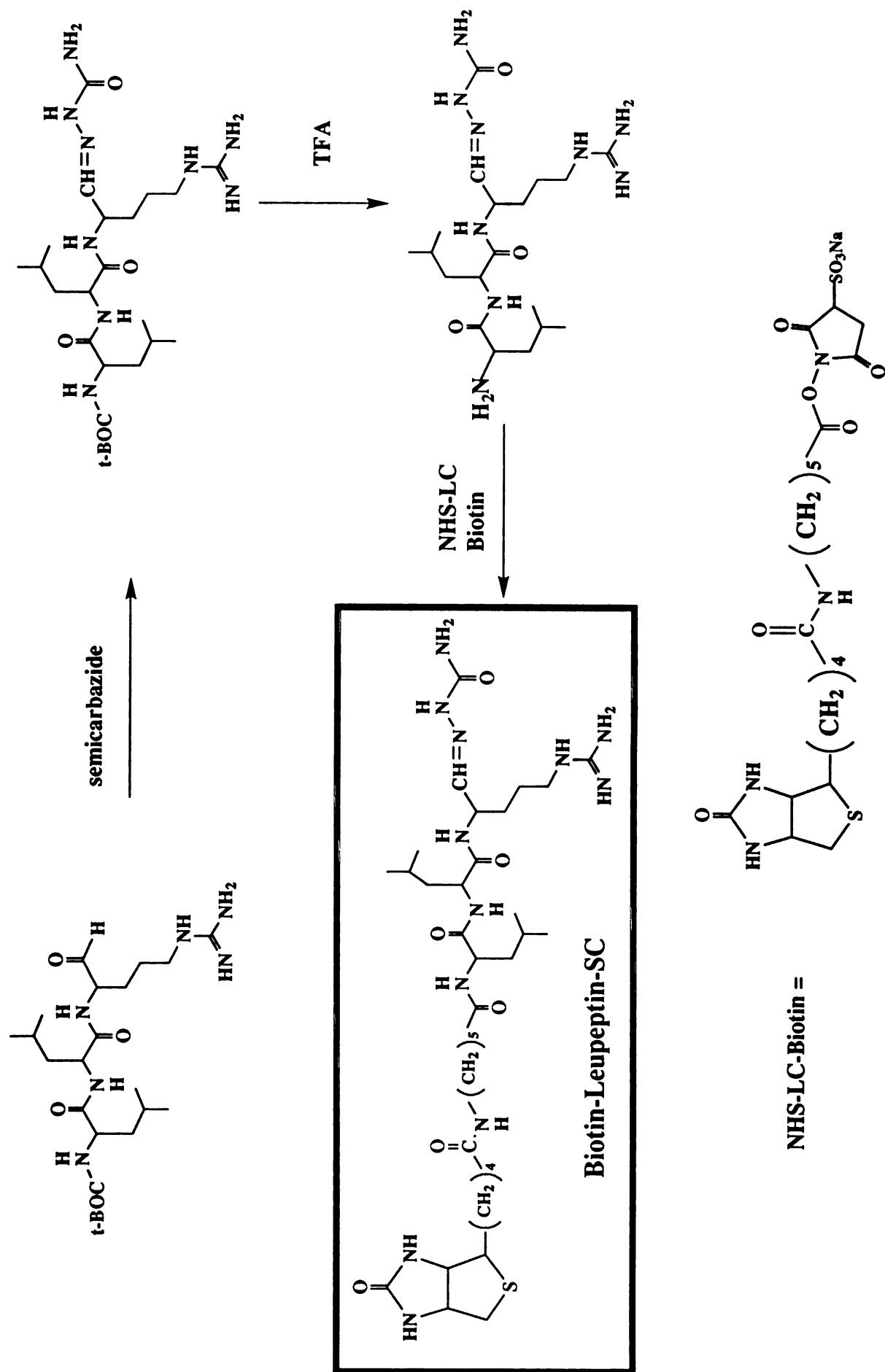
Biotin-Leupeptin Semicarbazone as a Ligand for *in vitro* Selection of Trypsin Phage

Phage display relies on non-covalent binding between the macromolecules displayed and the immobilized ligands. This non-covalent *in vitro* selection criteria is appropriate for studies involving antibody-antigen or inhibitor development. However, for protein engineering of enzymes, this intrinsic limitation could become an obstacle. Variant enzymes that bind tightly to the immobilized ligand will not necessarily exhibit high catalytic efficiencies. One potential solution to this problem is to take advantage of specifically designed transition-state analogs (Shaw *et al.*, 1965) and use these as the immobilized ligands. A transition-state analog forms a covalent bond with the active site residue of the target enzyme, trapping the enzyme in an unproductive tight-binding complex.

To explore the alternate conformation of the binding pocket of trypsin, an aldehyde inhibitor of microbial origin, leupeptin, was chosen for the *in vitro* selection of trypsin phage. Leupeptin is a potent tri-peptide inhibitor (Leu-Leu-Arg) that binds preferentially to active trypsin (Aoyagi *et al.*, 1969). Under normal conditions, over 90% of the aldehyde group is in an inactive hemiacetal form (Schultz *et al.*, 1989). To improve the activity of leupeptin, a semicarbazone group was attached to its C-terminus of leupeptin (Patel *et al.*, 1983). In addition, a biotin group was coupled to the N-terminus of leupeptin to provide the linkage between leupeptin and the solid support via binding to streptavidin. This bi-functional biotin-leupeptin-semicarbazone (Figure 4.1) can thus form very strong non-covalent interactions between biotin and streptavidin and a reversible covalent bond between leupeptin and trypsin. In theory, only active variant trypsin phage can bind to biotin-leupeptin-semicarbazone through nucleophilic attack by active site serine 195 and thus be enriched. In addition, the Arg at the P1 position of leupeptin maximizes the recognition by the negatively-charged wild-type binding pocket of trypsin. Subsequently, this P1 residue can be replaced with other non-polar residues to selectively enrich trypsin phage variants with altered substrate specificities.

Biotin-Leupeptin-Semicarbazone was synthesized by Dr. Cheng-I Wang, a postdoctoral researcher in the laboratory. Diagram of the synthesis procedure is shown in Figure 4.1 (Wang and Craik, unpublished results). The product was characterized for its inhibitory activity against trypsin phage. Different concentrations of biotin-leupeptin-semicarbazone were incubated with purified trypsin phage. The rates of hydrolysis are shown in Figure 4.2. The percent of inhibition is calculated and shown in Table 4.2. Leupeptin's inhibition against trypsin was not that of typical competitive inhibition since a covalent bond was formed between the enzyme and the inhibitor. The kinetic equations of slow tight binding inhibitor was used to estimate the K_i of biotin-leupeptin-semicarbazone

Figure 4.1 Synthesis of the Biotin-Leupeptin Semicarbazone Conjugate



**Figure 4.2 Inhibition of Trypsin Phage Activity
by Biotin-Leupeptin-Semicarbazone**

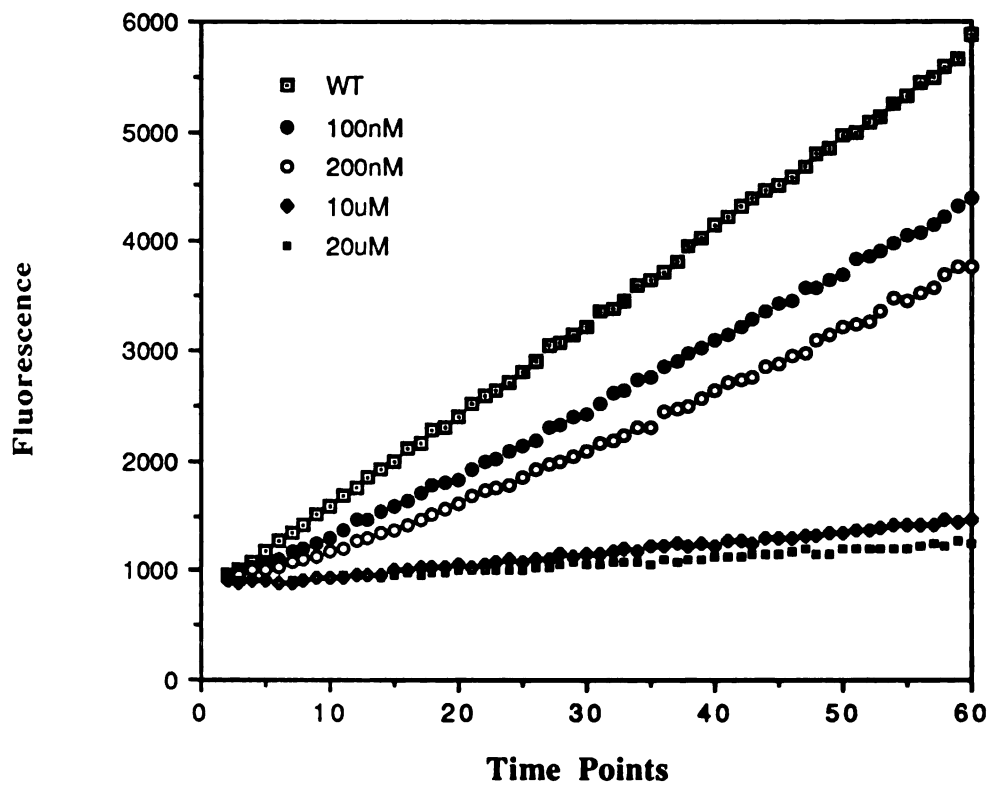


Figure 4.2 Inhibition of Trypsin Phage Activity by Biotin-Leupeptin-Semicarbazone

An aliquot of 10 μ l of trypsin phage was incubated with 5 to 20 μ l of concentrated biotin-leupeptin-semicarbazone stock solution in a total volume of 1000 μ l Tn100 (50 mM NaCl/50 mM Tris/10 mM CaCl₂, pH 8.0) buffer for 10 min. 10 μ l of 10 mM Z-GPR-AMC was added and the rate of substrate hydrolysis was measured at 25 °C for 5 min.

Table 4.2 Inhibition of Trypsin Phage Activity by Biotin-Leupeptin-Semicarbazone

Inhibitor Concentration (μ M)	Reaction Rate	Percent of Inhibition
0	83.67	0
0.1	60.31	27.9
0.2	50.28	39.9
10.0	10.81	87.8
20.0	6.23	92.6

Three solid matrices were used to test the effectiveness of biotin-leupeptin-semicarbazone as a ligand for trypsin phage panning. In addition to the leupeptin conjugate, another class of irreversible serine protease inhibitor, chloromethylketone was also tested with a similar solid support. These ligands and solid supports are listed in Table 4.3. None of these combinations resulted in significant enrichment of trypsin phage over background phage. Control experiments showed that a significant number of background M13 phage bound to the agarose beads, polystyrene plates, or magnetic beads. The potential advantage of having a specific and potent ligand on the surface of these matrices was not realized. Given the fact that almost all the ligands strongly inhibited trypsin phage in solution, the unfavorable physical and biochemical characteristics of the surface of the solid supports became the major hurdle to the development of an effective *in vitro* selection system.

Table 4.3 Leupeptin and Chloromethylketone as Ligands for Trypsin Phage Panning

Matrix	Ligand	MW (Da)	P1 Side chain	Enrichment
Agarose Beads	Leupeptin -semicarbazone	475	- Arg - CH=N-NH-CO-NH ₂	<2
Streptavidin-coated Poly-Styrene Plate	Biotin-Leupeptin- -semicarbazone	781	- Arg - CH=N-NH-CO-NH ₂	<2
	Biotin-Phe-Pro-Arg- Chloromethylketone	831	- Arg -CH ₂ Cl	<2
Streptavidin-Magnetic Beads	Biotin-Leupeptin- -semicarbazone	781	- Arg - CH=N-NH-CO-NH ₂	<2
	Biotin-Arg -Chloromethylketone	434	- Arg --CH ₂ Cl	<2

Design and Synthesis of an Indoxyl Chromogenic Protease Substrate

The development of a synthetic substrate with sufficient sensitivity to distinguish between active and inactive trypsins in an activity screen could be a useful tool to evaluate the activity of trypsin phage. The normal trypsin substrates for kinetic assays have AMC or pNa leaving groups, which form soluble products and diffuse away from the enzyme. A chromogenic protease substrate was designed to take advantage of the unique reactivity of the indoxyl group. X-gal (5-bromo-4-chloro-3-indoxyl- β -D-galactoside) is an essential reagent for the blue/white selection assay by α -complementation (Langley *et al.*, 1975; Langley & Zabin, 1976). Upon cleavage by β -galactosidase, a free indoxyl group is released and immediately oxidized to form a dimeric molecule, indigo blue. Indigo blue is a deep blue colored dye with extremely low solubility in water. Indigo blue is precipitated close to the *E. coli* cells that express active β -galactosidase. The diffusion of indigo blue is

negligible, thus the blue bacteria colonies have very sharp contrast to the background and are easily identified. Based on the same principle, a single amino acid substrate with a free indoxyl group was designed (Figure 4.3). An ester bond links the indoxyl group to the C-terminus of the amino acid. An unsubstituted indoxyl group was used because of lower cost. In addition, previous studies showed that the unsubstituted indoxyl dimerizes and precipitates, although the color of the oxidation product is not as dark as the 5-bromo-4-chloro-substituted indoxyl product. Although an arginine or lysine substrate is desirable for screening trypsin-like activities, a leucine substrate was chosen for the initial synthesis since this residue did not require substantial protection during the synthesis procedure. In addition, crab collagenase, a serine protease with broad specificity, was readily available in the laboratory for leucine-like activity assays (Tsu *et al.*, 1994).

The substrate Suc-leu-indoxyl was synthesized by Dr. Cheng-I Wang, a postdoctoral researcher in the lab. The synthesis procedure is outlined in Figure 4.4. The final product was characterized through assays with collagenase. Upon incubation with collagenase, a dark blue precipitate appeared in solution within 10 to 30 minutes depending on the substrate concentrations. The precipitate had the typical indigo blue color, suggesting that the oxidation and dimerization reaction proceeded as expected. *E. coli* cells expressing active collagenase were grown on a nitrocellulose filter. Upon incubation of the filter with Suc-Leu-indoxyl solution at room temperature for 2 to 6 hours, the colonies on the filter turned blue. Suc-Leu-indoxyl was then tested on LB plates based on similar protocols for X-gal. *E. coli* colonies expressing active collagenase were grown on LB plates pre-soaked with Suc-Leu-indoxyl. Results showed that the colonies remained colorless even after prolonged incubation at 37 °C. The chromogenic substrate had either diffused away or decomposed without oxidation and dimerization. These results precluded the use of indoxyl substrate from LB plate-based screening, and thus greatly reducing its potential use as a

Figure 4.3 Design of a Chromogenic Trypsin Phage Substrate

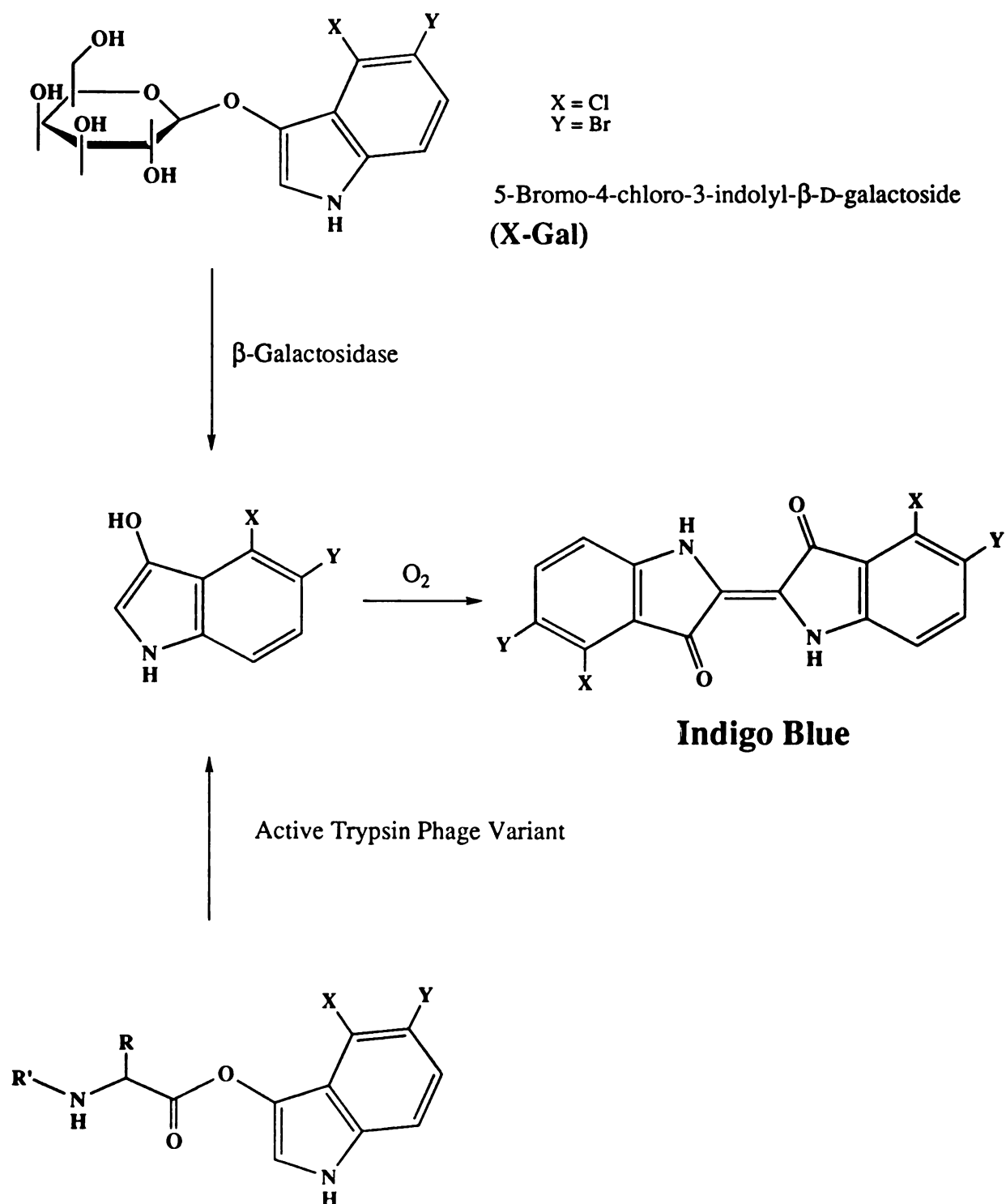
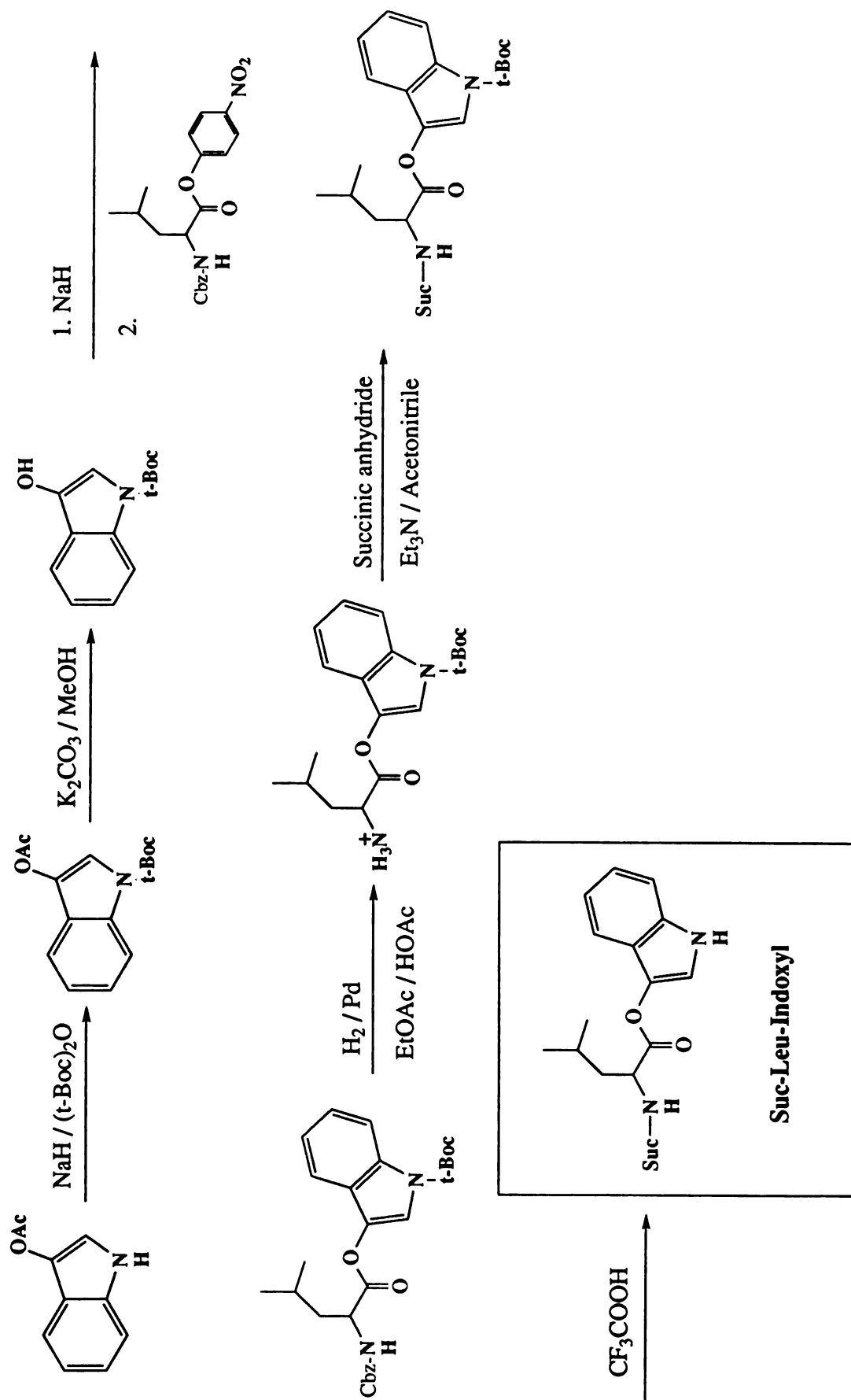


Figure 4.4 Synthesis of the Chromogenic Substrate Suc-Leu-indoxyl



Biotinylated Inhibitors on Oxirane Beads for *in vitro* Selection of Active Trypsin Phage

The results of the Suc-Leu-indoxyl experiment suggested that there was no easily accessible screening tool to characterize the trypsin phage library. The search for an optimal methodology to isolate trypsin phage variants was refocused on *in vitro* selection and panning procedures. Previous experiments had demonstrated that the non-specific binding interactions at the surface of the solid-support were the main obstacle to a highly specific protocol. Non-porous oxirane beads were identified as a solid matrix with extremely low level of non-specific binding activity (Schmitt *et al.*, 1990). This bead was chemically inert and stable under a wide range of conditions. The non-porous nature of the polymer eliminated the majority of the hydrophobic binding sites on the surface of the beads. Protein ligands of interest were covalently coupled to the beads through nucleophilic attack of the oxirane group by the N-terminal or side chain amines (Fleminger *et al.*, 1990). The excess unreacted oxirane groups were blocked by incubation with a strong base such as ethanolamine. The oxirane beads were used for the direct coupling of macromolecular inhibitors (described in later sections) or the indirect coupling of biotinylated small molecular serine protease inhibitors through streptavidin.

Streptavidin was covalently coupled to oxirane beads as shown by ELISA with a biotinylated-horseradish peroxidase assay(data not shown). Three biotinylated compounds were coupled to the streptavidin oxirane beads. These were:

- | | |
|-------------|---|
| Compound B: | Biotin-Arg-Gln-Arg-CH ₂ Cl from Biosyn Diagnostics |
| Compound E: | Biotin-Ile-Glu-Gly-Arg-CH ₂ Cl from Enzyme Systems Products |
| Compound N: | 7-[6-[[6-(Biotinylamino)caproyl]amino]caproyl]amino-4-chloro-3-methoxyisocoumarin from Neurex |

A four step procedure was developed to test the streptavidin oxirane beads carrying the biotinylated trypsin inhibitors (Table 4.4). First, the three inhibitors were first assayed for inhibition of rat trypsin in solution. Compound B and E achieved close to 100 % inhibition within 20 min at concentrations of 186 μ M (B) and 137 μ M (E). Compound N, at 100 μ M, reached 58 % inhibition after 50 min and 87 % inhibition after 8.5 hours. Then, the streptavidin oxirane beads coupled with the three inhibitors were then incubated with free rat trypsin for 100 minutes. The beads were removed from the solution. The remaining rat trypsin activity was measured by fluorescence assay using substrate Z-GPR-AMC and compared with that of the control reaction. Compounds B and E captured nearly all the rat trypsin activity, while compound N captured 26% of the trypsin activity. The third experiment measured the inhibition of the three biotinylated inhibitors against trypsin phage MTsn::III. Compound B and E showed 100 % inhibition, while compound N showed only 36%. In the final test, trypsin phage was captured by immobilized inhibitors over a period of 5 hours. Compound B captured 93% of the phage activity, and compound E 62%. Compound N, however, failed to capture any trypsin phage activity. The trypsin inhibition and capture assays were repeated using MTsn::VIII phage. The results were very similar to the previous results. Compound B, Biotin-Arg-Gln-Arg-chloromethylketone, was the best of the three for *in vitro* selection of trypsin phage.

Table 4.4 Development of Biotinylated Inhibitors as Ligands for Trypsin Phage Panning

	Compound B	Compound E	Compound N
Inhibition of Rat Trypsin by Free Inhibitors (%)	100.0	99.7	57.7
Capture of Rat Trypsin by Immobilized Inhibitors (%)	99.8	99.0	26.1
Inhibition of MTsn::III Phage by Free Inhibitors (%)	100.0	100.0	36.1
Capture of MTsn::III Phage by Immobilized Inhibitors (%)	93.1	61.7	< 0.1
Inhibition of MTsn::VIII Phage by Free Inhibitors (%)	100.0	100.0	< 0.1
Capture of MTsn::VIII Phage by Immobilized Inhibitors (%)	94.3	90.0	< 0.1

Note: Compound B: Biotin-Arg-Gln-Arg-CH₂Cl from Biosyn Diagnostics
Compound E: Biotin-Ile-Glu-Gly-Arg-CH₂Cl from Enzyme Systems Products
Compound N: 7-[6-[[6-(Biotinylamino)caproyl]amino]caproyl]amino-4-chloro-3-methoxyisocoumarin from Neurex

Chloromethylketone inhibitors react with active trypsins and form an irreversible bond with the enzyme (Shaw et al., 1965). The covalent nature of this interaction required another approach to release the phage from the solid support for iterative amplification. Trypsin phage captured by Biotin-RQR-chloromethylketone oxirane beads were incubated with fresh lawn cells. The covalently linked phage however failed to infect the *E. coli* cells. A Polymerase-Chain-Reaction (PCR) based method was developed to amplify the DNA fragment that included the trypsin gene within the phage genome. Phage bound to the oxirane beads were washed extensively in PBST and low pH buffer to remove non-specifically bound phage. The beads were extracted with phenol to release the phage DNA. The DNA sample was amplified through the PCR reaction using primers BstX1 and StyI. These two restriction sites are located at the N and C terminal of the trypsin gene

respectively, covering over 80% of the trypsin gene, especially the region encoding the substrate binding pocket. The fragment containing trypsin gene was religated into the trypsin phage display vector through BstX1/StyI sites and transformed into *E. coli*. Although time and labor consuming, this amplification procedure allowed the effectiveness of the experimental system to be assessed.

An advantage of using chloromethylketone for *in vitro* selection was that the selection might be based on the activity of the protease. A direct competition experiment was designed to evaluate this hypothesis. A mixture of wild-type trypsin phage and trypsin D189S phage or trypsin D102N phage was incubated with Biotin-RQR-chloromethylketone on oxirane beads. The population of captured phage was amplified through the PCR. Random clones from both the input and output phage population were sequenced to determine the ratio of WT and active -site variant phage. The enrichment ratio of this panning procedure is calculated and is shown on Table 4.5. Biotinylated chloromethylketone on oxirane beads offered modest enrichment of fully active WT trypsin phage over trypsin D102N or D189S phage. The discrimination against trypsin D102N phage was more significant, probably reflecting the greater activity difference between WT and D102N trypsin phage. The results of this competitive panning experiment confirmed the hypothesis that the *in vitro* selection was activity-based.

Table 4.5 Enrichment Ratio of Trypsin WT Phage over Trypsin D189S or D102N Phage against Biotin-RQR-CH₂Cl on Oxirane Beads

	Input Phage Number of Clones characterized	Output Phage Number of Clones characterized	Enrichment
Trypsin WT	2	6	
Trypsin D189S	8	6	
WT/D189S	0.25	1	4
Trypsin WT	1	6	
Trypsin D102N	9	6	
WT/D102N	0.11	1	9

Enrichment of Trypsin Phage over Background by Ecotin

The oxirane beads offered a versatile platform to immobilize different inhibitor ligands. Ecotin, a macromolecular serine protease inhibitor in the periplasm of *E. coli* (Chung *et al.*, 1983; McGrath *et al.*, 1991) was coupled to oxirane beads. Ecotin oxirane beads were an alternative to small molecule inhibitors such as chloromethylketones and leupeptin. Based on a superior solid support, the focus of the project was redirected towards the optimization of the inhibitor ligand, ecotin. Ecotin M84R was an active site variant in which the nonpolar Met at position 84 was replaced with a favorable Arg for trypsin binding (Chapter Five). Ecotin M84R was coupled to oxirane beads. The ecotin M84R beads were tested with a series of different concentrations of WT and D102N trypsin phage mixed with a large excess of background M13 phage. The phage mixtures were incubated with ecotin M84R oxirane beads overnight at 4 °C with gentle shaking and rotation. The beads were extensively washed and then eluted with low pH buffer. The ratio of trypsin phage vs.

M13 phage was characterized through blue/white selection assay. The enrichment ratio of panning was calculated and is shown in Figures 4.5 and 4.6.

Panning against ecotin M84R oxirane beads significantly enriched trypsin phage. WT trypsin phage was enriched up to 15,700 fold in a single round even when the ratio of WT trypsin to background was 7.6×10^{-6} in the initial phage population. In fact, enrichment was observed in the late fractions of washing as well as in the second elution. Nevertheless, the most significant enrichment was achieved in the first elution. The enrichment of D102N trypsin phage over background was similar to that of the WT trypsin, implying that the catalytic activity of the trypsin phage was not essential for the *in vitro* selection. In addition, this experiment showed that enrichment could be accomplished under a wide range of concentrations of trypsins phage. The protocol was sensitive enough to select one target phage from over a million background phage.

Figure 4.5 Enrichment Ratio of Trypsin WT Phage over M13 Phage

Three different mixtures of trypsin WT phage and M13 phage were panned against ecotin M84R coupled to oxirane beads. The relative abundance of trypsin phage in three mixtures are 6.8×10^{-2} , 7.6×10^{-4} , and 7.6×10^{-6} , respectively. The fractions of the last two washes and the first, second elutions were titrated using blue/white selection assay. The enrichment ratio was normalized using the trypsin phage/M13 phage ratio of input as the standard.

Figure 4.6 Enrichment Ratio of Trypsin D102N Phage over M13 Phage

The experimental procedure was the same as described for Figure 4.5, except trypsin D102N phage was used instead of the fully active trypsin WT phage in the mixture.

Figure 4.5 Enrichment Ratio of Trypsin WT Phage over M13 Phage

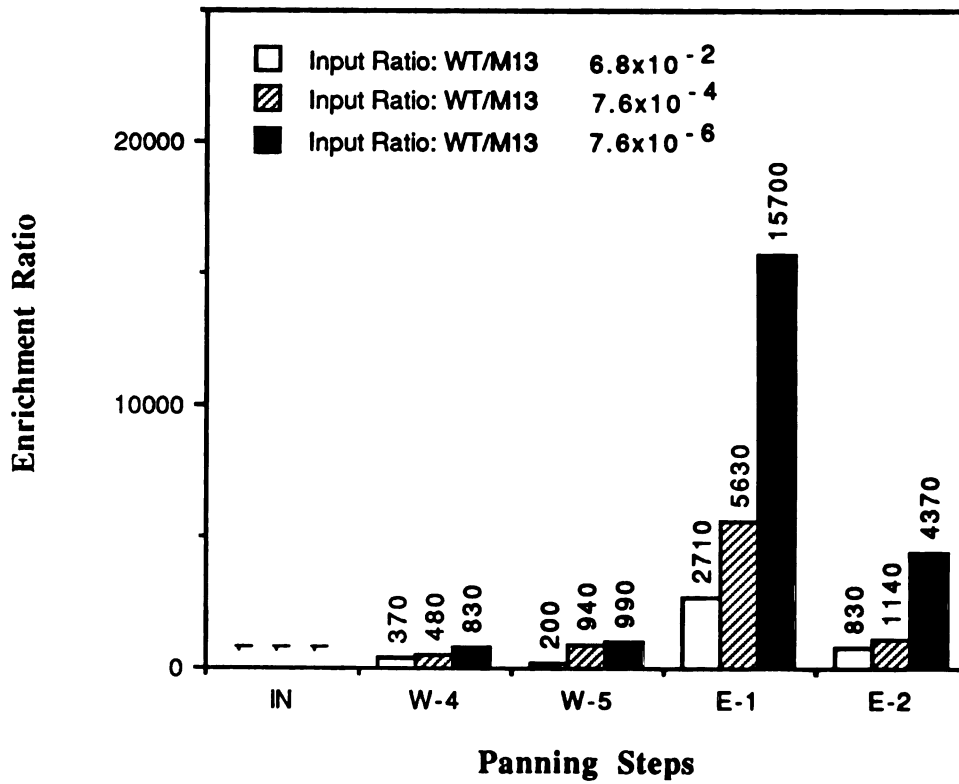
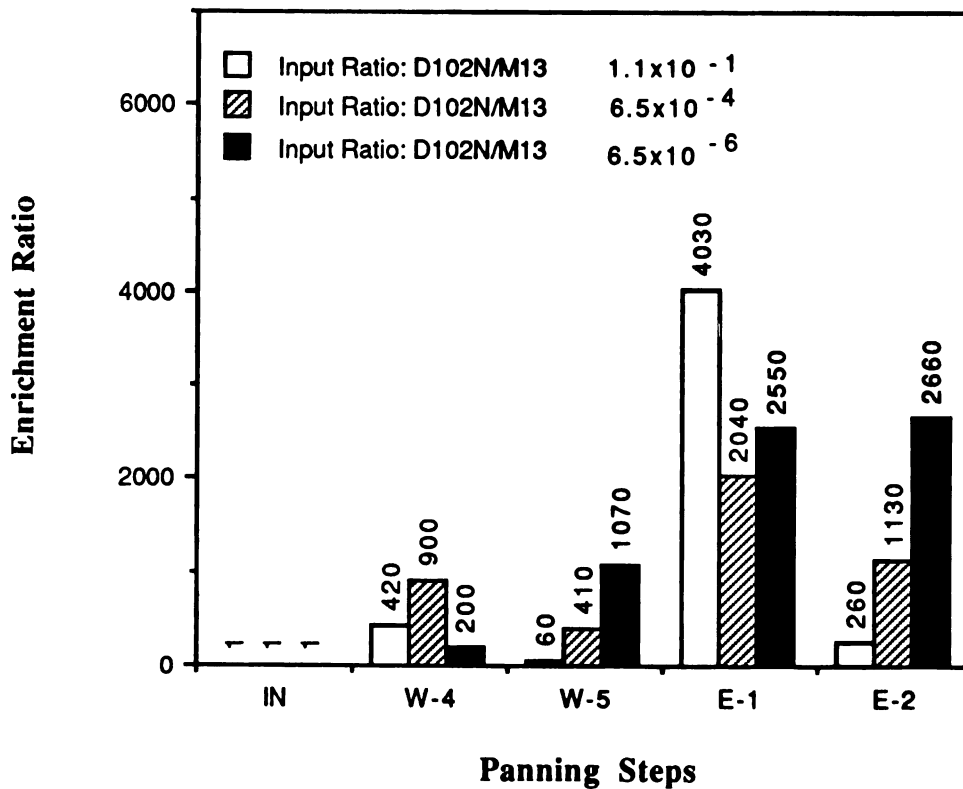


Figure 4.6 Enrichment Ratio of Trypsin D102N Phage over M13 Phage



Ecotin is a potent inhibitor of a large number of chymotrypsin-like enzymes with distinct substrate specificities, such as trypsin, chymotrypsin and elastase. The crystal structure of an ecotin-rat trypsin complex revealed a tetramer formed between the protease and the inhibitor (McGrath *et al.*, 1994). The dimerization of two ecotin monomers generates two unique secondary binding sites providing additional recognition and binding towards trypsin-like enzymes. Ecotin's secondary binding site was the focus of systematic investigation using site-specific and region-specific mutagenesis tools. These experiments are described in detail in Chapter Five.

Four different ecotin oxirane beads were prepared from ecotin WT and three ecotin variants, M84R, WT Δ , and M84R Δ . Both WT Δ and M84R Δ had a ten amino acid truncation at the C-terminal dimer interface of ecotin. The truncation significantly reduced the interaction between ecotin and rat trypsin (Chapter Five). The WT ecotin and the three variants allowed the impact of ecotin's primary and secondary binding sites on trypsin phage panning to be investigated in detail. Trypsin WT phage or trypsin D189S phage were mixed with M13mp18 phage and panned against the ecotin variants on oxirane beads to determine the selectivity of the potential of immobilized ecotin. The enrichment ratios of four ecotin oxirane beads against the WT and variant form of trypsin phage are presented in Table 4.6

Table 4.6 Enrichment Ratio of Trypsin WT and D189S Phage against Ecotin Variants on Oxirane Beads

Ligands	Trypsin WT/M13 Phage Mixture	Trypsin D189S /M13 Phage Mixture
Ecotin WT	392	181
Ecotin M84R	3531	3358
Ecotin Δ	18	21
Ecotin M84R Δ	41	13

Ecotin M84R oxirane gave more than 3,000 fold enrichment over background in a single round of panning for both trypsin WT and trypsin D189S phage. Ecotin WT beads also provided significant enrichment for the two trypsin phage. The two ecotin truncation variants were not as effective as the full length ecotins in capturing and enriching trypsin phage, confirming that the C-terminal dimer interface was critical for rat trypsin binding (Chapter Five). Since the enrichment ratios of trypsin WT and trypsin D189S phage were very similar using four different ecotin beads, the negative charge at the bottom of trypsin's binding pocket was certainly not critical for binding of trypsin phage to ecotin oxirane beads. In fact, mutagenesis at the P1 position of ecotin suggested that this residue did not play a dominant and independent role in determining the inhibitory specificity of ecotin (Chapter Five). The same conclusion was reached through the trypsin phage panning experiment. Although ecotin oxirane beads gave an unprecedented high enrichment ratio for trypsin phage, the broad inhibitory specificity of ecotin restricted its use in remodeling and redesigning the primary substrate binding pocket of trypsin. However, it can be used for capturing active trypsin variants redesigned for other purposes.

Selective Enrichment of Active Trypsin Phage by Soybean Trypsin Inhibitor (STI) and Bovine Pancreatic Trypsin Inhibitor (BPTI)

Ecotin oxirane beads gave 2,040 to 15,700 fold enrichment of trypsin phage over background. This was a very significant improvement compared with the results of the previous enrichment experiments with low molecular weight protease inhibitors. The advantage of a macromolecular inhibitor on oxirane beads was further exploited by coupling two other serine protease inhibitors to oxirane beads. Both soybean trypsin

inhibitor (STI) (Sweet *et al.*, 1974) and bovine pancreatic trypsin inhibitor (BPTI) are well-characterized classical competitive inhibitors of trypsin. Their potency of inhibition ranges from picomolar (BPTI) (Perona *et al.*, 1993b) to nanomolar (STI) (Sweet *et al.*, 1974). Previous mutagenesis experiments have shown that the P1 active residue Lys of BPTI and Arg of STI are the most significant determinant of inhibitory specificity. Replacing these positively-charged residues with uncharged residues greatly reduced the potency of these inhibitors towards trypsin. Therefore, wild-type BPTI and STI were ideal ligands for the panning of trypsin phage. P1 variants of these inhibitors could become the ligand of choice to select variants trypsins with altered substrate specificities.

BPTI and STI were covalently linked to oxirane beads. A trypsin capture assay indicated that both ligands captured the majority of the trypsin activity in solution and released them upon the addition of low pH buffer (Table 4.7). The dissociation of trypsin from these two inhibitors were different. BPTI oxirane beads slowly released the trypsin activity in three elution fractions. However, over 99% of the trypsin activity was released from soybean trypsin inhibitor beads in the first elution. This experiment showed that the STI bead was a better ligand matrix because active trypsin could be readily dissociated from the solid support.

Table 4.7 Capturing Trypsin Activity by BPTI and STI Oxirane Beads (Unit: mO.D./min)

Ligands Trypsin	BPTI Oxirane Beads		STI Oxirane Beads	
	Bovine	Rat	Bovine	Rat
Input	41.40	19.31	41.40	19.31
Supernatant	0.05	0.17	0.13	0.62
Last Wash	0.02	0.02	0.11	0.44
Elution One	4.03	5.83	29.51	18.24
Elution Two	1.29	1.83	0.04	0.09
Elution Three	1.50	3.31	0.08	0.00

To compare the efficacy of STI and BPTI oxirane beads on trypsin phage panning, trypsin WT, D189S, or D102N phage were mixed with M13mp18 phage and the mixture was panned against these two oxirane beads. Table 4.8 showed the enrichment ratio of the panning experiment. STI oxirane beads were very selective towards active trypsin phage, while giving no enrichment for the binding pocket variant trypsin D189S phage. BPTI was less discriminating in the panning. Although it enriched trypsin phage more than either trypsin D102N phage or trypsin D189S phage, the difference between the enrichment ratio of WT and trypsin D189S phage was very small. This indirect comparison experiment supported the idea that STI was a better ligand for the *in vitro* selection experiment.

Table 4.8 Enrichment Ratio of Trypsin WT, D102N and D189S Phage against STI and BPTI

Ligands	Trypsin WT/M13 Phage Mixture	Trypsin D102N/M13 Phage Mixture	Trypsin D189S/M13 Phage Mixture
STI	40	6	1
BPTI	53	31	11

A direct competition experiment was designed to compare STI and BPTI's selectivity between trypsin WT and trypsin D189S phage. A 1:1 ratio of trypsin WT and trypsin D189S phage were mixed and panned with STI and BPTI beads. The input and output phage populations were characterized for the ratio of WT vs. D189S phage. A unique *SacI* site in the genome of trypsin D189S phage allowed the ratio to be easily obtained through restriction digest analysis of random individual clones in the input and output sample. The relative enrichment of WT trypsin over trypsin D189S phage is calculated and presented in Table 4.9. This direct competition experiment confirmed the early observation that STI coupled to oxirane beads was more effective in binding to the active trypsin phage than

BPTI coupled to oxirane beads. Of the sixteen phage clones dissociated from STI beads after elution, none was the inactive trypsin D189S phage. Due to the small sample size, the relative enrichment was estimated to be more than 20 fold. This ratio could be much higher. On the contrary, BPTI on oxirane beads failed to distinguish between the active and inactive trypsin phage. The establishment of STI beads as a potent and selective ligand for the *in vitro* selection of active trypsin phage was a major progress for the search for novel trypsin substrate specificities by phage display.

Table 4.9 Enrichment Ratio of Trypsin WT Phage over Trypsin D189S Phage against STI and BPTI

Ligands		Input Phage Number of Clones	Output Phage Number of Clones	Enrichment
STI	Trypsin WT	7	16	
	Trypsin D189S	9	0	
	WT/D189S	0.8	> 16	> 21
BPTI	Trypsin WT	7	9	
	Trypsin D189S	9	7	
	WT/D189S	0.8	1.3	1.7

Discussion

Different approaches for searching and sorting the trypsin phage library were investigated and are presented here. A chromogenic substrate was designed for an activity-based screen, one alternative approach to *in vitro* selection process. Unlike the panning procedure, which only enriches trypsin phage that bind to the immobilized ligands, an activity screen shows both active and inactive clones. A genetic screen has been previously developed to evaluate trypsin variants with altered specificities (Evnin *et al.*, 1990; Perona *et al.*, 1993a). However, the screen required the construction of an auxotroph strain for each different amino acid tested. It was also limited to the naturally occurring amino acids. A synthetic substrate with sufficient sensitivity to distinguish between active and inactive tryptins is a useful tool to evaluate the activity of trypsin phage. Furthermore, the screen can be coupled to the panning process and be used as the final step before individual clones are picked for characterization. The synthesis and characterization of Suc-Leu-indoxyl proved the principles of the screen design. Although this chromogenic substrate could not be applied to the LB plate as expected, it still suggested a valuable approach to characterize the variant protease library. In summary, both selections and screens offer distinct advantages for specific goals. Screens are suitable for the characterization of a limited set of variants. The goal was to understand the fundamental rules that determine whether a specific variant was active or inactive under certain criteria. Selections are more suitable for exploratory research where the main objective is to isolate any variant with a desirable characteristic. Ideally, both approaches can be combined for the macroscopic and microscopic study of a molecular recognition and interaction process.

Among the *in vitro* selection systems studied, none was the ideal solution to the trypsin protein engineering problem. For example, the biotinylated chloromethylketone inhibitor on oxirane beads was very effective in capturing active trypsin phage from solution.

However, the covalent bond between the phage and the solid support demanded an intermediate PCR step to amplify the phage DNA. Both the lengthy experimental procedure and the low relative enrichment ratio suggested that the class of inhibitor was not a preferred system to isolate the desired trypsin phage variants since many rounds of amplification would be required to achieve a significant enrichment of a dominant species. In general, the low molecular weight inhibitors did not bind to trypsin phage as well as macromolecular inhibitors. The length of the linker moiety between the low molecular weight inhibitor and the solid support might be a key element that deserves further investigation. It is possible that most of the low molecular weight inhibitors simply are not long enough to be accessible to the trypsin phage in solution.

Three macromolecular serine protease inhibitors, ecotin, BPTI and STI, all gave significant enrichment in the trypsin phage panning experiments. Immobilized ecotin enriched the trypsin phage as high as 15,700 fold in a single round of panning. However, the networked interactions between ecotin and trypsin made the separation and isolation of any major determinants of molecular recognition virtually impossible. BPTI, on the other hand, bound to trypsin WT and other variants too tightly, failing to discriminate the active trypsin phage from other inactive trypsin phage variants. Soybean trypsin inhibitor presents an intermediate case. The enrichment ratio was above background but not very high. The relative enrichment between WT trypsin and trypsin D189S phage was statistically significant. Based on a 20-fold enrichment ratio in one round, it may take ten or more rounds of panning to obtain a highly clustered and homogeneous final population from the original random library. The time and effort involved in the panning process will be greatly increased. In addition, a recombinant mutagenesis and expression system has to be established for soybean trypsin inhibitor to make P1 variants carrying other amino acid residues if altering trypsin's substrate specificity is the ultimate goal of the project.

Designing and developing an effective and selective *in vitro* selection system for trypsin phage demands a thorough understanding of the mechanism of inhibition and the protein chemistry of low molecular weight and macromolecular serine protease inhibitors. A solid support matrix with a low level of non-specific binding interactions was identified. Soybean trypsin inhibitor immobilized on oxirane beads specifically enriched active trypsin phage over inactive trypsin phage variants, thus providing a useful *in vitro* selection for the characterization of the trypsin phage library.

Materials and Methods

Materials

The *E. coli* strain JM101 F⁺ was from Stratagene (La Jolla, CA). The *E. coli* ecotin gene deletion strain IM Δ ecoJ was derived from JM101 and was the generous gift of Dr. Iain Murray (Murray and Craik, unpublished results). Enzymes and reagents for the manipulation of DNA were purchased from Promega (Madison, WI) or New England Biolabs (Beverly, MA) and were used following the manufacturer's instructions. Bovine trypsin, nonporous oxirane beads (O-9269), soybean trypsin inhibitor (T-9003), and bovine pancreatic trypsin inhibitor (BPTI) were from Sigma (St. Louis, MO). Rat trypsin was expressed in *E. coli* using the expression vector pZ3 and purified as described (Graf *et al.*, 1987). Wild-type and variant forms of ecotin were produced in bacteria from the expression vector pTacTacEcotin following the protocol described in Chapter 5. The chromogenic substrate N α -benzyloxy-carbonyl-L-glyylproylarginine-7-amino-4-methylcoumarin (Z-GPR-AMC) and N α -benzyloxy-carbonyl-L-glyylproylarginine-paranitroaniline (Z-GPR-pNa) were purchased from Bachem Biosciences, Inc (King of Prussia, PA). Biotin-Arg-Gln-Arg-CH₂Cl was from Biosyn Diagnostics (Belfast, Northern Ireland). Biotin-Ile-Glu-Gly-Arg-CH₂Cl was from Enzyme Systems Products (Dublin, CA). 7-[6-[[6-(Biotinylamino) caproyl]amino] caproyl] amino-4-chloro-3-methoxyisocoumarin from Neurex (Menlo Park, CA). Bovine serum albumin (BSA) was purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN). DNA oligonucleotides were synthesized on a Perkin Elmer/Applied Biosystems 391 DNA synthesizer (Foster City, CA) using the phosphoramidite method and reagents from the same company. Synthetic oligonucleotides were purified with NENSORB cartridges from DuPont NEN (Boston, MA). DNA sequencing was performed with a Sequenase 2.0 kit from U.S. Biochemical Corp. (Cleveland, OH). α -³⁵S-dATP was from DuPont NEN (Boston, MA). GeneClean[®] was from Bio101, Inc (La Jolla, CA). Filtron concentrator

were from Filtron (Beverly, MA). LABQUAKE rotator was from . All other chemicals were of reagent grade or better and were used without further purification.

Polystyrene Plate Panning and Blue / White Selection of Trypsin Phage

The bottom of a 35 mm polystyrene dish (Falcon 1008) was coated with a ligand in the presence of 1 M NaHCO₃ (pH 8.6) with light shaking at 4 °C overnight. Excess binding sites were blocked with 5% non-fat, dry milk solution in 0.1 M NaHCO₃ for 2 hours. The blocking solution was discarded, and the dishes were washed six times with PBST solution (137 mM NaCl/2.7 mM KCl/10 mM Na₂HPO₄/1.8 mM K₂HPO₄ pH 7.5, 0.5% Tween-20). A mixture of trypsin MTsn::III phage and M13mp18 phage sample was loaded onto the plates and allowed to incubate while shaking at 4 °C overnight. Dishes were washed nine times with 10 ml of PBST. Bound phage were serially eluted with 1 ml of a 0.1 N HCl/Glycine pH 2.2 solution, and allowed to incubate at room temperature for 1 min. The ratio of trypsin of MTsn::III phage to M13mp18 phage was determined by mixing 150 µl of log phase JM101 F cells with serial dilutions of the eluate. The mixture was incubated at 37 °C for 30 min to allow for infection, then plated with top agar, 75 µl of 2 % X-Gal and 40 µl of 0.5 M IPTG on LB plates. Background M13 phage appeared on the top agar as blue plaques in the presence of X-Gal due to α -complementation. Trypsin phage, on the other hand, formed white/clear plaques because the reading frame of LacI was disrupted by the insertion of trypsin gIII (or gVIII) fusion into the multiple cloning site of pMTsn::III and pMTsn::VIII. The enrichment ratio was calculated as the ratio of the blue/white colonies of the input phage sample vs. that of the eluates.

Immobilization of Protein Ligands on Oxirane Beads

The immobilization protocol was based on the manufacturer's (Sigma) suggested procedure. Briefly, 50 mg of oxirane beads were mixed with 1 mg of ligand protein

dissolved in water. 100 μ l of concentrated reaction buffer, 2M NaH_2PO_4 , pH 7.6, was added to the oxirane bead and protein mixture. Water was added to a final volume of 200 μ l. The final buffer concentration was 1 M NaH_2PO_4 . The reaction mixture was incubated in a siliconized Eppendorf tube at room temperature in a LABQUAKE rotator for 48 hours. The beads were vortexed periodically to assure optimal mixing. After 2 days of coupling reaction, the beads were spun down in a microcentrifuge for 2 minutes. The supernatant was removed and saved to quantify the amount of ligand remaining in solution. The beads were washed twice with 1 ml of water and then resuspended in 500 μ l of 1 M ethanolamine solution to block the uncoupled oxirane groups. The blocking mixture was incubated at room temperature on a LABQUAKE rotator for 24 hours. The beads were spun down to remove the blocking solution and then washed three times with PBST solution. PBST solution 300 μ l was added to the beads to make a 400 μ l slurry of oxirane beads. This slurry was stored at 4 °C before use.

Panning of Trypsin Phage with Oxirane Beads

Trypsin phage and M13 phage were prepared as described in Chapter Two. A 20 μ l aliquot of ligand-coupled oxirane beads (ecotin oxirane beads or STI oxirane beads) was washed with PBST solution. The beads were resuspended in 500 μ l of 5% dry milk in PBST solution and incubated at room temperature for 6 hours to block the non-specific binding site on the beads. The beads were washed six times with PBST then mixed with 200 μ l of the trypsin phage/M13 phage mixture. The phage mixture was incubated with ligand-coupled oxirane beads at 4 °C with rotation. After 16 hours, the beads were washed with 400 μ l of PBST 9 times at room temperature. The total wash time was between 45 minutes and 1 hour. The elution was carried out in 200 μ l of 0.1 N HCl/Glycine pH 2.2 solution for 30 minutes. The beads were spun down to remove the eluate, which was neutralized with 70 μ l of 1 M Tris, pH 8.0 solution. The final sample was characterized through blue/white selection as described in the previously section.

PCR Amplification of Trypsin Phage

A mixture of active and inactive trypsin phage was panned against Biotin-RQR-chloromethylketone coupled to streptavidin-coated non-porous oxirane beads prepared as described above. After overnight incubation, the beads were washed three times with PBST, two times with 0.1 N HCl/Glycine/1 mg/ml BSA, pH 2.2, two times with 0.1 N HCl/Glycine/1 M NaCl/1 mg/ml BSA, pH 2.2, then three more times with PBST. The beads were resuspended in TE, extracted with phenol/chloroform three times. The supernatant was precipitated with 100% ethanol at -20 °C, washed with 70% ethanol, air dried and resuspended in 10 µl of water. A 2 µl aliquot of the sample was PCR amplified using the following conditions: 94 °C 1 min, 50 °C 2 min, 72 °C 3 min for 35 cycles. The two primers used for PCR amplification of the trypsin gene were:

primer BstXI : 5'-GTG TCC CTG AAC TCT GGC TAC CAC TTC TGT GGA G-3'

primer StyI : 5'-CTG AAT CCA GTC CAC ATA GTT GCA GAC CTT GGT G-3'

The PCR product was purified and digested with BstXI and StyI. The trypsin gene fragment was ligated into pMTsn::III vector linearized with the same two enzymes. The ligation product was purified with Geneclean and was transformed into IMΔecoJ competent cells. Individual phage plaques were amplified and characterized by kinetics analysis and restriction digest.

Chapter Five: Ecotin : A Serine Protease Inhibitor with Two Distinct and Interacting Binding Sites

ABSTRACT

Ecotin is a dimeric macromolecular serine protease inhibitor that forms a tetrameric complex with its target protease. The X-ray crystal structure of the ecotin-rat trypsin complex revealed that in addition to a primary binding site there is a unique secondary binding site between the proteins. The interactions between ecotin and target protease with trypsin-like specificity has been systematically dissected to understand the structural basis of ecotin's broad inhibitory specificity and the role of the secondary binding site. Site-directed and region-specific mutagenesis were performed at ecotin's primary site P1 residue, the C-terminal dimer interface, and two surface loops of the secondary binding site.

Substitutions at P1 position had little impact on the inhibition strength of ecotin against rat trypsin, suggesting that the primary site does not play a dominant and independent role in controlling inhibitory specificity. The dimer interface was disrupted by a ten amino acid deletion at the C-terminal of ecotin. The truncation variants retained the characteristics of inhibitors even though the self-association was weakened. The interactions of ecotin's secondary binding site with bovine trypsin, rat trypsin and human urokinase-type plasminogen activator (uPA) were investigated through multiple alanine substitutions in ecotin at the presumed amino acids that formed the contacts between the inhibitor and protease. A comprehensive set of ecotin variants at the primary and secondary sites was generated to probe the molecular recognition between ecotin and its target serine proteases. The important role of ecotin's unique secondary binding site has been firmly established.

Ecotin binds to proteases through the distinct combinations of primary and secondary site surface loops.

Introduction

Serine protease inhibitors are one of the most diverse families of macromolecules that achieve similar biological functions with entirely different scaffolds (Neurath, 1985). They have been studied extensively due to their ubiquitous presence in numerous biological functions. The rich mechanistic and structural database available on the interactions between serine protease inhibitors and their target enzymes make these inhibitors an excellent model system to investigate the fundamental biochemical and biophysical principles of protein-protein recognition. Serine protease inhibitors have been classified into at least twenty sub-families based on amino acid sequence and mechanism of interaction (Laskowski & Kato, 1980). Four common reaction mechanisms have been postulated to describe the different chemical and kinetic pathways of the inhibition.

The substrate-like standard mechanism accounts for the inhibition by a large number of canonical small macromolecular inhibitors with less than 200 amino acids. A classic example of this type of inhibition is that achieved by the binding of Bovine Pancreatic Trypsin Inhibitor (BPTI) to trypsin. In this case, the P1 (Schechter & Berger, 1967) residue of the inhibitor occupies the S1 binding pocket of the protease and the amino acid residues flanking P1 bind to the enzyme in a substrate-like conformation.

Serpins, a family of serine protease inhibitors consisting of plasma proteins of more than 400 amino acids, inhibit their target enzymes through a loop insertion followed by a conformational change (Bode & Huber, 1992; Engh *et al.*, 1995). Upon cleavage of the serpin reactive loop by the protease, the newly formed terminal fragments are incorporated into the β -sheets of the inhibitor to induce its refolding. This structural rearrangement step is followed by the generation of a tightly bound inhibitory complex.

In contrast to the above mentioned mechanisms, the other two serine protease inhibition mechanisms are more dependent on steric occlusion of the enzyme substrate binding pocket by the inhibitor. The non-specific "molecular trap" inhibition by α 2-macroglobulin involves breaking a thiol ester bond and "engulfment" of the target protease. Finally, the inhibition of thrombin by hirudin involves an extended binding site that bridges the active site nucleophile of the enzyme and blocks it from productive substrate binding.

A common structural theme among these reaction mechanisms is that the protease and the inhibitor form a one-to-one complex. The majority of these inhibitors bind to the target protease at a single site. This binding site is usually well-defined structurally and confined to a specific surface of the inhibitor. In the case of substrate-like serine protease inhibitors, an extended surface loop of the inhibitor binds to the active site of the protease with a substrate-like conformation. This so-called "reactive site loop" provides all the binding interactions between the inhibitor and protease. The rest of the inhibitor plays the structural role of maintaining the proper conformation of the reactive site loop. Solution kinetics studies have shown that, in many cases, the amino acid side chain at the P1 position of this reactive site loop is the primary determinant of the inhibitory specificity. For example, BPTI, with a lysine at the P1 position, preferably inhibits proteases with trypsin-like fold and Arg/Lys specificity. Replacement of the P1 Lys in BPTI with large hydrophobic residues changes its specificity to chymotrypsin. Replacement of the Lys with small non-polar residues changes its specificity to elastase.

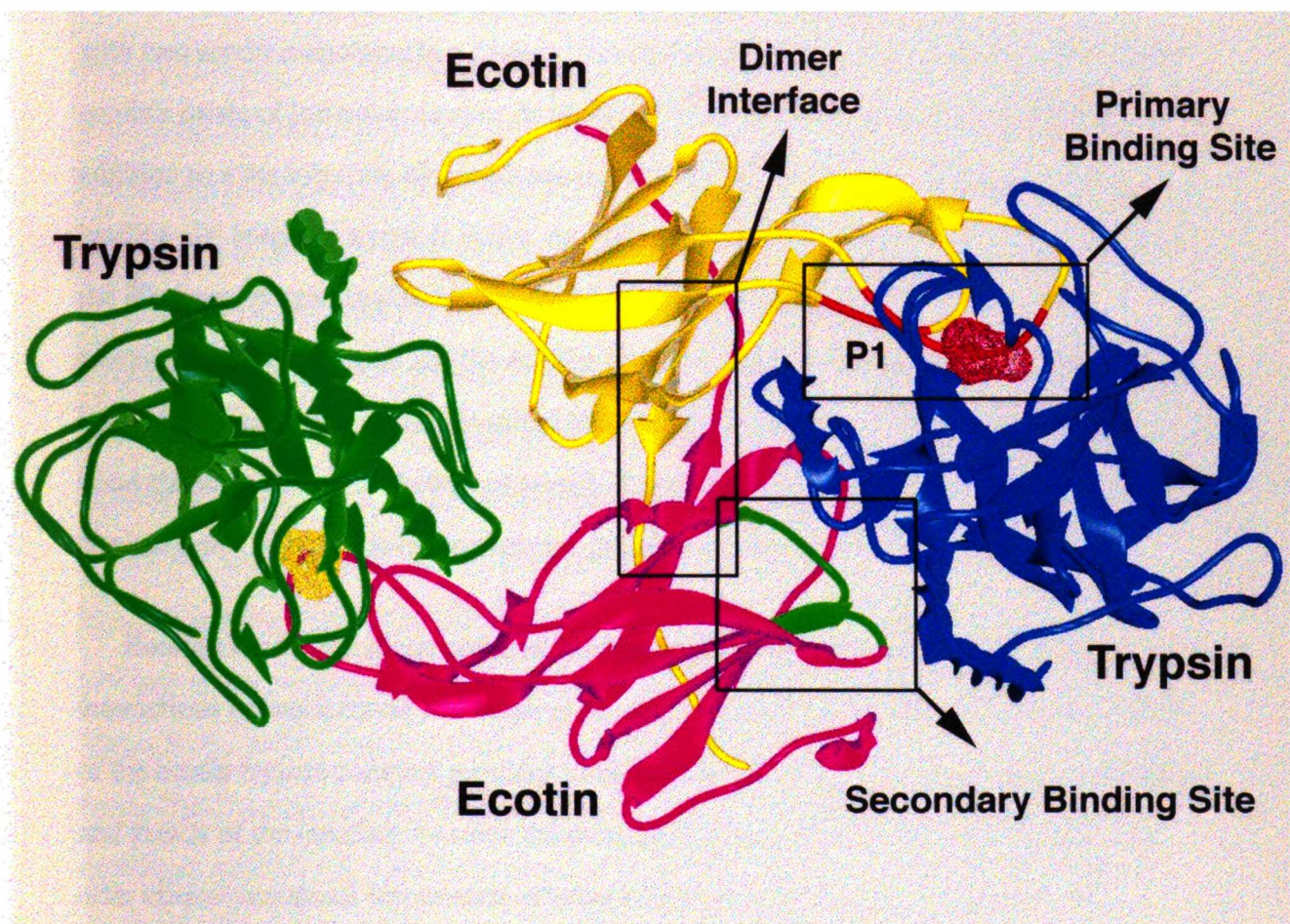
Ecotin, a serine protease inhibitor found in the periplasm of *Escherichia coli*, is a noticeable exception to these observations. Ecotin is a competitive inhibitor that strongly inhibits trypsin, chymotrypsin and elastase and many other serine proteases with comparable potencies (Chung *et al.*, 1983). This well characterized inhibitor has been purified; its cDNA has been cloned and expressed recombinantly in *E. coli*. (Erpel *et al.*,

1992; McGrath *et al.*, 1991a; McGrath *et al.*, 1991b) The 142 amino acid monomeric ecotin, with a P1 Met at residue 84, dimerizes in solution at submicromolar concentrations. The high-resolution crystal structure of ecotin complexed with rat anionic trypsin shows two ecotin monomers form a dimer through interactions at their C-termini (McGrath *et al.*, 1994) (Figure 5.1). The ecotin dimer binds to two trypsin molecules at opposite ends to form a heterotetramer with a two-fold symmetry axis. The crystal structure also reveals a network of interactions between ecotin and trypsin. The protein-protein interaction region between the inhibitor and the protease consists of two distinct areas, each provided by one of the two ecotin molecules. The first area, known as the "primary binding site", involves the reactive site loop of ecotin, i.e. the 80s loop (residues 80-88), and the active site of trypsin. The second area, known as the "secondary binding site", includes two surface loops of ecotin, the 60s loop(residues 66-70) and 100s loop(residues 108-113), and the C-terminal region of the protease (including part of the C-terminal helix and part of the 90s loop of trypsin).

Figure 5.1 Tetrameric Complex of Ecotin-Trypsin

The high-resolution crystal structure of rat anionic trypsin-ecotin complex illustrated A network of interactions between ecotin-ecotin monomers, between ecotin-trypsin at the primary binding site, and between ecotin-trypsin at the secondary binding site.

Figure 5.1 Tetrameric Complex of Ecotin-Trypsin



The heterotetramer formed by two ecotin monomers and two trypsins is unique among the known structures of serine protease and inhibitor complexes. Although *Streptomyces* subtilisin inhibitor (SSI), an inhibitor also following the standard mechanism, forms a tetramer with subtilisin, each subtilisin is bound to one SSI in a "chain-like" configuration (Takeuchi *et al.*, 1991). However, in the case of ecotin, each trypsin makes direct contacts with two ecotin monomers in a "matrix" configuration. From a mechanistic point of view, ecotin's mode of inhibition against its target protease shows novel characteristics in addition to a the substrate-like mechanism. The interaction between ecotin's primary binding site loop and trypsin's active site follows the classic model. The conformation of the 80s loops is superimposable on that of the reactive site loop of BPTI and other canonical macromolecular inhibitors. The inhibitor displays competitive inhibition kinetics. Biochemical characterization identified Met84 as the P1 residue with direct evidence of bond cleavage between Met84 and Met85. However, ecotin's distinct secondary binding site has rarely been observed in protease-inhibitor complexes.

Ecotin's broad specificity and dual binding mode led us to investigate the binding interactions between ecotin and target proteases. Although the three-dimensional structure of the ecotin-trypsin complex provides some insight regarding the mechanism of inhibition and details of the interface between the protease and its inhibitor, this structure does not offer critical functional information relating to the activity of ecotin. Therefore, the contribution of the various interactions of the protease-inhibitor complex must be determined through other means. We have used a combination of site-directed and region-specific mutagenesis to address the relationship between the structure and function of this unique inhibitor and specifically ask how the inhibitor uses two distinct binding sites to recognize its target.

Results

Ecotin Variants Preparation and Characterization

In this study, ecotin variants with amino acid substitutions and deletions at the primary or secondary binding site were generated by site-directed mutagenesis (Kunkel, 1985), expressed to high-levels in *E. coli* and subsequently purified to homogeneity. The final yield ranged from 20 to 100 mg protein per liter of liquid culture. The same expression system and purification protocol using heat treatment and reverse phase HPLC steps were applied to both primary and secondary site ecotin variants. Activity assays were performed at each purification step to monitor for activity loss of the samples. All variants were expressed in *E. coli* at comparable levels and were stable under the purification conditions. Table 5.1 is a summary of the nomenclature of ecotin variants appeared in this study. The SDS-PAGE analysis of purified aliquots of ecotin variants is shown on Figure 5.2 and 5.3. Ecotin WT and six variants at the P1 position and dimer interface (WT, M84R, WTA, M84RΔ, M84K, M84F, M84W) are shown on Figure 5.2; while figure 5.3 shows ecotin WT and seven variants at the secondary site loops (WT, 60A₄, 100A₄, 60A₄+100A₄, 50A₃+80A₅, M84R+60A₄, M84R+100A₄, M84R+60A₄+100A₄). Each variant exhibited single-band homogeneity on SDS-polyacrylamide gels and migrated at the expected molecular weight. All variants cross-reacted with anti-ecotin polyclonal antibodies on immunoblots. Selected variants were analyzed with electrospray mass spectroscopy to confirm their compositions. Ecotin variants 50A₃+80A₅, 60A₄+100A₄, and M84R+60A₄ had molecular weights of 15713.40±2.14, 15636.67±3.22, 15886.95±1.79 respectively.

Table 5.1 Definition of Ecotin Variants

Variant Symbol	Mutations
M84X (X=R, K, F, W)	Single residue substitution at position 84
WTΔ	Deletion of residues 133-142
M84RΔ	Deletion of residues 133-142 combined with Arg substitution at position 84
50A₃	Multiple alanine substitutions at residues 52-54
60A₄	Multiple alanine substitutions at residues 67-70
80A₅	Multiple alanine substitutions at residues 81-85
100A₄	Multiple alanine substitutions at residues 108, 110, 112, 113

Figure 5.2 SDS-PAGE Gel of Ecotin Primary Site and Dimer Interface Variants

Figure 5.3 SDS-PAGE Gel of Ecotin Secondary Site Variants

Ecotin was purified as described in Methods. An aliquot of about 1 ng of HPLC purified ecotin variant sample was loaded in each lane of the SDS-PAGE gels.

**Figure 5.2 SDS-PAGE Gel of Ecotin
Primary Site and Dimer Interface Variants**

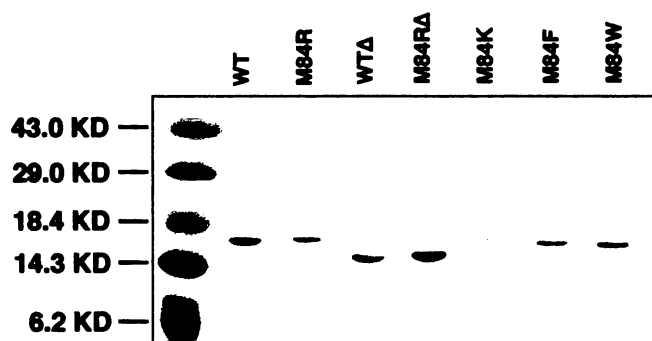
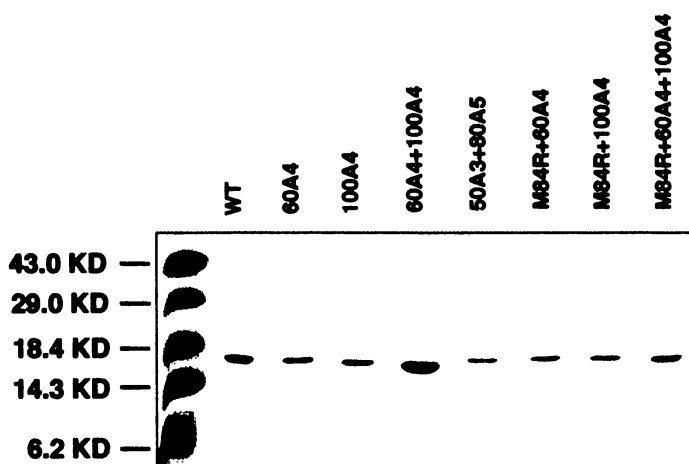


Figure 5.3 SDS-PAGE Gel of Ecotin Secondary Site Variants



The Role of the Primary Binding Site

Since the P1 amino acid side chain is often a major determinant of the inhibitory specificity of many substrate-like serine protease inhibitors, the role of Met 84 of ecotin was explored by substituting this position with Arg, Lys, Phe and Trp. Met is not a common residue at P1 for serine protease inhibitors that target enzymes with Arg or Lys specificity. For classic canonical inhibitors of trypsin, Arg and Lys are preferred at the P1 site, while both Phe and Trp are not well accommodated in the S1 pocket of trypsin. The inhibition constants (K_i) of ecotin P1 substitution variants against rat trypsin were measured and are listed in Table 5.2. Significant changes of the electrostatic and hydrophobic properties of the P1 residue of ecotin had relatively little effect on inhibition of rat trypsin by these variants. Although the favorable substitutions M84R and M84K had lower K_i values compared to that of WT ecotin, the magnitude of the change was only two fold. Similarly, ecotin variants with unfavorable mutations M84F and M84W were still potent inhibitors of rat trypsin with the greatest effect being five fold weaker association between rat trypsin and ecotin M84F compared to WT ecotin.

Table 5.2 K_i of Ecotin P1 Variants at Position 84 against Rat Trypsin (unit: nM)

WT	M84F	M84W	M84K	M84R
0.93	4.46	2.20	0.55	0.45

The Role of the C-terminal Dimer Interface

The C-terminal dimer interface between the two ecotin monomers is a key structural element of the tetrameric complex as it forms one of the three types of protein protein

interfaces in the tetramer. The other two are the primary and secondary interactions with the protease. The dimerization of ecotin not only dramatically increases the contact region between the inhibitor and the protease, but also adapts to maintain the proper orientation of the two ecotins as a molecular "hinge" to permit binding to diverged proteases. The combined surface areas of the primary and secondary sites of ecotin that become buried upon binding to trypsin exceed 2800 Å². This area is substantially greater than most of the other protease-inhibitor interfaces (e.g. BPTI-Trypsin interface is only 1390 Å²). If the hinge region is disrupted or destroyed, the relative positions of the primary and secondary binding sites may change as they derive one each from two ecotins. We investigated the role of the C-terminal dimer interface on ecotin's unique broad specificities by disrupting it.

Based on the crystal structure of the ecotin-trypsin complex, residues 133-142, which consist of about one half of the contact amino acids of the C-terminal arm, were deleted while still maintaining a stub. The ecotin variants with residues 133-142 truncated was denoted ecotin WTΔ. The other variant with residues 133-142 truncated and the P1 Met at position 84 replaced with an Arg was denoted ecotin M84RΔ. These two variants were designed to probe the importance of the dimer interface under the context of "neutral" or favorable P1 interactions. The monomer-dimer dissociation constant, K_d , was measured by change in fluorescence upon titration following the protocol of Seymour et al. (Seymour *et al.*, 1994). The ecotin monomer contains two tryptophans with Trp 130 being located near the dimer interface. The fluorescence of Trp130 is sensitive to the oligomerization state of the molecule. The K_d 's of WT ecotin and several truncation variants are shown in Table 5.3. The K_d of WT ecotin was close to the value of 390±150 nM measured by Seymour et al. using the same methodology (Seymour et al., 1994). However, ecotin WTΔ and M84RΔ had K_d values of only 2.5 and 3.7-fold higher respectively, suggesting that the truncation variants of ecotin still formed dimers, even though their dimer

interactions was weaker, as indicated by the modest increase of the K_d 's. Interestingly, replacement of Met at position 84 with an Arg also increased the K_d slightly relative to WT ecotin (1.7-fold). The subtle change at one surface loop is conveyed to the C-terminal of the same molecule, implying ecotin has remarkable structural flexibility.

Table 5.3 K_d of Ecotin C-terminal Truncation Variants (unit: nM)

WT	220 ± 130
WTΔ	540 ± 210
M84R	380 ± 170
M84RΔ	820 ± 310

To probe the inhibitory activities of the ecotin variants, three homologous serine proteases, bovine cationic trypsin, rat anionic trypsin, and catalytic domain of human urokinase-type plasminogen activator (uPA) were chosen as model targets. These enzymes share the same three dimensional fold and have very similar substrate binding pockets. However, trypsin and uPA have different *in vivo* physiological functions and catalytic efficiencies. Together, they form a sample set that represents the large family of serine proteases recognizing Arg/Lys at the P1 position. The K_i values of the ecotin C-terminal truncation variants against the three serine proteases are shown in Table 5.4. These data are also plotted in Figure 5.4 to illustrate the relative contributions from the P1 residue and the dimer interface. The juxtaposition of the K_i data of the same group of ecotin variants against the three serine proteases provides a clear and direct graphic comparison of the inhibitory activity of the ecotin variants under different circumstances. We used these K_i plots to visualize the distinct patterns of molecular recognition between ecotin variants and their target proteases.

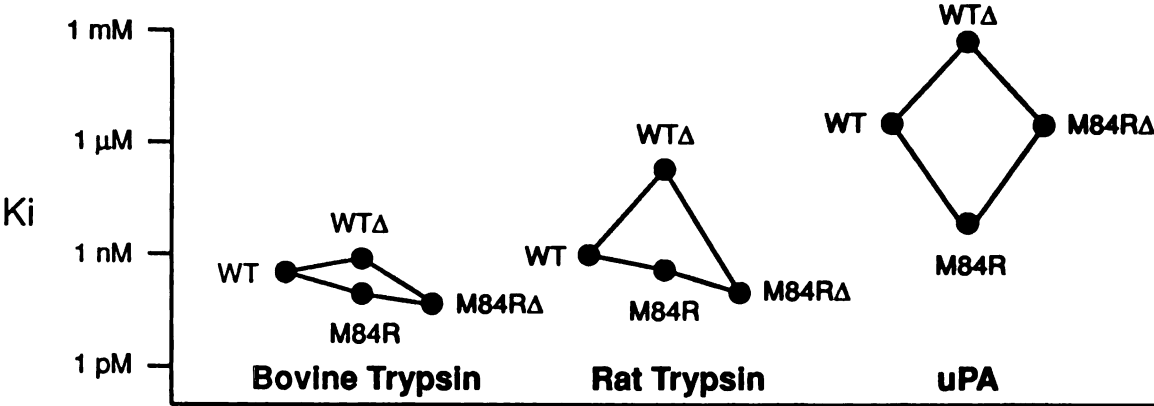
Table 5.4 *K_i* of Ecotin C-terminal Truncation Variants (unit: nM)

	Bovine Trypsin	Rat Trypsin	uPA
WT	0.31	0.93	2800
WTΔ	0.72	181.3	434000
M84R	0.09	0.38	3.6
M84RΔ	0.04	0.09	2700

Figure 5.4 *K_i* Plot of Ecotin Truncation Variants

The *K_i*'s of ecotin truncation variants are plotted in log scale along Y-axis.

Figure 5.4 Ki Data of Ecotin Truncation Variants



One immediate and striking implication from the *K_i* measurement was that deletion of ten amino acids at the C-terminus of ecotin has minimum impact on its structural and functional integrity. This observation is seen in the case of bovine trypsin where the C-terminal truncations in either WT or M84R variant only had small effects on *K_i*, suggesting that the deletion did not destabilize the ecotin molecules. In contrast, the inhibition of rat trypsin by WTΔ was almost two hundred fold weaker, indicating that the truncation could have dramatic consequences. This surprising result also demonstrated that a change distal from the primary binding site of ecotin could transmit through the molecule and affect inhibition. However, once the M84R substitution was introduced and combined with the C-terminal deletion, ecotin M84RΔ not only regained nanomolar potency against rat trypsin, but also became an even better inhibitor of this enzyme. This result suggests that the favorable electrostatic force at P1 position could compensate for the weakened interactions due to the C-terminal truncation. Finally, the *K_i* data of uPA inhibition revealed yet another pattern. In this case, the M84R substitution only partially compensated for the unfavorable interactions created by the C-terminal deletion. However, the *K_i* of the double variant M84RΔ against uPA remained at micromolar levels. In contrast to that of rat trypsin, the favorable electrostatic interaction at the P1 alone was not sufficient for tight binding with uPA. Overall, these results show that the inhibition of rat trypsin and uPA by ecotin are very sensitive to the perturbation of the dimer interface. However, the inhibition of bovine trypsin, an enzyme that has 74% sequence identity with rat trypsin, was insensitive to mutations at the P1 and C-terminal region.

Region-Specific Mutagenesis at the Primary and Secondary Binding Sites

Although investigations of the P1 site and dimer interface had revealed the importance of the secondary site, more direct evidence is necessary to establish its role in protease binding. A structure-based approach was adopted to dissect and isolate the relative contributions from ecotin's two binding sites. By examining the ecotin-trypsin interface in the crystal structure of the tetrameric complex, a series of candidate residues was generated and validated through interactive computer modeling (Table 5.5). We focused on residues that are within 4.2 Å of the protease, have large buried surface areas (over 50 Å²) or large percentage change of buried surface area upon binding (over 50% buried), and residues with right geometry and location for intermolecular hydrogen bonding. These residues could be divided into two distinct groups. The primary site group includes residues 81-86 and 52-54 located on two surface loops (50s and 80s loops). The secondary site group consists of residues 67-70, 108, 110, 112, and 113, also from two surface loops (60s and 100s loops).

Table 5.5 Identification of Ecotin Secondary Site Residues

Residue	A _{uncomplex} (Å ²)	A _{complex} (Å ²)	A _{buried} (Å ²)	Percentage Buried %	Hydrogen Bond Distance (Å)
Leu52	146	50	96	66	3.1
His53	38	22	16	42	
Arg54	133	8	125	94	
Val81	83	6	77	92	2.9
Ser82	57	17	40	70	
Thr83	62	0.3	61.5	99.5	
Met84	209	0.5	208.5	99.8	3.0
Met85	133	9	124	93	3.0
Ala86	100	8	92	92	
Trp67	117	13	104	89	
Gly68	60	3	57	95	3.1
Tyr69	51	14	37	72	2.9
Asp70	71	43	28	40	
Arg108	180	87	93	52	
Asn110	24	13	11	47	2.9
Lys112	161	129	32	20	
Leu113	44	17	28	62	

A_{uncomplex} : Solvent accessible surface area in uncomplexed ecotin dimer
 A_{complex} : Solvent accessible surface area in ecotin-trypsin tetrameric complex
 A_{buried} : Calculated as the difference between A_{uncomplex} and A_{complex}
 Percentage Buried : Calculated as A_{buried} / A_{uncomplex} *100%
 Hydrogen Bond Distance : Putative hydrogen bond distance with closest side chain or main chain atoms in trypsin.

To investigate the roles of these residues in binding, we used an "alanine shaving" method (Jin & Wells, 1994), by which the amino acids within a specific region of the molecule of interest are substituted by alanines simultaneously. The multiple alanine substitutions may reduce the hydrophobic and electrostatic interactions provided by the side chains to a minimum if the conformation of the polypeptide backbone remains unaltered. We made two multiple alanine substitution variants of ecotin, designated as variant

50A₃+80A₅ and 60A₄+100A₄. Ecotin 50A₃+80A₅ has eight alanine substitutions at residues 81-85 (residue 86 is already an alanine) and 52-54. Variant 60A₄+100A₄ also has eight alanine substitutions at residues 67-70, 108, 110, 112 and 113. These two multiple alanine substitution variants were investigated to define the roles of the primary and secondary binding sites of ecotin. Their inhibition constants against bovine and rat trypsins are shown in Table 5.6.

Table 5.6 *K_i* of Ecotin Multiple Alanine Substitution Variants (unit: nM)

	Bovine Trypsin	Rat Trypsin
WT	0.31	0.93
50A ₃ +80A ₅	98.3	27900
60A ₄ +100A ₄	1.59	15300

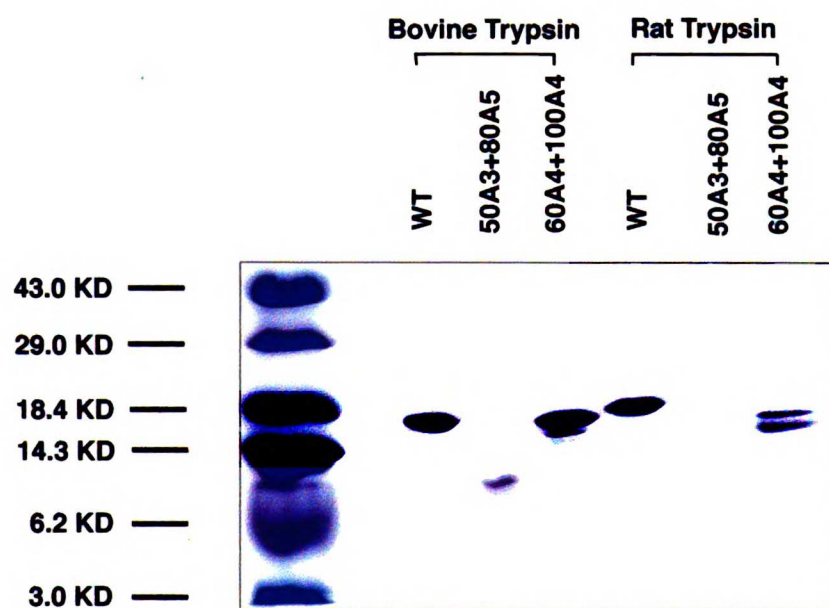
The *K_i* of ecotin 50A₃+80A₅ against rat trypsin increased 30,000-fold compared to that of WT ecotin, from 0.93 nM to 27.9 μM. Its *K_i* against bovine trypsin also increased significantly (about 300-fold). These results confirmed the essential role of the reactive site loop as a major determinant of the strength of protease inhibition. At the secondary binding site, ecotin 60A₄+100A₄'s *K_i* for bovine trypsin increased slightly (about 5-fold) compared to that of the WT ecotin. The remarkable potency of this variant demonstrated that ecotin remains competent as an inhibitor even after a significant part of the molecule was replaced with alanines. Nevertheless, the inhibition towards rat trypsin by variant 60A₄+100A₄ decreased over 16,000-fold, indicating that this loop contributes greatly to binding. WT ecotin inhibits both rat and bovine trypsin equally well, but ecotin 60A₄+100A₄ has acquired near 10,000-fold greater potency for bovine trypsin over rat trypsin.

This dramatic contrast highlighted the importance of the 60s and 100s loops for different proteases. To ensure that the loss of inhibition of ecotin 60A₄+100A₄ was not due to its intrinsic destabilization as monitored by proteolysis, variant 50A₃+80A₅, 60A₄+100A₄ and WT ecotin were incubated with rat and bovine trypsin at 37 °C for 16 hours and then loaded onto a denaturing SDS-PAGE gel (Figure 5.5). WT ecotin was stable to proteolysis under all conditions. Ecotin 50A₃+80A₅ was degraded by both bovine and rat trypsin. Ecotin 60A₄+100A₄ was stable in the presence of bovine trypsin under the assay condition, but became a transient inhibitor of rat trypsin and was subsequently proteolyzed into several smaller fragments. The protease over-digestion results correlated with the steady-state inhibition data with 20 minutes of incubation and confirmed the importance of the secondary binding site for rat trypsin binding.

Figure 5.5 Proteolytic Digestion of Ecotin Multiple Alanine Substitution Variants

The same amount of WT and variant ecotins (3 ng) was incubated with rat and bovine trypsin at 200 μM in pH 8.0 buffer at 37 °C for 16 hours and then loaded onto a denaturing SDS-PAGE gel.

**Figure 5.5 Proteolytic Digestion of
Ecotin Multiple Alanine Substitution Variants**



The Role of the 60s Loop and 100s Loop

These experiments established that the secondary binding site in ecotin plays a major role in the inhibition of serine proteases, as shown in the case of rat trypsin. The substitution of the eight residues in the 60s and 100s loops was sufficient to weaken the binding interactions between ecotin and one target protease by at least four orders of magnitude. Since these residues are located at two discontinuous regions of the surface loops, their contributions can be further dissected and analyzed. Ecotin variants 60A₄, M84R+60A₄, 100A₄, M84R+100A₄, and M84R+60A₄+100A₄ were designed to address this question. Their *K_i* values against the three serine proteases are listed in Table 5.7 and are also plotted in Figures 5.6, 5.7, and 5.8 for ecotin variants with alanine mutations at the 60s loop, the 100s loop, and both loops respectively. Together with variants M84R and 60A₄+100A₄ generated from the previous experiments, they formed a complete set of variants that allowed the separation and definition of the individual role of the 60s and 100s loops.

Table 5.7 *K_i* of Ecotin Combinatorial Variants (unit : nM)

	Bovine Trypsin	Rat Trypsin	uPA
WT	0.31	0.93	2800
M84R	0.09	0.38	3.6
60A ₄	0.26	4420	579400
M84R+60A ₄	0.14	0.18	1470
100A ₄	0.04	0.75	77800
M84R+100A ₄	0.03	0.04	4040
60A ₄ +100A ₄	1.59	15300	99100
M84R+60A ₄ +100A ₄	0.16	0.50	28200

60A₄: multiple alanine substitutions at positions 67-70.

100A₄: multiple alanine substitutions at positions 108, 110, 112, 113.

60A₄+100A₄: multiple alanine substitutions at positions 67-70, 108, 110, 112, 113.

Figure 5.6 Ki Data of Ecotin 60s Loop Variants

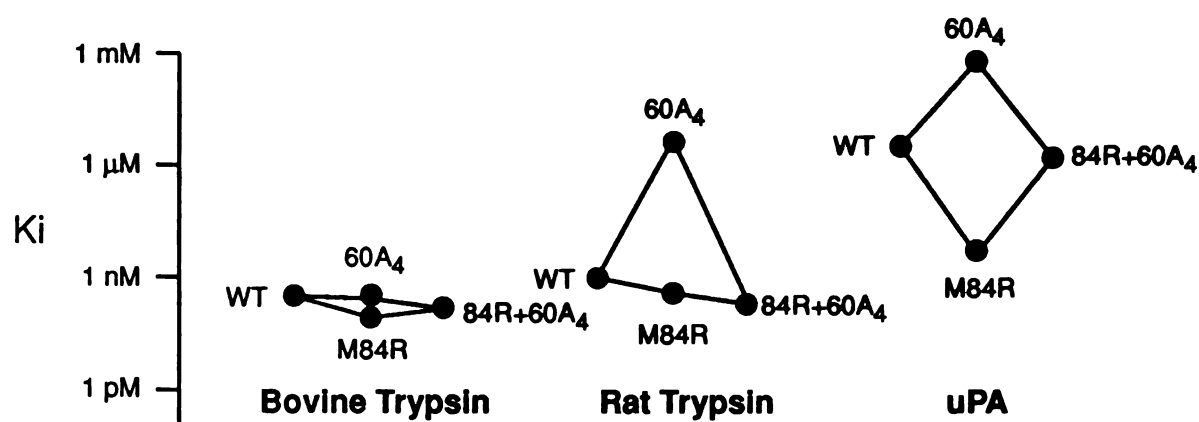


Figure 5.7 Ki Data of Ecotin 100s Loop Variants

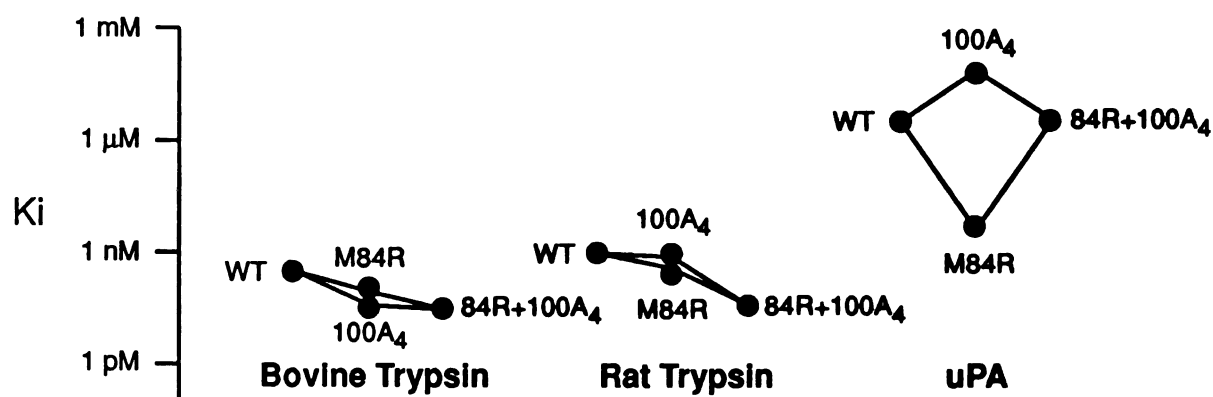
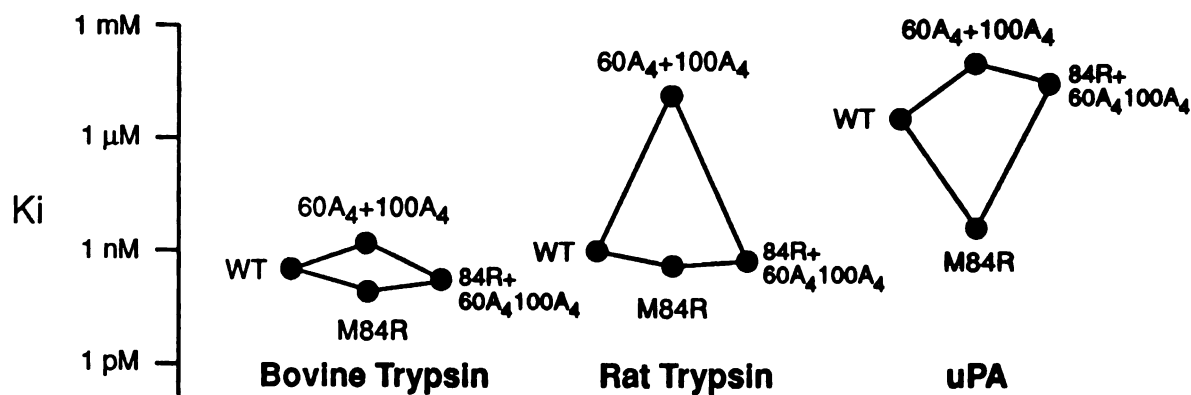


Figure 5.8 Ki Data of Ecotin 60s and 100s Loops Variants



Figures 5.6, 5.7, 5.8 provide an overview of the interactions among ecotin's two surface loops and various proteases. The *K_i* plots were used both horizontally and vertically to evaluate either the impact of specific surface loops on different enzymes, or the inhibition of a particular enzyme by different variants. Similar to the results from the previous experiment, the inhibition of bovine trypsin by ecotin was not affected by drastic changes to the inhibitor. The *K_i*'s of all variants were approximately 1 nM. Ecotin 100A₄ bound even slightly tighter than ecotin M84R (40 pM vs. 90 pM). Thus, the only dominant factor for bovine trypsin binding was the proper conformation of the primary site reactive loop. In the case of inhibition of uPA, it was clear that both the primary site (P1) and the two secondary site loops were critical for binding. The difference between the *K_i*'s of the alanine substitution variants and the *K_i* of WT ecotin against uPA were significant, ranging from 75 μM to 576 μM. The three double variants which carry simultaneously a favorable M84R mutation and unfavorable multiple alanine substitution variants failed to achieve the low nanomolar binding of M84R. Their *K_i* values range from 1.5 to 28.2 μM, strongly suggesting that the proper amino acid sequences at the 60s and 100s loops were essential for optimal interaction between uPA and ecotin. The *K_i* plot of rat trypsin indicated yet another pattern of secondary site inhibition. In this case, the 60s loop was responsible for providing most of the binding energy as indicated by the over 4,000-fold increase of *K_i* with ecotin 60A₄, while the effect of the 100s loop was minimal. In fact, for all three enzymes, the 60s loop played a more predominant role in determining the strength of the inhibition at the secondary site.

It was remarkable that the inhibition of each enzyme was achieved through the differential contributions from these two surface loops. It was also intriguing that the contributions from these loops were dependent on the identity of the P1 residue and of the target enzyme. For example, if the target enzyme was rat trypsin, the favorable substitution

M84R alone was sufficient to compensate for the adverse effects of the alanine substitution at the secondary binding site. Thus, the 60s loop was only important when P1 84 was a Met. However, the inhibition of uPA required not only a favorable P1 residue (Arg) at 84 position, but also key residues at both 60s and 100s loops.

Nonadditivity among Ecotin Variants

The K_i data set of the three groups of ecotin variants allow us to analyze the interrelationships among the key determinants of ecotin's inhibition potency and specificity. For example, we can ask the question of whether the separate mutations are additive by comparing the K_i values of ecotin 60A₄+100A₄ versus those of ecotin 60A₄ or ecotin 100A₄ against one particular protease. In the case of rat trypsin inhibition, the variant with alanine-substitutions at two loops bound more weakly than either of the two single alanine-substituted loop variants. For uPA binding, the 60s loop had a more significant effect. The K_i of ecotin 60A₄ against uPA was even higher than the K_i of the double mutant ecotin 60A₄+100A₄. It is clear that the contributions of these two loops were not independent of each other.

The K_i plot was effective in directly comparing the strength of inhibition by ecotin variants and illustrating the relationships between the P1 mutation and the secondary site multiple alanine substitutions (or C-terminal truncation). Alternatively, we can define additivity as the absence of interaction energy term ΔG_I° in

$$\Delta\Delta G_{(A,B)}^\circ = \Delta\Delta G_{(A)}^\circ + \Delta\Delta G_{(B)}^\circ + \Delta G_I^\circ \quad (5.1),$$

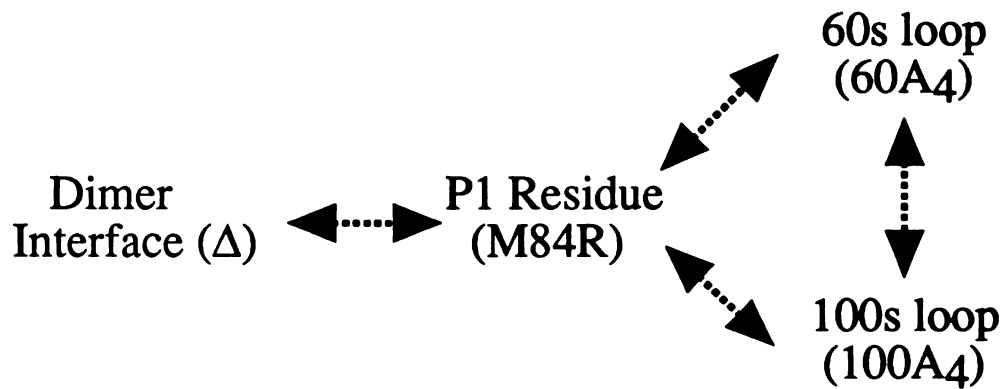
where $\Delta\Delta G_{(A,B)}^\circ$ is the ΔG difference between double mutant A, B and WT, and $\Delta\Delta G_{(A)}^\circ + \Delta\Delta G_{(B)}^\circ$ are the differences between the single site variant A or B and the wild-

type, respectively (Wells, 1990). By converting K_i to free energy using $\Delta G_{(X)}^{\circ} = -RT \ln K_i$ (2), we can calculate the ΔG_I° term for mutations at different parts of ecotin. Fifteen double mutation cycles were constructed from our original K_i data. These cycles were based on three different target proteases and between ecotin M84R and ecotin WT Δ , between ecotin M84R and ecotin 60A₄, between ecotin M84R and ecotin 100A₄, and between ecotin 60A₄ and ecotin 100A₄. The calculated ΔG_I° values are listed in Table 5.8. Mutational additivity has been seen in other model systems such as growth hormone and its receptor (Wells, 1996), as well as subtilisin (Wells *et al.*, 1987) and the third domain of turkey ovomucoid (Lu *et al.*, 1997; Qasim *et al.*, 1997). In those cases, the interaction energy ΔG_I° was shown to be small, usually less than 0.4 Kcal/mol. In our system, however, the ΔG_I° was much larger. Half of the ecotin interaction energy values were greater than 1 Kcal/mol. The nonadditivity was most pronounced between ecotin M84R and ecotin 60A₄ mutations when the target enzyme was rat trypsin. The ΔG_I° for this particular mutation cycle was 5.46 Kcal/mol. The large interaction energy highlighted the dynamic linkage among the key structural and functional elements of ecotin as shown in Figure 5.9. These data provide evidence for cooperative and synergetic interactions among the P1 residue of ecotin and the two surface loops and the dimer interface. It also suggests that the unique flexibility and adaptability of ecotin are the structural basis for the observed broad specificity.

**Table 5.8 Relative Change in Interaction Energy among Ecotin's
Primary, Secondary Sites and Dimer Interface (unit: Kcal /mol)**

Mutations		Bovine Trypsin	Rat Trypsin	uPA
A:	M84R	0.98	3.98	-0.93
B:	Δ			
A:	M84R	-0.37	5.46	-0.40
B:	60A ₄			
A:	M84R	-0.56	1.21	-2.19
B:	100A ₄			
A:	M84R	0.63	5.59	-3.20
B:	60A ₄ +100A ₄			
A:	60A ₄	-2.29	-0.86	3.02
B:	100A ₄			

Figure 5.9 Nonadditivity among Ecotin Mutations



The interdependent relationships among ecotin's four key structural and functional elements are illustrated. The network of interactions contributes to the high affinity and broad specificity of ecotin against a wide range of serine proteases within the chymotrypsin subfamily.

Discussion

Ecotin's Secondary Site as a Key Element in Macromolecular Recognition

We have used several approaches to establish the role of the secondary site of ecotin in protease recognition and inhibition. The experimental results can be interpreted based on the unique network of interactions between ecotin and its target protease. The impact of P1 substitutions in ecotin was very different from that of the same substitutions in canonical serine protease inhibitors such as BPTI or turkey ovomucoid third domain (Lu *et al.*, 1997). For these familiar serine protease inhibitors, there is usually a 100 to 1000-fold differences in K_i 's against trypsin between variants carrying a positively-charged and an uncharged apolar residue at the P1 position. However, the K_i of ecotin against rat trypsin was insensitive to alterations at this position. Other research groups have also obtained similar results using ecotin against other serine proteases (Pal *et al.*, 1994; Seong *et al.*, 1994; Seymour *et al.*, 1994). These results demonstrate that the side chain of the P1 residue does not play a predominant and independent role in determining the inhibitory specificity of ecotin. The binding energy of ecotin must come from other parts of the molecular interface. Structural analysis of the trypsin-ecotin complex suggests that a logical candidate is the secondary binding site which is composed of the two surface loops covering residues 66-70 and 108-113. These loops derive from the contralateral ecotin in the tetramer with respect to the primary loops.

The C-terminal truncation mutagenesis experiment which affects the hinge between the ecotins reinforced the importance of the secondary site and its alignment. The differential inhibition of the three proteases by ecotin WT Δ was a novel characteristic of this inhibitor.

A plausible explanation could be made based on two assumptions. First, the secondary binding site might play different roles in ecotin binding to different enzymes. Second, the perturbation of the dimer interface changed the positions and/or conformations of the secondary binding site relative to other parts of the molecule. Therefore, if rat trypsin (or uPA) required the secondary site for tight binding, and if the C-terminal truncation dislocated the secondary binding site, the interactions between ecotin WTΔ and rat trypsin (or uPA) would be severely weakened. In the network of interactions between the protease and ecotin, ecotin's C-terminal hinge region may affect the binding by modulating the relative contributions from the primary and secondary binding sites. The dramatic increase of K_i values of WTΔ against rat trypsin and uPA offered indirect evidence of the critical roles of the secondary site surface loops.

Finally, the multiple alanine substitution experiments strongly support the notion that the 60s and 100s loops could provide the critical binding energy necessary for protease inhibition. The loss of function of the variants not only establish the essential role of the secondary site in rat trypsin and uPA binding, but also reveal an interdependent relationship between the 60s and 100s loops, suggesting the dynamic nature of the secondary site interaction. The K_i plot of the 60s loop variants was almost superimposable onto that of the truncation variants. These two different large scale modifications of ecotin have actually generated strikingly similar consequences in terms of the inhibitory activity. It implies a fundamental connection between the dimer interface and the secondary binding site, two regions structurally far from each other. These two experiments further support the hypothesis that the dimer interface affected binding through dislocating the secondary site loops at the protease inhibitor interface.

Monomer Dimer Equilibrium

The interaction of the protease with the secondary site of ecotin requires the formation of a stable tetrameric complex. Although the K_d of ecotin's monomer dimer equilibrium is much higher than K_i values against most target proteases, the K_d of ecotin dimerization could be dramatically perturbed in the presence of proteases. Proteases may serve as templates to facilitate the dimerization of ecotin. If there are strong cooperative interactions upon the binding of the second inhibitor and the second enzyme molecule, the reaction pathway may proceed from $I \rightarrow EI \rightarrow EI_2 \rightarrow E_2I_2$, where the last two steps are rapid and drive the tetrameric complex formation to completion. Analytical centrifugation is being used to pursue this question. Preliminary data show that trypsin monomer, ecotin dimer, and ecotin-trypsin heterotetramer were the only three coexisting homogenous species in solution when a sample of trypsin and ecotin mixture was subjected to velocity sedimentation (Doyle, unpublished results). Recently, a C-terminal truncation ecotin variant was shown to be monomeric by gel filtration (Pal *et al.*, 1996). This discrepancy with our results indicates that the solution state of ecotin truncation variants in the absence of protease may vary with buffer and other variables. However, that study also showed that upon the addition of the protease, the tetrameric complex was formed. All the data in our study support the conclusion that ecotin protease complex is tetrameric under a wide range of concentrations and that the secondary binding site has crucial roles in the complex formation and stabilization.

Protease Inhibitors with Multiple Binding Sites

Ecotin offers unique opportunities to study the complex network of interactions between serine proteases and the bi-dentate macromolecular inhibitors. It is also an ideal scaffold to design and engineer protease inhibitors. Potency and specificity toward target

serine proteases may be introduced through the secondary binding site, a special structural feature of ecotin that does not resemble any previously known binding motifs within the families of macromolecular serine protease inhibitors. There are few serine protease inhibitors with multiple binding sites for one protease molecule. Hirudin, rhodniin and ornithodorin are three examples of bi-dentate inhibitors against thrombin (Rydel *et al.*, 1990; van de Locht *et al.*, 1995; van de Locht *et al.*, 1996). They bind to thrombin at both the active site and a highly charged exosite in a one-to-one complex. Unlike the predominant electrostatic interaction between these three inhibitors and the thrombin exosite, the recognition between the secondary binding site of ecotin and its target protease is driven by a combination of hydrophobic, electrostatic, and hydrogen bond forces. Ecotin is also different from hirudin, rhodniin and ornithodorin because ecotin's two binding sites are provided by two monomers through dimerization. This unique bi-dentate binding mode has two advantages for macromolecular recognition for diverged proteases. First, it creates a buried surface area that is nearly 50 percent larger than the hirudin/rhodniin/ornithodorin-thrombin interface (about 1900 Å²), allowing a large variety of interdependent factors to contribute to the formation of the ecotin-protease complex. The macromolecular recognition is very sensitive to the specific residues located at the binding interface. Second, the tetrameric network of interactions creates greater flexibility to modulate the strength of inhibition by introducing new controlling elements such as the "hinge" region at the dimer interface. In this aspect, the ecotin-protease interaction is reminiscent of the antibody-antigen interaction, in which the hyper-variable regions of the six CDR loops of immunoglobulin provide all the possible surface landscapes to recognize any given antigen through an astronomical number of combinatorial side chain conformations (McGrath *et al.*, 1995). Ecotin's four surface loops, the 50s, 60s, 80s, and 100s loops, have great potential to be tailored to provide a complementary fit with different

protease surfaces that are in direct contact with both the primary and secondary binding site.

The dramatic *K_i* differences of ecotin variants containing multiple alanine substitutions at the secondary site against bovine trypsin, rat trypsin, and uPA highlight the novel structural and functional features of ecotin. Rat and bovine trypsins are closely related enzymes with the same substrate specificity. Their crystal structures are very similar. Although rat trypsin is anionic, while bovine trypsin is cationic, it is still not clear whether the electrostatic interaction is solely responsible for the selective inhibition of rat trypsin by the ecotin 60s loop variants. Other crucial structural factors may contribute to the binding energy of complex formation, for example, hydrophobic packing of the interface, or the flexibility and rigidity of the surface loops.

A systematic approach has been taken to investigate the dynamic relationships among the four major determinants of the interactions between ecotin and its target protease, the primary reactive site loop, the 60s and 100s loops, and the dimer interface at the C-terminal hinge region of ecotin. Our data suggest that the mutations at these four regions are nonadditive. Furthermore, the importance of ecotin's unique secondary binding site in recognizing and differentiating between target protease has been firmly established. A comprehensive set of ecotin variants at the primary and secondary sites have been generated to probe the molecular recognition between ecotin and target serine proteases. Ecotin binds to proteases through distinct combinations of interactions between the primary and secondary site surface loops. The primary and secondary sites are adapted to the surface features by their intrinsic flexibility and by the C-terminal hinge which permits relative adjustments between the primary and secondary sites. The discriminating power of the individual surface loops at the secondary site of ecotin suggests a novel opportunity to exploit the subtle difference among proteolytic enzymes with identical primary substrate

specificity and to design selective and potent macromolecular inhibitors against these enzymes.

Materials and Methods

Materials

The *E. coli* strain JM101, XL-1 Blue F' and the VCSM13 helper phage were from Stratagene (La Jolla, CA). The *E. coli* ecotin gene deletion strain IM Δ ecoJ was derived from JM101 and was the generous gift of Dr. Iain Murray (Murray and Craik, unpublished results). Enzymes and reagents for the manipulation of DNA were purchased from Promega (Madison, WI) or New England Biolabs (Beverly, MA) and were used following the manufacturer's instructions. Low molecular weight uPA (LMuPA) was obtained from American Diagnostica (Greenwich, CT). Bovine trypsin was from Sigma (St. Louis, MO). Rat trypsin was expressed in *E. coli* using the expression vector pZ3 and purified as described (Higaki *et al.*, 1989). Wild-type and variant forms of ecotin were produced in bacteria from the expression vector pTacTacEcotin following protocol described below. The chromogenic substrate N α -benzyloxy-carbonyl-L-glyylproylarginine-7-amino-4-methylcoumarin (Z-GPR-AMC) was purchased from Bachem Biosciences, Inc (King of Prussia, PA). The chromogenic substrate Z- γ -Glu(α -t-butoxy)-Gly-Arg-*p*-nitroanilide (Spectrozyme UK) used for LMuPA kinetic analysis was from American Diagnostica (Greenwich, CT). 4-Methylumbelliferyl *p*-guanidinobenzoate (MUGB) was from Sigma (St. Louis, MO). DNA oligonucleotides were synthesized on a Perkin Elmer/Applied Biosystems 391 DNA synthesizer (Foster City, CA) by using the phosphoramidite method and reagents from the same company. Synthetic oligonucleotides were purified with NENSORB cartridge from DuPont NEN (Boston, MA). DNA sequencing was performed with a Sequenase 2.0 kit from U.S. Biochemical Corp. (Cleveland, OH). α -³⁵S-dATP was from DuPont NEN (Boston, MA). GeneClean[®] was from Bio101, Inc (La Jolla, CA). Spectra/Por[®] molecularporous dialysis membrane was from Spectrum (Laguna Hills, CA).

Amicon Centriprep-10 and Centricon-10 were from Amicon (Beverly, MA). All other chemicals were of reagent grade or better and were used without further purification.

Computer Modeling of Ecotin

Ecotin WT Δ (133-142): The ecotin-ecotin interface was determined by using the program InsightII (Dayringer *et al.*, 1986) to search for all ecotin residues in one monomer that were within 4.2 Å of the other monomer. In the 2.4 Å X-ray structure of ecotin bound to D102N trypsin (McGrath *et al.*, 1994), the ecotin-ecotin dimer interface buries 3000 Å² of surface area as calculated with the program Access using a 1.4 Å probe size. In addition it makes 10 monomer-monomer hydrogen bonds as calculated using InsightII with a distance cut off of 3.0 Å and an angle cut-off of 120°. Removal of residues 133-142 was calculated to remove 2000 Å² buried surface area in the dimer interface leaving a possible dimer interface of only 1000 Å² with only 4 of the 10 original monomer-monomer hydrogen bonds.

Ecotin 50A₃+80A₅ and Ecotin 60A₄+100A₄: Residues in the ecotin-trypsin interface were determined from the 2.4 Å X-ray structure of ecotin-D102N trypsin (McGrath *et al.*, 1994) using InsightII by finding all residues from one molecule within 4.2 Å of the other molecule and vice versa. Area buried was computed according to Lee and Richards as implemented by the program Access using a probe size of 1.4 Å (Lee & Richards, 1971). Candidate residues for mutation to alanine were all ecotin residues whose solvent accessible surface decreased 50% or more upon binding to trypsin in either the primary binding site (ecotin 50A₃+80A₅) or secondary binding site (ecotin 60A₄+100A₄) with the exception of Gly66 which had an unusual *phi psi* angle such that its changing to another residue would destroy the secondary structure. This resulted in 8 residues in the primary site which were changed to alanine. This group of residues is comprised of Leu52, His53, Arg54, Val81,

Ser82, Thr83, Met84, Met85, and Ala86 (left unchanged). The resulting construct was designated ecotin 50A₃+80A₅. In the secondary site residues Trp67, Gly68, Tyr69, Asp70, Arg 108, Asn110, Lys112 and Leu113, were all substituted with alanines. This variant was referred to as ecotin 60A₄+100A₄.

Ecotin Mutagenesis

Site-directed mutagenesis was performed by the method of Kunkel (Kunkel, 1985). All mutations have been confirmed by DNA sequencing of the mutation region.

Plasmids for the production of M84X (where X=R, K, F, W)

The codon in plasmid pBS Ecotin that codes for Met84 was changed to code for Arg84, Lys84, Phe84 or Trp84 by site-directed mutagenesis using oligonucleotides M84R, M84K, M84F, or M84W respectively. The following is a list of these primers (underlined nucleotides were base changed, italics codons were residues changed).

Primer M84R : 5'-GT TCC CCG GTT AGT ACT AGG ATG GCC TGC C-3' [with a unique ScaI site]

Primer M84K : 5'-GT TCC CCG GTT AGT ACT AAG ATG GCC TGC C-3'

Primer M84F: 5'-GT TCC CCG GTT AGT ACT TTC ATG GCC TGC C-3'

Primer M84W : 5'-GT TCC CCG GTT AGT ACT TGG ATG GCC TGC C-3'

Plasmids for the production of the C-terminal deletion variants WTΔ and M84RΔ

pBS Ecotin and pBS Ecotin M84R were used as templates in a PCR reaction with oligonucleotides EcoN and EcoC to generate C-terminal deletion variants at residues 133-142. The amplified region was digested with BamHI and HindIII, purified and ligated to the BamHI/HindIII fragment of pTacTacEcotin . The PCR condition used were: 94 °C for

1 min, 40 °C for 2 min, 72 °C for 3 min, 35 cycles. The sequences of the two primers were:

Primer EcoN: 5'-AAATTAACCCTCACTAAAGGG-3'

Primer EcoC: 5'-TTGTCAATTTAAGCTTACGCCTTCCAGACGCGG-3'

Plasmids for the production of multiple alanine substitutions at the secondary binding site. The DNA in pBS Ecotin was changed to code for alanines at positions 52-54, 67-70, 81-85, 108,110, 112-113 by site-directed mutagenesis using oligonucleotides 50A₃, 60A₄, 80A₅, or 100A₄ respectively. Each of these primers carries a unique restriction site, their sequences are as follows,

Primer 50A₃: 5'-CTG GAA GTC GAT TGC AAT *GCG GCT GCC TTG* GGC GGG AAG CTG GAA AAC-3' [unique StyI site]

Primer 60A₄: 5'-AAC AAA ACG CTG GAA GGG *GCC GCC GCG GCC* TAT TAT GTC TTT GAT AAA GTC-3' [unique SfiI site]

Primer 80A₅: 5'-AAA GTC AGT TCC CCG *GCT GCA GCG GCG GCG GCA* TGC CCG GAT GGC AAG-3' [unique SphI site]

Primer 100A₄: 5'-GC GAT GCT GGA ATG CTG *GCT TAC GCT AGC GCG GCG* CCG ATC GTG GTG TAT AC-3' [unique NheI site].

The region in the vector containing the desired mutation was digested with BamHI and HindIII and ligated to the BamHI/HindIII fragment of expression vector pTacTacEcotin.

Expression and Purification of Recombinant Ecotin

IMΔecoJ cells were transformed with pTacTac Ecotin and transformants were selected by plating on LB/ampicillin plates. A single colony was used to inoculate 3 ml of LB containing 60 µg/ml ampicillin. The cultures were grown at 37 °C for 9 hours and diluted to 1 liter of LB/ampicillin. Following growth at 37 °C for 1 hour, IPTG was added to the

cultures to a final concentration of 0.2 mM, and continued to grow for 12 hours at 37 °C. Cells were harvested and treated with lysozyme in a solution containing 25% sucrose/10 mM Tris, pH 8.0. The periplasmic fraction was dialyzed against 10 mM Tris, pH 8.0 with a Spectra/Por molecularporous dialysis membrane of 12-14 K molecular weight cut-off. Following dialysis, the supernatant was acidified to pH 2.8 by addition of 1M HCl to the sample. The sample was incubated on ice for 30 minutes and the acid-insoluble material was removed by centrifugation. The supernatant was neutralized to pH 7.4 with 1 M Tris, pH 8.0, and adjusted to a NaCl concentration of 0.3 M. The solution was heated at 100 °C for 10 minutes, and then cooled to room temperature. The precipitate was removed by centrifugation, and the supernatant was dialyzed against deionized distilled water overnight at 4 °C. The ecotin sample was concentrated with a 10 K molecular weight cut-off concentrator(Amicon Centriprep-10). The concentrated sample was loaded onto a Vydac C4 reverse-phase high performance liquid chromatography column (2.2 x 25 cm) that had been equilibrated with 0.1% trifluoroacetic acid. The column was washed and ecotin was eluted with a linear gradient of 34-37% acetonitrile/0.1% trifluoroacetic acid at a flow rate of 10 ml/minute over 30 minutes. Fractions were analyzed individually by SDS-PAGE and those containing pure ecotin fractions were pooled, and lyophilized. Purified ecotin was redissolved in water and stored at 4 °C. The concentrations of ecotin and ecotin variants were determined using a calculated molar extinction coefficient of 14890 cm⁻¹ M⁻¹ for variants 60A₄, M84R+60A₄, 60A₄+100A₄, WTΔ, M84RΔ, 21860 cm⁻¹ M⁻¹ for wild-type (WT) and variants 100A₄, M84R+100A₄, M84R, M84K, and M84F, and 27550 cm⁻¹ M⁻¹ for M84W.

Determination of Ecotin Ki's against Various Proteases

Rat and bovine trypsin activity assays were using the same substrate and procedure. Trypsin was titrated with 4-methylumbelliferyl *p*-guanidinobenzoate (MUGB) to obtain an accurate concentration of the enzyme's active sites. Various concentrations of ecotin or ecotin variants were incubated with trypsin in a total volume of 990 μ l of buffer containing 50 mM NaCl/50 mM Tris/10 mM CaCl₂, pH 8.0. Following a 10-minute equilibration at room temperature, 10 μ l of 2.5 mM Z-GPR-AMC substrate was added and the rate of substrate hydrolysis was measured by monitoring the change of emission at 460 nm in a 2-minute period at 25 °C. LMuPA assays were performed as described. All the data were fit to the equation derived for kinetics of reversible tight-binding inhibitors by nonlinear regression analysis using the program Kaleidagraph, and the values for apparent K_i and true K_i were determined.

Chapter Six: Engineering Bidentate Macromolecular Inhibitors for Trypsin and Urokinase

ABSTRACT

Ecotin, a dimeric serine protease inhibitor from *Escherichia. coli*, is a novel platform for inhibitor design. A new approach using three-dimensional structural information to guide the design efforts was taken to create potent bidentate inhibitors of trypsin and human urokinase-type plasminogen activator(uPA). The ecotin surface loop encompassing residues 67 to 70 (60s loop) which bind to the target protease at a region distal from the enzymes' active site was redesigned. Two ecotin phage display libraries were constructed to exploit the binding interactions at the 60s loop. The ecotin 60X₄ library, in which residues 67-70 of ecotin were randomized, was panned against rat and bovine trypsin in parallel for four rounds. Panning against bovine trypsin resulted in enrichment of ecotin phage but did not yield a consensus sequence. Panning against rat trypsin resulted in enrichment as well as the ecotin consensus sequence WGFP at positions 67-70. The variant ecotin encoded by this sequence inhibited rat trypsin at 80 pM, a 12 fold improvement over ecotin wild-type(WT). A second generation library M84R+60X₄ including an additional methionine to arginine substitution in the primary binding site of ecotin was generated for panning against uPA and rat trypsin. Panning against rat trypsin resulted in enrichment but no consensus sequence. Panning against uPA resulted in enrichment as well as a different ecotin consensus sequence WGYD. The ecotin variant M84R+D70R bound to uPA at 50 pM, a 56,000-fold increase in binding compared to WT ecotin. Furthermore, ecotin M84R+D70R achieved 13,680-fold improvement of specificity towards uPA vs. rat trypsin. That the 60s loop of ecotin plays different roles in

binding to trypsin and uPA suggests this site can be used to introduce specificity and potency for other members of the serine protease family.

Introduction

The serine proteases are a large family of enzymes involved in a wide variety of vital biological processes. The crucial physiological functions of these enzymes in blood coagulation, fibrinolysis, complement pathways, viral maturation, apoptosis, and cancer make them important targets for efforts to design and engineer potent and specific inhibitors. A highly selective protease inhibitor can serve as a powerful tool to block key proteolytic activities for the dissection of proteolytic pathways and cascades and elucidation of the *in vivo* roles of particular proteases in complex biological processes. Ultimately, this may lead to the development of innovative therapies for life-threatening diseases.

Previous studies have established that, for the chymotrypsin family of serine proteases, the electrostatic and hydrophobic characteristics of the substrate binding pocket are critical factors determining the substrate specificity (Perona & Craik, 1995). This region has been the major focus of structural, functional, and mechanistic studies as well as inhibitor design efforts. Since many mammalian serine proteases have the same primary specificity, a small molecule inhibitor that only exploits the interactions at the primary binding pocket may not have sufficient discrimination towards similar enzymes. Macromolecular substrate-like serine protease inhibitors, such as bovine pancreatic trypsin inhibitor (BPTI), bind to the serine protease through a single surface loop that includes the critical P1 residue. This residue fits into the binding pocket of the target protease in a substrate-like conformation to lock the enzyme in a complex formed between the protease and the inhibitor. The common contact area between the protease and the inhibitor is small and clustered around the active site of the enzyme. The highly homologous structures of the enzyme active site and the S1 pockets make inhibitor design a difficult and challenging task.

Ecotin, a macromolecular serine protease inhibitor found in the periplasm of *Escherichia. coli*, offers a unique platform to explore the protease-inhibitor interactions that are distal from the active site (Chung *et al.*, 1983). Ecotin has extremely broad specificity, inhibiting a large number of serine proteases in the chymotrypsin family such as trypsin, chymotrypsin and elastase to different degrees (McGrath *et al.*, 1991). The three-dimensional crystal structure of an ecotin-rat trypsin complex reveals a heterotetramer with a two-fold symmetry formed among two ecotin monomers and two trypsins (McGrath *et al.*, 1994). Two ecotin monomers dimerize in a head-to-tail fashion providing two binding sites at opposite ends of the dimer. Each trypsin binds to two ecotins at the two well-separated binding sites. One site is composed of the reactive site loop of ecotin containing the P1 Met at position 84. The other site that is 25 Å away ($\pm 3-5$ Å dependent upon the ligand) binds to the serine protease at the C-terminal loops and helix region. In the protease-inhibitor structure, four ecotin surface loops are in direct contact with trypsin, including the 50s(residues 52-54) and 80s(residues 81-86) loops at the primary reactive site, and the 60s(residues 67-70) and 100s(residues 108-113) loops at the secondary binding site. These four loops form two interface regions between ecotin and the protease resulting in a combined surface area of 2800 Å². The enormous buried interface area between ecotin and its target protease is far greater than that of most other protease-inhibitor complexes. In addition, ecotin's unique secondary binding site plays a major role in determining the strength of interaction between ecotin and the protease. Through systematic mutagenesis at the 60s and 100s loops, we have demonstrated the significance of this secondary binding site for selective inhibition against several proteases, such as rat trypsin and urokinase-type plasminogen activator (uPA) (Yang & Craik, 1997).

uPA is a serine protease that plays an active role in extracellular proteolysis, cell migration, and tissue remodeling processes (Fazioli & Blasi, 1994). Because of its

implication in cancer metastasis and tumor invasion, uPA has become an important target for drug design and inhibitor development efforts. To investigate the macromolecular recognition between uPA and ecotin, we have shown previously that a high affinity inhibitor of uPA could be isolated from a library of phage-displayed ecotin variants at the P1 and P1' positions (Wang *et al.*, 1995). The current study is focused on ecotin's secondary binding site to determine if potency and specificity against target proteases could be obtained through this unique surface region. To achieve this goal a structural-based approach coupled to phage display has been developed to design uPA inhibitors through stepwise optimization.

Results

Identifying Key Residues at the 60s loop of the Secondary Binding Site

Taking advantage of its large binding interface and the unique bidentate mode of interaction, ecotin effectively recognizes a broad range of serine proteases with different substrate specificities. In the accompanying paper (Yang & Craik, 1997), the roles of the four surface loops (the 50s, 60s, 80s and 100s loops) have been systematically dissected and the key elements responsible for the binding energy of the complex formation have been identified. In particular, the contributions from individual amino acid side chains at ecotin's secondary site surface loops were probed by substituting a cluster of residues with multiple alanines simultaneously. The multiple alanine substitution variants were designed to remove the majority of side chain interaction at the binding interface, assuming the main chain conformation of ecotin remained the same (Jin & Wells, 1994). The key residues were identified through a structure-based approach by examining the crystal structure of the ecotin-rat trypsin complex with multiple criteria, such as distance to the protease, buried surface area, and potential to form hydrogen bonds or other electrostatic interactions. Eight residues at ecotin's secondary site surface loops, all within 4.2 Å of the protease, were picked as the final candidates to provide the potential binding energy for the recognition and binding of the protease. They were residues Trp67, Gly 68, Try69, Asp70 at the 60s loop, and residues Arg108, Asn110, Lys112 and Leu113 at the 100s loop. The results of this region-specific mutagenesis study firmly established the significance of the secondary binding site and showed that this site played different roles with different target enzymes. The complete experimental results and detailed analysis are described in the proceeding paper (Yang & Craik, 1997).

Table 6.1 summarizes the results of multiple alanine substitution experiments at the 60s loop, a central focus for these inhibitor design efforts. Three highly homologous serine proteases in the chymotrypsin family, bovine trypsin, rat trypsin and uPA were used to test the activities of ecotin variants. Since these three enzymes have the same Arg/Lys specificity at the P1 site, they offered an ideal opportunity to concentrate on the interactions at ecotin's secondary binding site. The experimental data showed that the three enzymes were inhibited to different extents by the same set of ecotin variants. The inhibition of bovine trypsin was insensitive to changes at position 84 and the 60s loop of ecotin. For rat trypsin, multiple alanine substitutions at ecotin's 60s loop had a dramatic effect on binding; the K_i difference between ecotin 60A₄ and WT was over three orders of magnitude. However, the loss of inhibition due to the 60s loop alternation towards rat trypsin was context dependent. Once the favorable M84R mutation at the P1 residue was combined with 60A₄, the potency of inhibition by the double variant, ecotin M84R+60A₄ against rat trypsin was completely recovered, suggesting the dominant role of electrostatic interactions in rat trypsin binding. On the other hand, optimal binding of uPA required proper residues at both the 60s loop and the P1 position of ecotin. The loss of side chain interaction at the 60s loop increased the K_i towards uPA to over 500 μ M. Meanwhile, a single M84R mutation was able to increase affinity by almost three orders of magnitude. Furthermore, without the wild-type residues at the 60s loop, the favorable M84R mutation alone was not sufficient to restore the high affinity towards uPA. The sharp contrasts of the K_i 's of the two ecotin alanine substitution variants (60A₄ and M84R+60A₄) towards bovine trypsin, rat trypsin and uPA illustrated that there were substantial differences among the corresponding C-terminal surface regions of the proteases, where the 60s loop bound to the target enzyme. These experiments confirmed the potential of residues 67-70 to serve as novel determinants of specificity and potency of protease inhibition.

Table 6.1 Ki of Ecotin Combinatorial Variants (unit : nM)

	Bovine Trypsin	Rat Trypsin	uPA
WT	0.31	0.93	2800
M84R	0.09	0.38	3.6
60A ₄	0.26	4420	579400
M84R+60A ₄	0.14	0.18	1470

Through site-directed and region-specific mutagenesis, a four amino acid residue cluster at the 60s loop has been located as the starting point for engineering ecotin to be a highly specific and potent inhibitor. With limited understanding of the nature of the interactions at this region, it was unrealistic to conduct *de novo* design of a specific ecotin variant that would enhance its binding towards a target protease. On the other hand, it was equally difficult to test all the amino acid combinations on this surface loop individually since the total number of variants is 160,000 (20^4), beyond the limit of our current assay methodology. A combinatorial method was developed to approach this problem by taking advantage of the *in vitro* selection power of phage display, to directly isolate strong binding ecotin variants from a comprehensive 60s loop library.

Designing and Constructing Ecotin Phage Display Libraries at the 60s loop

Phage display has been successfully applied to a wide range of peptides and proteins, including antibodies (McCafferty *et al.*, 1990), growth hormone (Bass *et al.*, 1990), DNA binding proteins (Jamieson *et al.*, 1994; Rebar & Pabo, 1994), enzymes (Corey *et al.*, 1993; McCafferty *et al.*, 1991; Soumillon *et al.*, 1994), and macromolecular protease inhibitors (Markland *et al.*, 1996a; Markland *et al.*, 1996b; Pannekoek *et al.*, 1993; Roberts

et al., 1992; Wang et al., 1995). We have developed an ecotin phage display system that presents fully functional ecotin on the surface of filamentous phage (Wang et al., 1995). The ecotin phage display vector pBSeco-gIII expresses the fusion protein of full length ecotin connected to the C-terminal domain of filamentous phage minor coat protein pIII via a GlyGlyGly linker. With the addition of VCSM13 helper phage, the ecotin-pIII fusion protein is assembled onto phage particles. Phage carrying this fusion protein has ecotin activity and can bind to the immobilized protease on the solid surface. The three enzymes, bovine trypsin, rat trypsin, and uPA were coated onto polystyrene petri dishes and remained active as monitored by para-nitroanilide release of Z-GPR-pNa substrate after 30 min incubation. Ecotin phage offer a convenient vehicle to carry a large number of variant ecotins for panning against any target protease because of its direct physical linkage between the genotype and the phenotype. The ecotin phage display system allows the thorough investigation of the molecular recognition between ecotin and serine proteases without extensive *a priori* knowledge of the secondary binding site.

Two different ecotin libraries were designed, taking advantage of three dimensional structure information to randomize key residues thereby permitting isolation of the optimal cognate inhibitor for a target protease. The first library, 60X4, had four residues randomized at positions 67-70 of the 60s loop. This library was used to pan against bovine trypsin and rat trypsin separately. From the alanine substitution experiments, we predicted that bovine trypsin would bind to many ecotin 60s loop variants equally well regardless of the identities of the residues at positions 67-70. In contrast, only a small subset of ecotin variants within the 60s loop library should inhibit rat trypsin with high affinity. For rat trypsin inhibition, the side chain contribution from the 60s loop was important in the absence of strong favorable electrostatic interactions at the P1 position. Selective binding towards rat trypsin would certainly put greater constraints upon the nature of the amino acid

moieties at positions 67-70. Panning the ecotin 60X₄ library with rat trypsin would likely produce one or a few strongest binders from the millions of library members. The second library, ecotin M84R+60X₄, combined a favorable P1 Arg residue with the randomized 60s loop. This library was designed to encode ecotin variants that inhibit uPA with great potency. The K_i of ecotin M84R+60A₄ against uPA is 1470 nM, several hundred fold higher than the K_i of ecotin M84R, suggesting that four amino acid substitutions 45 Å away from the active site were sufficient to induce a dramatic change in the strength of the interaction. Since the inhibition of uPA was so sensitive to the residue substitutions at the 60s loop even in the presence of M84R, we reasoned that the uPA panning experiment with the ecotin M84R+60X₄ library had a high probability of generating a subset of ecotin variants that recognize uPA with increased potency and specificity.

The two ecotin libraries were constructed through Kunkel mutagenesis. The NNS (N=A,C,G,T, S=C,G) encoding scheme allowed a maximum of 1.0×10^6 possible DNA sequences that represent 1.6×10^5 possible amino acid residue combinations. The ecotin 60X₄ library contained approximately 5 million individual clones; the ecotin M84R+60X₄ library contained approximately 50 million individual clones. The completeness of the library was calculated using the equation.

$$N = \ln(1-p) / \ln[1-(1/n)] \quad (1)$$

where N is the number of total individual clones in the library, n is the number of possible combinations, and p is the probability that any clone can be found in the library given library size N. In both cases, the sizes of the final libraries indicated that these libraries were well over 99 % complete in representing all the possible four amino acid sequence combinations at positions 67-70. Both libraries were characterized for completeness as follows. Random individual clones were isolated from the libraries. Their

plasmid DNA and phage were purified using standard procedures. A BamHI/HindIII restriction digest of the sample plasmid DNA was used to monitor the total size of the ecotin phage display vector and the size of the DNA fragment containing the ecotin-gIII fusion. All the individual clone characterized contained full length insert. Double-strand nucleotide sequencing of the ecotin M84R+60X₄ library members revealed that 60% of the library members were variants. The heterologous nucleotide sequences within the correct reading frame at the designed positions from 67 to 70 encoded random amino acid residues without significant deviation from the expected frequency distributions. An immuno-blot assay of the library phage sample using rabbit anti-ecotin antibodies detected that ecotin was expressed on phage. Bovine trypsin activity inhibition assays of the phage library mixture further assured that the ecotin variants displayed on phage were still active. Selected ecotin variants were over-expressed, purified and characterized kinetically(see below).

Panning of the ecotin 60X₄ Library with Bovine and Rat Trypsin

Throughout this study, two criteria have been adopted to assess the effectiveness of a panning experiment. One was the total recovery of phage eluted from the plates. A rigorous panning protocol should allow the enrichment of strong binding clones over background in the course of iterative panning and amplification. The enrichment could be indirectly measured by comparing the phage recovery in the library panning experiment to the recovery in a control experiment where no protease was coated onto the polystyrene plate solid support. The number of phage recovered from background nonspecific binding with the plate and the blocking agent was less than 10^5 under our normal experimental conditions. A typical round of ecotin phage library panning with significant enrichment

would yield 10^6 to 10^8 phage from the elution, given that the input phage was in the range of 10^{10} to 10^{11} . Although an elevated recovery usually suggested an increase of positive clones in the pool of panning intermediates, other factors that were not directly related to the *in vitro* binding between ecotin and the immobilized protease might also lead to an artificially high recovery. These factors included the emergence of certain deletion phage variants that possessed a significant growth advantage in the liquid amplification procedure, or a sub population of the phage library that bound to the solid support or the blocking agent. To assure the *in vitro* selection was solely based on the strength of the interactions between ecotin phage and immobilized protease, another more stringent criteria has also been applied to evaluate the outcome of a panning experiment: the emergence of one or a few consensus sequences in the final population of the ecotin phage library. Ecotin variants encoded by these consensus sequences were then made and characterized kinetically for their affinities towards the immobilized protease used in the panning experiment.

The ecotin 60X4 library was panned against two serine proteases, bovine and rat trypsin, which were coated onto separate polystyrene plates. After the first round of panning, the elution fraction from the two plates were pooled and amplified separately on multiple LB/ampicillin plates to generate two intermediate libraries for the next round of panning. The two panning experiments were then carried out in parallel for the subsequent rounds. Careful precautions were taken to avoid cross-contamination between the bovine trypsin and rat trypsin panning. Four rounds of panning were completed before the final sequencing of the individual clones. For both ligands, there was significant increase of both total recovery of phage eluted and the percentage of phage recovered after each round (data not shown), suggesting the enrichment of phage clones carrying ecotin variants that bind to the two immobilized proteases. The individual clones from the final round of

bovine trypsin and rat trypsin panning were sequenced. The DNA sequencing results agreed extremely well with the predictions. Panning the ecotin 60A₄ library against bovine trypsin resulted in a heterologous population of sequences (data not shown). It confirmed that the impact of 60s loop side chains on binding was minimal under such conditions. On the other hand, as shown in Table 6.2, panning with the ecotin 60A₄ library against rat trypsin generated a clear consensus sequence W, G, F/L, P for positions 67-70. The strong selections for W and G at positions 67 and 68 were highly significant. At position 69, all 18 clones contained large, hydrophobic residues such as Phe, Leu, Ile, Tyr, Trp and Met. Phe and Leu were the most predominant residues of the consensus with five and four occurrences, respectively. Pro emerged from the library as a predominant residue at position 70. Not surprisingly, the wild-type sequence WGYD appeared three times, confirming that this sequence was also a good solution for rat trypsin binding. Another result that further substantiated the consensus sequence was that the selection operated at the amino acid level instead of nucleotide level. Since the library was encoded by a mixture of NNS codons, amino acids, such as G, L and P, could be encoded by two or three codons. Indeed, among the 18 sequences, we found different codons with comparable frequencies for G (12 GGG and 5 GGC), L (2 CTG, 2 CTC) and P (4 CCC and 4 CCG).

Table 6.2 The Consensus Sequence from 60X₄ Library Panning against Rat Trypsin

Position	67	68	69	70
	Y	G	F	I
	W	G	I	Q
	W	G	F	T
	W	G	L	P
	W	Q	L	P
	W	G	L	P
	W	G	W	G
	W	G	F	N
	R	G	Y	P
	W	G	F	S
	W	G	Y	D
	W	G	Y	D
	W	G	F	P
	W	G	Y	D
	W	G	L	W
	W	G	M	P
	W	G	M	P
	W	G	I	P
Consensus	W	G	F	P
Occurrence	16	17	5	8
(P _o -P _e)/σ	20.9	15.5	6.0	6.7
Consensus			L	
Occurrence			4	
(P _o -P _e)/σ			1.9	
WT	W	G	Y	D

Note: P_e, the expected frequency of possible NNS (N=A,C,G,T, S=C,G) codons. P_o, the observed frequency of codons in the clones sequenced, n, number of clones sequenced, $\sigma = [P_e(1-P_e)/n]^{1/2}$. (Braisted & Wells, 1996)

Inhibition Kinetics of Ecotin Variants Y69F+D70P and Y69L+D70P

To show that the panning experiments selected for strong binders, the variant ecotins encoding the consensus sequences WGFP and WGLP at positions 67-70 were cloned into the expression vector pTacTacEcotin. Ecotin Y69F+D70P and Y69L+D70P were expressed and purified. Their K_i 's against bovine, rat trypsins and uPA are listed in Table 6.3. The kinetic data shows that for both trypsins, ecotin Y69F+D70P and ecotin Y69L+D70P bound tighter than WT ecotin. Ecotin Y69F+D70P was the stronger binder with a lower pM K_i . The similar increase in affinity by ecotin Y69F+D70P for both trypsins suggested a general optimization of the hydrophobic packing at the secondary binding site. For uPA, ecotin Y69F+D70P bound better than WT and ecotin Y69L+D70P bound worse, implying that the inhibition against uPA was extremely sensitive to changes at the 60s loop and that the uPA-ecotin interface differed from the trypsin-ecotin interface at the secondary site. Clearly, high affinity ecotin secondary site variants can be selected from a phage display library.

Table 6.3 K_i of Ecotin Variants from Library 60X₄

	WT	Y69F+D70P	Y69L+D70P
Bovine Trypsin (unit : pM)	308	32	185
Rat Trypsin (unit : pM)	926	79	464
uPA (unit : μ M)	2.8	0.67	20.2

Liquid vs. Plate Amplification of Library M84R+60X₄ with uPA

The isolation of an inhibitor with improved binding towards rat trypsin suggested that this approach could be used to identify a high affinity inhibitor for uPA. To reduce the panning cycle time and simplify the experimental procedure, the phage amplification process was changed from plate amplification to liquid amplification. Although panning of the ecotin 60X₄ library was conducted with plate amplification, no serious consequences were expected in switching the amplification protocols. The ecotin M84R+60A₄ library was panned with uPA through liquid amplification for four rounds. However, there was no significant increase of phage recovery in each round of the experiment. When samples from the final round were characterized, a significant fraction of the clones were deletions of the full length ecotin phage clones. In fact, the outcome of panning process was very sensitive to both the fraction of initial positive clones in the library and the panning protocol. Due to their selective growth advantage, the spontaneously occurring deletion mutant phage grew faster in liquid culture, thereby gaining a significant advantage in competition with the full length ecotin phage. By iterative enrichment, the clones that carry the deletion sequences could be amplified at the early stage of library propagation and dominated the final population.

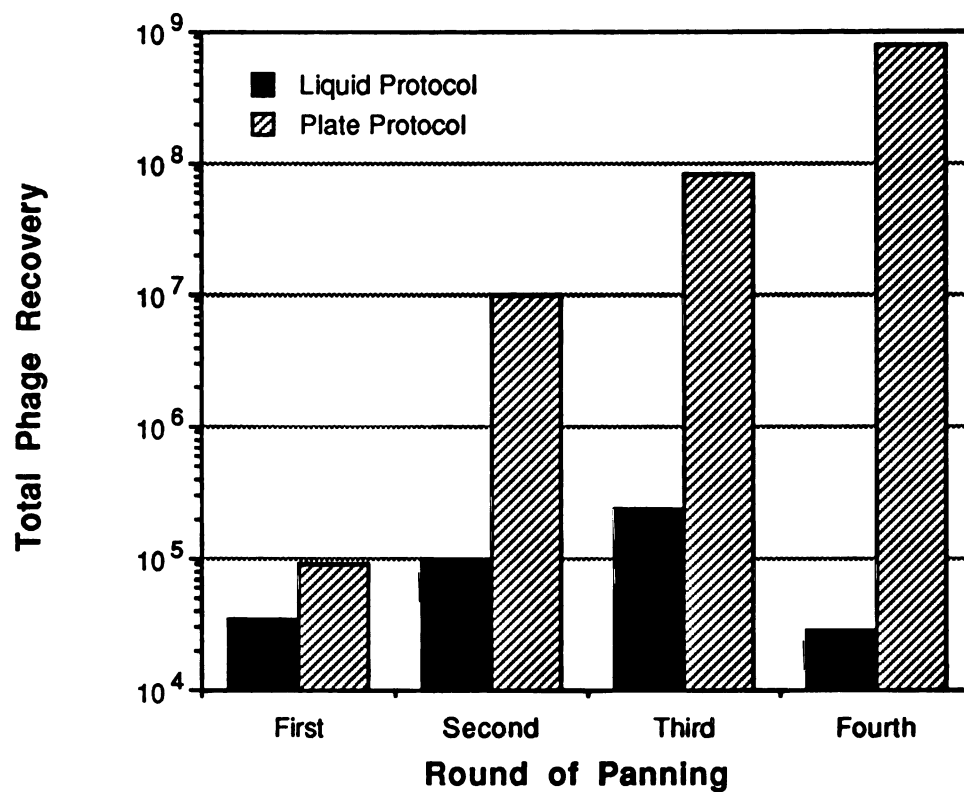
Since the amplification procedure was actually a critical factor for the success of the panning experiments, the plate amplification protocol was further refined to eliminate the nutrient selection pressure for phage growth in liquid culture. Extra care has been taken to develop a proper "nursing" protocol to preserve the small fraction of positive clones in the initial input library, especially at the early rounds of panning when the selection pressure had a disproportionally large impact. During amplification, *E. coli* colonies infected with phage were grown on multiple large LB/ampicillin plates with appropriate density (2.5 to

6.0×10^4) to ensure the separation of individual colonies. The same ecotin M84R+60A₄ library was panned with uPA again using the new method. The plate amplification significantly increased the number of phage recovered from the elution, indicating the enrichment of clones that bind to uPA. The comparison of the phage yields from the two methods is shown in Figure 6.1. By adopting this "nursing" procedure, we have not only obtained a final phage sample that was highly enriched over the background (10^7 to 10^8 in total vs. less than 10^5), but also isolated a clear consensus sequence (see the following section). These results strongly suggested that the conditions of the amplification protocol had a direct impact on the outcome of the *in vitro* selection process.

Figure 6.1 Comparison of Liquid and Plate Amplification

The total phage yields from the liquid and plate amplification procedures are compared. The panning experiment is conducted with ecotin M84R+60A₄ library panning against uPA for four rounds. The phage yield from liquid amplification round one to four are: 3.5×10^4 , 1.0×10^5 , 2.4×10^5 , and 2.9×10^4 , respectively. The phage yield from plate amplification round one to four are: 9.2×10^4 , 9.8×10^6 , 8.2×10^7 , and 8.0×10^8 , respectively.

Figure 6.1 Comparison of Phage Recovery from Liquid and Plate Amplification Protocols



Panning of the Ecotin M84R+60X₄ library with Rat Trypsin and uPA

The ecotin M84R+60A₄ library was panned against rat trypsin and uPA in parallel with intermediate plate amplification steps. Similar to the results from panning the ecotin 60A₄ library, a significant increase of phage recovery was observed in the eluates in each of the four rounds of panning for both ligands. Again, the panning results confirmed our observations from prior mutagenesis experiments. In rat trypsin binding, the dominant role of electrostatic interaction at Arg84 completely masked the impact of the 60s loop. Thus panning against rat trypsin did not generate a consensus sequence (data not shown), even though in the final round of panning, the phage recovery from the acid elution exceeded 10⁸. As anticipated, a consensus sequence was found in panning against uPA (Table 6.4). Furthermore, this consensus sequence WGY(R/P) was different from the previous consensus WG (F/L) P from panning the ecotin 60A₄ library against rat trypsin, suggesting that the *in vitro* selection process was indeed dependent on the nature of the immobilized protease. At positions 67 and 68, the same residues Trp and Gly were selected from the random sequence pool with high efficiency. At position 69, the wild-type residue Tyr occurred in nearly 75% of the sequences. The Asp at position 70 was replaced with Arg (eight times) and Pro (four times). This was the only residue that differed from the wild-type sequence. Again, the nucleotide sequences of the 18 samples from uPA panning showed a mixture of codons encoding the selected amino acids such as Gly (14 GGG and 3 GGC), Pro (3 CCG and 1 CCC) and Arg (5 AGG and 3 CGG), further strengthening the conclusion that these residues were selected based on their contributions to increase the affinity towards uPA.

Table 6.4 The Consensus Sequence from M84R+60X₄ Library Panning against uPA

Position	67	68	69	70	Ki (unit :pM)
	W	G	Y	D	
	W	G	H	R	
	M	G	Y	P	
	W	G	Y	R	
	Y	G	Y	R	
	W	G	H	R	
	W	E	F	P	
	W	G	N	R	
	W	G	Y	R	
	W	G	Y	Q	
	N	G	Y	R	
	W	G	Y	P	
	W	G	F	G	
	W	G	Y	P	
	W	G	Y	P	
	F	G	Y	K	
	W	G	Y	G	
	W	G	Y	W	
Consensus	W	G	Y	R	50
Occurrence	14	17	13	8	
(P _o -P _e)/σ	18.2	15.5	16.8	5.1	
Consensus				P	80
Occurrence				4	
(P _o -P _e)/σ				2.8	
M84R	W	G	Y	D	3600

Inhibition Kinetics of Ecotin M84R+D70R and Ecotin M84R+D70P

Ecotin M84R+D70R and ecotin M84R+D70P were constructed by cloning the specific variant sequences from the phage clone into expression vector pTacTacEcotin, taking advantage of a pair of common restriction sites BamHI/AatII that covered 90% of the ecotin gene. The two variants were purified to homogeneity via reverse phase HPLC. Their K_i 's against uPA were determined and are listed in Table 6.4. The K_i of ecotin M84R+D70R was lower than that of ecotin M84R+D70P, mirroring their relative occurrence in the consensus. The D70R substitution alone reduced the K_i 72-fold, resulting in a 50 pM inhibitor of uPA, the tightest binding inhibitor of uPA reported up to date. The one amino acid substitution at position 70 generated a significant impact on the affinity of these two ecotin variants.

The increased potency and specificity of ecotin M84R+D70R validates the strategy to optimize ecotin's affinity towards uPA through a stepwise approach. The inhibition of rat trypsin can be used as a benchmark to assess the effectiveness of this methodology. Ecotin M84R+D70R, with a K_i against rat trypsin of 220 pM, was a better inhibitor of uPA than rat trypsin. This result was in sharp contrast with the preference of WT ecotin for rat trypsin over uPA by over 3,000-fold. Table 6.5 summarizes the ecotin K_i 's towards rat trypsin and uPA generated through the site-directed mutagenesis and phage display experiments. The specificity of ecotin has been successfully converted from one serine protease to the other with a significant increase in potency at the same time. The overall specificity improvement was 13680-fold.

Table 6.5 Designing Specificity & Potency of Ecotin

		WT	M84R	M84R+D70R
Ki (Rat Trypsin)	(nM)	0.93	0.31	0.22
Ki (uPA)	(nM)	2800	3.6	0.05
Ki (uPA)/Ki (Rat Trypsin)		3010	12	0.22
Improvement Value vs. WT ecotin		1	260	13680

Note : Improvement value is calculated by dividing the ratio of Ki (uPA) / Ki (Rat Trypsin) of ecotin wild type (WT) by that of ecotin M84R or M84R+D70R.

Discussion

The consensus sequences generated from rat trypsin and uPA panning revealed that residues 67 and 68 of ecotin were essential in high affinity binding against the target serine protease. In the tetrameric complex of ecotin-trypsin, Trp67 of ecotin provided the majority of the hydrophobic interactions at the binding interface. In fact, this tryptophan side chain was within direct van der Waals distance of residues 91, 233, 234, 236, 237 and 240 (chymotrypsin numbering system) of the protease. In all three proteases both 234 and 237 were aromatic side chains. Residue 234 is either Tyr (rat and bovine trypsin) or Phe (uPA); residue 237 is a conserved Trp. The stacking of aromatic rings between ecotin and the protease provides a tightly-packed hydrophobic interface. Due to tryptophan's large side chain volume and buried surface area, substituting Trp67 of ecotin to any other amino acid might create an unfilled cavity that destabilizes the ecotin-protease complex. Thus Trp67 appears to be an integral part of the hydrophobic "core" of the secondary binding site. On the other hand, Gly68 was also selected from both panning experiments. This residue probably plays a more structural role to maintain the proper flexibility and main chain conformation of the 60s loop. Since Trp67 and Gly68 were conserved at 60s loop, the other two residues, Tyr69 and Asp70 were the only candidates to provide differential recognition towards target proteases.

The ecotin consensus sequence WG (F/L) P at positions 67-70 that resulted when the ecotin library was panned against rat trypsin represented an overall improvement of the hydrophobicity at positions 69 and 70. Both Phe and Leu are more apolar than Tyr, which has a hydroxyl group. In addition to its hydrophobicity, proline was commonly found at various types of turns to lock the surface loops in stable conformations. Pro70 was presumably selected for this reason. The ten-fold increase of affinity of ecotin

Y69F+D90P towards both bovine and rat trypsin illustrated the modest improvement of side chain packing in the vicinity of amino acids 69 and 70. In the case of uPA inhibition, the charge reversion substitution of D70R suggests that electrostatic interaction could also be important for this position. In the crystal structure of the trypsin-ecotin complex, Asp70 of ecotin is very close to Asn93 of rat trypsin. In fact, the O δ of Asp70 is within hydrogen bonding distance (2.94 Å) of the N δ of Asn93. Although the crystal structure of the ecotin-uPA complex is not yet available, it is possible to build a complex model based on the high resolution structures of rat trypsin-ecotin complex and the catalytic domain of human uPA. The recently published 2.5 Å structure of the uPA protease domain has a similar topology to trypsin (Spraggon *et al.*, 1995). By superimposing the catalytic triad of uPA and rat trypsin, a model of the ecotin-uPA complex was constructed. The back bone main-chain conformation of uPA overlaid with that of rat trypsin very well. Most of the differences occurred in several surface loops. However, the region near the secondary binding site and close to position 93 of the protease were highly conserved (Figure 6.2). Residue 70 of ecotin is close to residue 93 of uPA in the model of the ecotin-uPA complex. Under such a context, the results of the uPA panning are easily interpreted. In uPA, residue 93 was an Asp instead of Asn. The consensus sequence Arg70 from the ecotin M84R+60A₄ library panning against uPA was indeed an ideal choice to provide a counter charge to stabilize Asp93 by forming a salt bridge between these two side chains. Thus, we have not only shown affinity improvement through the creation of alternative intermolecular electrostatic interaction, but also refined the structural model of the ecotin-uPA complex. Since D70R decreased the K_i of ecotin variant 72-fold from 3.6 nM to 50 pM, this favorable charge substitution alone was responsible for about 2.5 Kcal/mol (RT ln72, where T=298 K) of binding energy improvement.

**Figure 6.2 Modeling Electrostatic Interactions
at the 60s Loop of Ecotin**

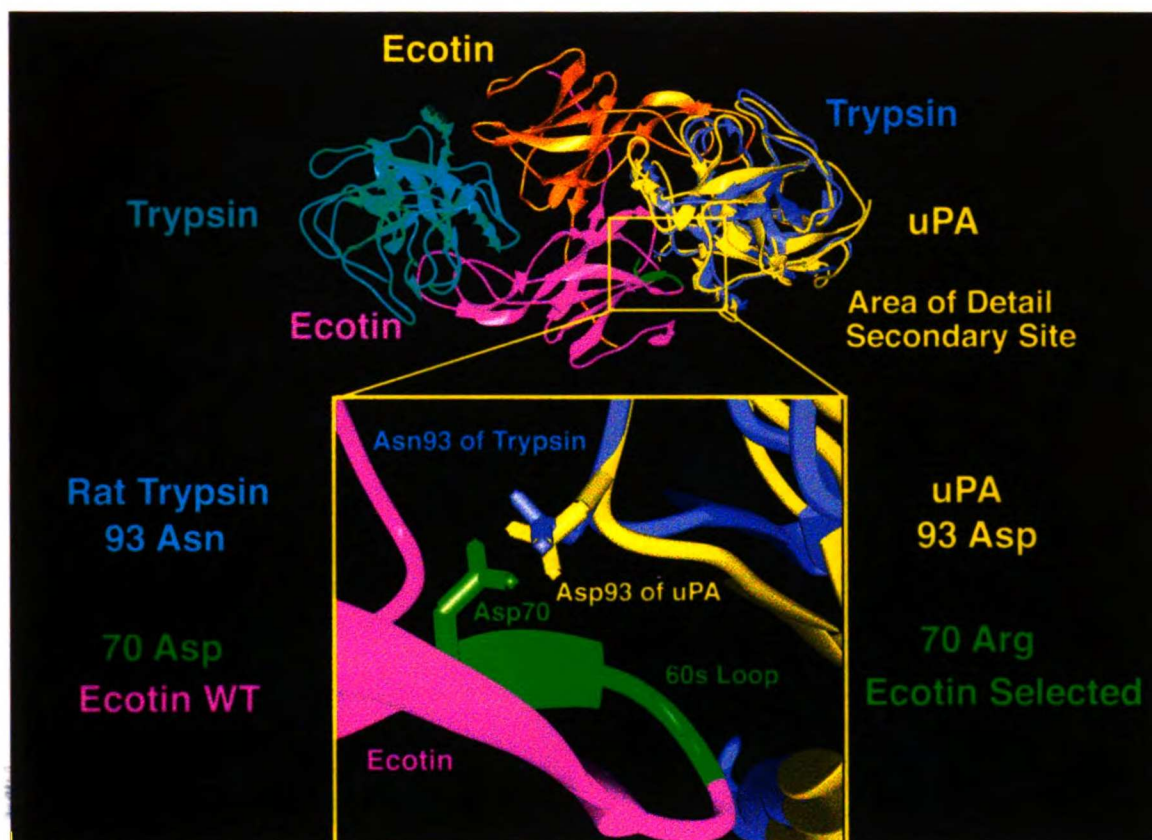


Figure 6.2 Modeling Electrostatic Interaction at the 60s Loop of Ecotin

The crystal structure of uPA was superimposed to one trypsin molecule in the ecotin-trypsin tetrameric complex by matching the residues Ser195, His57, Asp102, and Asp189 of each enzyme using the program MidasPlus (Computer Graphic Laboratory, UCSF) with RMS deviation of 0.66 Å.

The agreement between the results from the multiple alanine substitution experiments and the phage display experiments suggests that mutagenesis data of the 60s loop can be used to design ecotin phage libraries and predict the possibility of a consensus sequence. To generalize these findings and extend the scope of our efforts to other parts of the secondary site, the impact of a particular surface loop under different contexts was determined using the expression of free energy difference ΔG (Wells, 1990). The K_i 's of all ecotin variants were converted to free energy terms with the equation $\Delta G = -RT \ln K$. The impact of the 60s loop when the P1 residue is an Arg can be calculated as the $\Delta\Delta G = \Delta G_{(M84R+60A4)} - \Delta G_{(M84R)}$. Based on our data in the previous study, the $\Delta\Delta G$ values of the 60s and/or 100s loop in the presence of the M84R substitution were calculated and listed in Table 6.6. A high value of $\Delta\Delta G$ will be a good indication that the specific surface loop is a key contributor to protease recognition. The outcome of the ecotin phage display experiment is summarized in Table 6.7 for the comparison between the free energy difference and the *in vitro* selection process. For the ecotin 60A4 library against rat trypsin, the $\Delta\Delta G$ was 5.02 Kcal/mol. For the ecotin M84R+60A4 library against uPA, the $\Delta\Delta G$ was 3.56 Kcal/mol. In both library selection processes, the emergence of a consensus sequence was associated with a high $\Delta\Delta G$ value of the corresponding mutations. Therefore, for uPA binding, we predict that the 100s loop alone or the 100s loop combined with the 60s loop will have a major impact on the strength of ecotin-uPA complex formation. Panning a phage display library at these regions will likely generate novel

consensus sequences encoding ecotin variants with high affinity towards uPA. The multiple alanine substitution experiments established a rigorous framework from which future combinatorial selection or screening approaches might be rationally designed and developed. Combining the defined control of region-specific mutagenesis and the comprehensive scope of phage display, this two step affinity optimization procedure has the potential to be applied to other systems of macromolecular recognition as well.

Table 6.6 Ecotin Variants $\Delta\Delta G$ (Kcal /mol)

$\Delta\Delta G$	Bovine Trypsin	Rat Trypsin	uPA
ΔG (60A ₄) - ΔG (WT)	0.10	-5.02	-3.16
ΔG (M84R+60A ₄) - ΔG (M84R)	-0.26	0.44	-3.56
ΔG (M84R+100A ₄) - ΔG (M84R)	0.65	1.33	-4.16
ΔG (M84R+ 60A ₄ +100A ₄) - ΔG (M84R)	-0.34	-0.16	-5.31

Note: ΔG was calculated by converting K_i to $RT\ln(K_i)$, where $T=298$ K (25 °C). The relative free energy was the ΔG difference between corresponding variants. According to this scale, a 100-fold difference in K_i will be equal to -2.73 Kcal/mol.

Table 6.7 Summary of Ecotin Phage Library Panning Results

Library	Bovine Trypsin	Rat Trypsin	uPA
Ecotin 60A ₄	High Recovery No Consensus	High Recovery Consensus	--
Ecotin M84R+60A ₄	--	High Recovery No Consensus	High Recovery Consensus

The highly differentiated and specialized roles of the mammalian serine proteases demand inhibitors that interact with their target enzymes with strong selectivity. Both

conventional small molecule and macromolecular inhibitor design approaches are constrained by the distinctive structural features at the proximity of the active site and binding pocket of the target enzyme. This limitation is partially due to the historical fact that the most of our knowledge on serine proteases have been based on the reactive mechanism and substrate specificity studies of the model enzymes such as trypsin and chymotrypsin. However, under physiological conditions, many mammalian serine proteases have actually taken advantages of a large array of cofactors to develop sophisticated mechanisms to control and regulate the potency and specificity of precise proteolytic events. For example, thrombin, a key enzyme that cleavages fibrinogen and forms fibrin clots in the blood coagulation pathway, is regulated by heparin, α_2 -macroglobulin, antithrombin III, thrombinmodulin (Stubbs & Bode, 1995) and monovalent ions such as Na^+ (Dang & Di Cera, 1996). Through several cofactor interactions that are distal to the active site, thrombin achieves a high level of fine tuning and balance between its coagulatory and anticoagulatory activities in an intertwined web of biological pathways in haemostasis, platelet aggregation, tissue remodeling, mitosis and chemotaxis. In the case of uPA, the serine protease domain can form a high-affinity complex with several key partners such as PAI-1, PAI-2, protease nexin-1, and α_2 -macroglobulin receptor (Andreasen *et al.*, 1994; Fazioli & Blasi, 1994). These examples illustrate the variety of structural features that are involved in the macromolecular recognition between serine proteases and other macromolecules *in vivo*. It also suggests unexploited opportunities of using other protease surface regions as the basis of inhibitor design and engineering.

Ecotin offers such a platform to investigate and utilize the contribution from the distal binding region for protease inhibition. Ecotin has special structural features for innovative methods of inhibitor design and engineering. By modulating the amino acid residues at the 60s loop, we have gained another level of control in designing the specificity and potency

of ecotin variants. The secondary binding site of ecotin not only facilitated the fine-tuning of the molecular recognition towards many known homologous enzymes, but also provided additional side chain conformational flexibility to accommodate other serine proteases with similar scaffolds. These experimental results demonstrated that ecotin had several distinct advantages as a generic starting point for inhibitor design. The first advantage is the tremendous electrostatic and hydrophobic surface diversity embedded in the contact regions between ecotin dimer and two protease molecules. Secondly, the combinatorial approach of phage display makes it feasible and highly efficient to search and sort the large repertoires of ecotin surface loop variants. Finally, the crystal structures of ecotin-protease complexes can serve as a framework for designing inhibitors against enzymes with unknown structures. By taking advantage of ecotin's unique secondary binding site, it is possible to improve the affinity against a target protease through stepwise optimization of various surface loops involved in the binding interaction. This approach will be especially powerful for enzymes that share the same primary substrate specificity.

A combination of site-directed mutagenesis and phage display approaches were taken to study the interactions between ecotin and several serine proteases. The secondary site of ecotin plays critical roles for certain proteases. Phage libraries were then made at these surface loops and used for panning against the target proteases. Two distinctive consensus sequences were identified from panning of phage libraries of ecotin variants at the secondary binding site against rat trypsin and uPA. In both cases, the consensus sequence encoded ecotin variants with higher affinity for the target protease. This study validated a general strategy to engineer inhibition potency and specificity of a macromolecular serine protease inhibitor by modulating various components of the network of interactions between the inhibitor and the protease.

Materials and Methods

Materials

Enzymes and reagents for the manipulation of DNA were purchased from Promega (Madison, WI) or New England Biolabs (Beverly, MA) and were used following the manufacturer's instructions. The *E. coli* strain JM101, XL-1 Blue F' and the VCSM13 helper phage were from Stratagene (La Jolla, CA). The *E. coli* ecotin gene deletion strain IM Δ ecoJ was derived from JM101 and was the generous gift of Dr. Iain Murray (Murray and Craik, unpublished results). Low molecular weight uPA (LMuPA) was obtained from American Diagnostica (Greenwich, CT). Rat trypsin was expressed in *E. coli* using the expression vector pZ3 and purified as described. Bovine trypsin was from Sigma (St. Louis, MO). α -³⁵S-dATP was from DuPont NEN (Boston, MA). Sequenase Version 2.0 sequencing kit was from U.S. Biochemical Corp (Cleveland, OH). GeneClean[®] was from Bio101, Inc (La Jolla, CA). Oligonucleotides were synthesized with an Applied Biosystems 391 DNA synthesizer (Foster City, CA). Falcon polystyrene petri dishes were from Becton Dickinson Labware (Lincoln Park, NJ). All other chemicals were of reagent grade or better and were used without further purification.

Plasmid and Library Constructions

Mutagenesis was performed by the method of Kunkel (Kunkel, 1985). All mutations have been confirmed at the DNA level by sequencing. The vector pBSeco-gIII was used to construct phage libraries 60X₄ and M84R+60X₄. A deletion and frameshift mutation was introduced at residues 67-70 of ecotin by primer 5'-C AAA ACG CTG GAA GG TAT TAT GTC TTT GAT-3' to make pBS eco-gIII Δ 60. This construct was used as template to make library 60X₄ by primer 5'-AAC AAA ACG CTG GAA GGC NNS NNS NNS NNS

TAT TAT GTC TTT GAT AAA GTC AG-3' (N=A/C/G/T, S=C/G). Primer 5'-GT TCC CCG GTT AGT ACT AGG ATG GCC TGC C-3' [unique *ScaI* site] was used to introduce an M84R mutation in pBSeco-gIII Δ 60 to generate pBSeco-gIII M84R Δ 60. This vector was then used to generate library M84R+60X₄ by the same 60s loop library primer 5'-AAC AAA ACG CTG GAA GGC *NNS NNS NNS NNS* TAT TAT GTC TTT GAT AAA GTC AG-3' (N=A/C/G/T, S=C/G). Both libraries of ecotin phage had four positions 67-70 randomized. The ecotin phage display vector pBSeco-gIII and expression vector pTacTacEcotin were mutated to carry an *AatII* site by primer 5'CA GAC AAT GTA GAC GTC AAG TAC CGC GTC-3' at amino acid 125 of ecotin to facilitate the cloning between the two vectors. All ecotin variants obtained from panning experiments could be directly cloned into the expression vector pTacTacEcotin. For library phage preparation, 10 μ g of uracil-laden template was used for Kunkel mutagenesis in a total synthesis reaction of 1000 μ l. The mutagenesis reaction mixture was purified by GeneClean, redissolved in water, electroporated into F' XL-1 Blue cells, and grow in 100 ml of 2YT/Ampicillin for 1 to 2 hours. The culture was then divided into two portions. One portion was transferred into fresh 2YT/Ampicillin medium and grown for 8-12 hours. The cells were harvested by centrifugation and the double strand plasmid was prepared using Promega Midiprep Kit. This DNA sample was the ecotin phage library stock. The other portion of the culture was subsequently diluted to an OD₆₀₀ = 0.25, and then infected with VCSM13 helper phage. This infected culture was grown for 6 hours at 37 °C with rigorous shaking, and the phage were harvested as described in the section of ecotin phage preparation below.

Ecotin Mutagenesis, Expression and Purification

Ecotin and ecotin variants were expressed and purified as described previously (Chapter Five).

Determination of Ecotin Ki Values against Target Proteases

Ecotin and ecotin variants Ki values against bovine, rat trypsin and uPA were determined as described previously (Chapter Five).

Ecotin Phage Preparation

For the preparation of pBSeco-gIII ecotin bacteriophage, plasmid DNAs were transformed into a male strain (F') of IM Δ ecoJ. A single colony selected on ampicillin plates was grown in 3 ml 2YT medium containing 60 μ g/ml ampicillin at 37 °C for 8 hours. The culture was diluted into 100 ml of 2YT/ampicillin, grown to OD₆₀₀ = 0.25, and infected with the helper phage VCSM13 at a multiplicity of infection of approximately 100 helper phage per cell. The infected culture was allowed to grow at 37 °C with shaking for approximately 6 hours. Phage particles were harvested by precipitation with one fifth volume of 20 % polyethylene glycol 8000, 2.5 M NaCl at 4 °C overnight, centrifugation at 6000 g for 40 min, and resuspended in 1 ml TE buffer. Phage stocks were stored at 4 °C for up to six months. Phage titers typically ranged from 10¹⁰ to 10¹¹ cfu/ μ l culture and were stable within six months.

Ecotin Library Panning

Polystyrene petri dishes (35 mm, Falcon 1008) were coated with 1 ml of 10 μ g/ml bovine trypsin, or rat trypsin, or LMuPA in PBS (137 mM NaCl/2.7 mM KCl/10 mM Na₂HPO₄/1.8 mM K₂HPO₄ pH 7.5) overnight at 4 °C, and excess binding sites were blocked with 5% non-fat dry milk PBS solution for 2 hours. Phage were added to the dishes in buffer containing 1 ml PBS/0.5% Tween 20 and were incubated for 2 to 24 hours with gentle agitation at 4 °C. Solutions containing the phage were then removed and the

dishes were washed 9 times with 5 ml PBS/0.5% Tween 20. Bound phage were serially eluted by incubation with 900 μ l of 0.1 N HCl/Glycine solution (pH 2.2) with gentle shaking for 15 minutes at room temperature. Three elutions were performed. The eluates were neutralized with 167 μ l 1 M Tris•HCl, pH 8.8.

Ecotin Library Amplification

Equal volumes of three eluates were pooled for subsequent amplification and characterization. Two amplification protocols were used in this experiment. In the liquid amplification procedure, 900 μ l of phage elution pool was incubated with 9 ml of fresh grown IM Δ ecoJ lawn cells for 30 min at 37 °C with gentle shaking. The infected lawn cells were transferred to 200 ml of 2YT/ampicillin liquid medium and continued to grow to an OD₆₀₀ = 0.25, and infected with VCSM13 helper phage. The infected culture was grown for 6 to 10 hours at 37 °C with shaking, then harvested and precipitated as described above. In the plate amplification procedure, 9 ml of fresh grown IM Δ ecoJ lawn cells was incubated with 900 μ l of phage elution pool at 37 °C for 30 min with gentle shaking. The cells were then plated on multiple LB/ampicillin plates (polystyrene petri dishes, 150 mm, Falcon 3025) and grown for 6 to 10 hours. Cells from these plates were recovered by soaking each plate with 5 ml of 2YT and scraping the cells into the medium. The 2YT medium containing cells from plates were collected in 100 ml of 2YT/ampicillin and continue to grow for one to two hours. This cell culture was then diluted in 100 ml 2YT/ampicillin to an OD₆₀₀ = 0.25, and infected with VCSM13 helper phage. The infected culture was grown for 6 hours at 37 °C with shaking, and the phage were harvested as described above.

Chapter Seven: Future Directions

Introduction

The establishment of trypsin and ecotin phage display systems have paved the way for detailed analysis and engineering of macromolecular recognition between serine proteases and their macromolecular inhibitors. Compared with previous methods such as conventional genetics or single-site mutagenesis, the coupling of structure-based site-directed mutagenesis and the combinatorial approach of phage display make the search for novel functions or altered specificities from diverse molecular repertoires more efficient, thorough, and focused. The current study has started to address some of the fundamental problems of protein-protein interactions and revealed a sophisticated and novel network of interactions in ecotin-protease binding (Chapters Five and Six). Both experimental systems can be further developed and improved. Since the main technical obstacles have been removed, many questions regarding substrate and inhibitory specificities are waiting to be explored and investigated. The following is a brief discussion on the potential of trypsin and ecotin phage display systems.

Discussion

Trypsin Phage Display

Overcome the Deletion Problem

A large random variant phage library and a specific *in vitro* selection protocol are essential for effective phage display. Both prerequisites have been met in the current trypsin phage display system. The only remaining technical difficulty is the spontaneous deletion of the full length phage vector. In some amplification experiments, up to 80 % of the trypsin phage clones carried deletions to different extents after a single round of amplification. The majority of these deletions occurred near or within the trypsin gene-pIII fusion as demonstrated by restriction analysis. The insert instability of the trypsin phage display vector is an intrinsic characteristic of M13-based vectors (Zinder & Boeke, 1982). This problem can not be overcome by using a different bacteria strain (for example, RecA⁻ strains). The deletion of the trypsin phage vector greatly reduces the yield of the full length phage clones after amplification, and thus makes it more difficult to obtain a clear consensus sequence from the background of non-specific binding phage. In addition, the impact of vector deletion is much greater when the enrichment ratio of the protease inhibitor ligand is low, as in the case of soybean trypsin inhibitor (STI) coupled to oxirane beads. If the single round enrichment ratio of the solid matrix is on the order of over 500 to 1,000 for wild-type trypsin phage, the efficiency of a particular panning procedure may compensate for the deletions and keep the positive clones within the amplified pools of phage libraries.

To solve the deletion problem, several approaches can be taken. A better solid matrix-ligand combination that enriches active trypsin phage over background for more than several hundred fold in a single round of panning is the most desirable solution. The considerable technical difficulties to mutate, express, and purify soybean trypsin inhibitor (STI) variants in the absence of a clone of the STI gene also call for a ligand that can be more easily modified recombinantly or chemically. The search for the optimal ligand and solid support will remain a major theme in the next stage of trypsin phage development. Another solution is to use a phagemid derived vector for trypsin phage display. Demonstrated by the success of ecotin phage display, phagemid vectors are less prone to spontaneous deletions during iterative amplification and propagation due to their smaller sizes. However, the construction of a reliable trypsin phagemid vector depends on trial and error testing since there are few theoretical principles to guide the vector design efforts. Last, but not least, although the PCR-base amplification protocol presented in Chapter Four is laborious and time consuming, it is still a successful alternative. The trade-off can be justified if significant results are expected through such a brute force approach.

Design Larger and More Diverse Trypsin Libraries

The three residue library at the binding pocket of trypsin is the first generation library and was used to establish an experimental procedure for *in vitro* selection of active protease variants. The number of residues and the location of the amino acid substitutions can be expanded. Structure-based mutagenesis suggested that residues on the peripheral surface loops of the trypsin binding pockets are also crucial determinants of substrate specificity of trypsin (Hedstrom *et al.*, 1994; Hedstrom *et al.*, 1992). Complete four residues libraries or partially complete five or six residue libraries can be constructed at these peripheral surface loops. The modification of these residues may provide more conformational flexibility and diversity accommodating novel or altered substrate specificities for trypsin. Once the optimal *in vitro* selection and amplification procedures are established, multiply-variant

trypsin phage libraries can be constructed. Furthermore, the libraries can be generated through other means such as error-prone PCR (Leung *et al.*, 1989) or sexual PCR (Stemmer, 1994a; Stemmer, 1994b), which introduces substitutions that are not limited to a particular structural or functional region. The enormous diversity of these complex libraries will facilitate the isolation and identification of trypsin variants with novel properties.

Display Other Trypsin-like Proteases on Phage

Rat anionic trypsin is the first protease displayed on phage. Although there is no general rule to predict the possibility of displaying other chymotrypsin family serine proteases on phage, the current experimental results are encouraging. Displaying other chymotrypsin family serine proteases with vital biological functions and therapeutic potentials, such as Hepatitis C Virus (HCV) protease (Matthews *et al.*, 1994), and characterizing the protease variant repertoires through *in vitro* selection will expand our knowledge and understanding of the structural and functional relationships of these important enzymes. In particular, studying serine proteases in such a system may provide new insights on the determinants of substrate specificity of this family of enzymes. The rules that emerge from these studies may be general enough to be applied to other enzymes and macromolecules as well. The ultimate goal of this work is to generate protease variants with desired novel specificities, to allow for more intelligent predictions in the *de novo* design of a macromolecule.

Ecotin Phage Display

Phage display of ecotin and the panning of two ecotin secondary site libraries revealed the novel structure function relationship of this dimeric serine protease inhibitor. The ease of mutagenesis, expression and purification of ecotin variants to large quantities for kinetic

and structural studies facilitates the systematic probing of the structural basis of ecotin's broad specificity. All the protocols of ecotin phage display and panning have been rigorously tested and validated through successful panning of the two 60s loop phage libraries (Chapter Six). The multiple alanine substitution study of ecotin's surface loops, together with ecotin phage display experiments, has laid down a well-structured framework for further investigation.

The 100s Loop Libraries at the Secondary Site

Multiple alanine substitutions at residues 108, 110, 112 and 113 show that these four positions play important roles in recognition and binding to the target protease uPA. Kinetic analysis of ecotin 100A₄ and M84R+100A₄ against uPA suggests that ecotin variants with potency and specificity can be isolated through panning of a random library at 100s loop alone or combined with the 60s loop. Further characterization of these 100s loop variants both kinetically and structurally can offer new insight into the nature and strength of the interactions between ecotin and the target protease.

The 50s and 80s Loop Libraries at the Primary Site

Similar phage library panning experiments can be conducted at the primary binding site by dissecting binding interactions between the 50s and 80s loops. Ecotin variant libraries were designed and constructed at the 80s loop. Panning of this 80s loop library against uPA resulted in consensus sequences encoding variants with improved binding towards uPA (Takeuchi and Craik, unpublished results). Based on the same principle, the 50s loop library can be designed and prepared to identify key residues among positions 52 to 54 of this surface loop.

The Loop Combination Libraries

Individual loop libraries will be instrumental in establishing the key residues within the primary and secondary site surface loops. The panning results from the 60s loop library suggest that only a subset of these residues is critical. Combinatorial libraries can be then designed to include the key residues from different surface loops. By panning these combined loop libraries, we can address whether binding interactions from different surface loops are additive under different contexts. By expanding the randomized region to four surface loops at the ecotin-protease interface, a much larger conformational space can be effectively searched to isolate potent inhibitors against uPA as well as other target enzymes with trypsin-like scaffolds. Ecotin's extraordinary binding interface will be engineered and fine-tuned to accommodate proteases with distinct surface electrostatic and hydrophobic characteristics.

Ecotin Loop Trimming and Remodeling

The current ecotin phage library panning strategy does not change the length or the conformation of the secondary surface loops. For some serine proteases, one or a few ecotin surface loops may have steric hindrance for binding. Trimming or remodeling these loops by transplanting a different loop from an existing protein structural scaffold can introduce additional conformational diversity at the ecotin-protease binding interface. Computer modeling techniques have been developed to model the conformation of loop grafting using a loop database extracted from the known crystal structures of proteins (Ring *et al.*, 1992). Using this approach, the 62-71 loop of ecotin was replaced by the residues of 19-22 of *Staphylococcal* nuclease, and the variant was capable of inhibiting trypsin at nanomolar concentration (Murray, Gillmor, Fletterick and Craik, unpublished results). This result indicates that loop remodeling is a valid approach for inhibitor design.

Ecotin Minimization

A recent development of the phage display technique is to use the *in vitro* selection approach to minimize the protein scaffold while preserving the key functional epitopes and the activities of the macromolecule of interest (Braisted & Wells, 1996; Li *et al.*, 1995). A stepwise approach was taken to reduce the size of a polypeptide hormone, atrial natriuretic peptide (ANP), from 28 residues to 15 while retaining high biopotency (Li *et al.*, 1995). Ecotin is an excellent candidate to develop a minimization strategy for serine protease inhibitors. Although the dimeric nature of the ecotin-protease interaction presents an additional challenge for this approach, the high-resolution three-dimensional structure of ecotin-trypsin complex can be used to guide the initial minimization design efforts.

Obligate Ecotin Monomer

Ecotin's dimer interface is a central element in the network of interactions between ecotin and its target protease. Repeated efforts to design an obligate ecotin monomer through C-terminal truncation and remodeling were unsuccessful. More radical changes may be introduced at this region to create a tailless version of an ecotin molecule. An obligate monomer ecotin molecule would be a useful tool to dissect and quantify the contributions of the primary and secondary binding sites.

Kinetic Characterizations of Monomer-Dimer Equilibria

Understanding ecotin's monomer-dimer equilibria will have a significant impact on protein engineering of ecotin. Detailed kinetic analysis of this equilibrium in the presence and absence of target proteases is critical for redesigning ecotin's large binding interface. Kinetic rate constants such as *k_{on}* and *k_{off}* of the multi-step ecotin-protease complex forming pathway will define the oligomerization state of ecotin and the target protease quantitatively. The kinetic characterization will also facilitate the inhibitor design efforts by understanding the impact of the mutation of interest on the rate of association and dissociation between ecotin and the target protease.

Structural Characterization of Ecotin Variants

X-ray crystallography and NMR spectroscopy can be used to obtain the high-resolution three-dimensional structure of the complex between a target protease and an ecotin variant. The consequences of radical loop remodeling and redesigning can be directly studied by comparing the structure of the WT ecotin with that of the variant in complex with different serine proteases. Investigating the flexibility and rigidity of the ecotin surface loops will be useful in understanding the conformational diversity of ecotin-protease interfaces and deriving general rules to predict and design specific changes.

In summary, ecotin is a versatile platform for testing novel hypothesis and approaches of protein design and engineering. Its potential as a specific and potent serine protease inhibitor against critical enzymes in cancer metastasis also demands a thorough understanding of the structural and functional relationships of this unique serine protease inhibitor. The dual approaches of structural-based, site-directed mutagenesis and phage display of combinatorial libraries can continue to uncover fundamental principles of macromolecular recognition and binding specificity through the study of the network of interactions between ecotin and serine proteases.

References

- Allen, J. B., Walberg, M. W., Edwards, M. C. & Elledge, S. J. (1995). Finding prospective partners in the library: the two-hybrid system and phage display find a match. *Trends Biochem Sci* 20(12), 511-6.
- Andreasen, P. A., Sottrup-Jensen, L., Kjoller, L., Nykjaer, A., Moestrup, S. K., Petersen, C. M. & Gliemann, J. (1994). Receptor-mediated endocytosis of plasminogen activators and activator/inhibitor complexes. *FEBS Lett* 338(3), 239-45.
- Aoyagi, T., Miyata, S., Nanbo, M., Kojima, F. & Matsuzaki, M. (1969). Biological activities of leupeptins. *J Antibiot (Tokyo)* 22(11), 558-68.
- Baca, M., Presta, L. G., O'Connor, S. J. & Wells, J. A. (1997). Antibody humanization using monovalent phage display. *J Biol Chem* 272(16), 10678-84.
- Barbas, C. d., Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991). Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc Natl Acad Sci U S A* 88(18), 7978-82.
- Barbas, C. d., Rosenblum, J. S. & Lerner, R. A. (1993). Direct selection of antibodies that coordinate metals from semisynthetic combinatorial libraries. *Proc Natl Acad Sci U S A* 90(14), 6385-9.
- Bartel, D. P. & Szostak, J. W. (1993). Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* 261(5127), 1411-8.
- Bass, S., Greene, R. & Wells, J. A. (1990). Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* 8(4), 309-14.
- Bender, M. L. & Killheffer, J. V. (1973). Chymotrypsins. *Crc Crit Rev Biochem* 1(2), 149-99.
- Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F. & Gething, M. J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell* 75(4), 717-28.
- Blow, D. M., Janin, J. & Sweet, R. M. (1974). Mode of action of soybean trypsin inhibitor (Kunitz) as a model for specific protein-protein interactions. *Nature* 249(452), 54-7.
- Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355(6360), 564-6.

- Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 204(2), 433-51.
- Bode, W., Schwager, P. & Huber, R. (1978). The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9 Å resolution. *J Mol Biol* 118(1), 99-112.
- Bone, R., Fujishige, A., Kettner, C. A. & Agard, D. A. (1991). Structural basis for broad specificity in alpha-lytic protease mutants. *Biochemistry* 30(43), 10388-98.
- Bone, R., Silen, J. L. & Agard, D. A. (1989). Structural plasticity broadens the specificity of an engineered protease. *Nature* 339(6221), 191-5.
- Botstein, D. & Shortle, D. (1985). Strategies and applications of in vitro mutagenesis. *Science* 229(4719), 1193-201.
- Bradshaw, R. A., Neurath, H., Tye, R. W., Walsh, K. A. & Winter, W. P. (1970). Comparison of the partial amino-acid sequence of dogfish trypsinogen with bovine trypsinogen. *Nature* 226(242), 237-9.
- Braisted, A. C. & Wells, J. A. (1996). Minimizing a binding domain from protein A. *Proc Natl Acad Sci U S A* 93(12), 5688-92.
- Brenner, S. & Lerner, R. A. (1992). Encoded combinatorial chemistry. *Proc Natl Acad Sci U S A* 89(12), 5381-3.
- Buchli, P. J., Wu, Z. & Ciardelli, T. L. (1997). The functional display of interleukin-2 on filamentous phage. *Arch Biochem Biophys* 339(1), 79-84.
- Cabibbo, A., Sporeno, E., Toniatti, C., Altamura, S., Savino, R., Paonessa, G. & Ciliberto, G. (1995). Monovalent phage display of human interleukin (hIL)-6: selection of superbinder variants from a complex molecular repertoire in the hIL-6 D-helix. *Gene* 167(1-2), 41-7.
- Carter, P. & Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature* 332(6164), 564-8.
- Chang, A. C., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. & Cohen, S. N. (1978). Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. *Nature* 275(5681), 617-24.
- Charbit, A., Molla, A., Saurin, W. & Hofnung, M. (1988). Versatility of a vector for expressing foreign polypeptides at the surface of gram-negative bacteria. *Gene* 70(1), 181-9.
- Chen, P., Tsuge, H., Almassy, R. J., Gribskov, C. L., Katoh, S., Vanderpool, D. L., Margosiak, S. A., Pinko, C., Matthews, D. A. & Kan, C. C. (1996). Structure

- of the human cytomegalovirus protease catalytic domain reveals a novel serine protease fold and catalytic triad. *Cell* **86**(5), 835-43.
- Chien, C. T., Bartel, P. L., Sternglanz, R. & Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* **88**(21), 9578-82.
- Chiswell, D. J. & McCafferty, J. (1992). Phage antibodies - will new coliclonal antibodies replace monoclonal antibodies. *Trends Biotechnol.* **10**(March), 80-84.
- Choo, Y., Sanchez-Garcia, I. & Klug, A. (1994). In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372**(6507), 642-5.
- Chung, C. H., Ives, H. E., Almeda, S. & Goldberg, A. L. (1983). Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases. *J Biol Chem* **258**(18), 11032-8.
- Chung, C. H., Seol, J. H. & Kang, M. S. (1996). Protease Ti (Clp), a multi-component ATP-dependent protease in *Escherichia coli*. *Biol Chem* **377**(9), 549-54.
- Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature* **352**(6336), 624-8.
- Clackson, T. & Wells, J. A. (1994). In vitro selection from protein and peptide libraries. *Trends Biotechnol.* **12**(May), 173-184.
- Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science* **267**(5196), 383-6.
- Climie, S., Ruiz, P. L., Gonzalez, P. D., Prapunwattana, P., Cho, S. W., Stroud, R. & Santi, D. V. (1990). Saturation site-directed mutagenesis of thymidylate synthase. *J Biol Chem* **265**(31), 18776-9.
- Corey, D. R. & Craik, C. S. (1992). An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J. Am. Chem. Soc.* **114**(5), 1784-1790.
- Corey, D. R., Shiau, A. K., Yang, Q., Janowski, B. A. & Craik, C. S. (1993). Trypsin display on the surface of bacteriophage. *Gene* **128**(1), 129-34.
- Craik, C. S., Choo, Q. L., Swift, G. H., Quinto, C., MacDonald, R. J. & Rutter, W. J. (1984). Structure of two related rat pancreatic trypsin genes. *J Biol Chem* **259**(22), 14255-64.
- Craik, C. S., Largman, C., Fletcher, T., Rocznia, S., Barr, P. J., Fletterick, R. & Rutter, W. J. (1985). Redesigning trypsin: alteration of substrate specificity. *Science* **228**(4697), 291-7.

- Craik, C. S., Roczniaik, S., Largman, C. & Rutter, W. J. (1987a). The catalytic role of the active site aspartic acid in serine proteases. *Science* **237**(4817), 909-13.
- Craik, C. S., Roczniaik, S., Sprang, S., Fletterick, R. & Rutter, W. (1987b). Redesigning trypsin via genetic engineering. *J Cell Biochem* **33**(3), 199-211.
- Crissman, J. W. & Smith, G. P. (1984). Gene-III protein of filamentous phages: evidence for a carboxyl-terminal domain with a role in morphogenesis. *Virology* **132**(2), 445-55.
- Cull, M. G., Miller, J. F. & Schatz, P. J. (1992). Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proc Natl Acad Sci U S A* **89**(5), 1865-9.
- Cunningham, B. C., Henner, D. J. & Wells, J. A. (1990). Engineering human prolactin to bind to the human growth hormone receptor. *Science*, 1461-5.
- Cunningham, B. C. & Wells, J. A. (1987). Improvement in the alkaline stability of subtilisin using an efficient random mutagenesis and screening procedure. *Protein Eng* **1**(4), 319-25.
- Cwirla, S. E., Peters, E. A., Barrett, R. W. & Dower, W. J. (1990). Peptides on phage: a vast library of peptides for identifying ligands. *Proc Natl Acad Sci U S A* **87**(16), 6378-82.
- Dang, Q. D. & Di Cera, E. (1996). Residue 225 determines the Na(+)-induced allosteric regulation of catalytic activity in serine proteases. *Proc Natl Acad Sci U S A* **93**(20), 10653-6.
- Dano, K., Andreasen, P. A., Grondahl-Hansen, J., Kristensen, P., Nielsen, L. S. & Skriver, L. (1985). Plasminogen activators, tissue degradation, and cancer. *Adv Cancer Res* **44**, 139-266.
- Day, L. A. (1969). Conformations of single-stranded DNA and coat protein in fd bacteriophage as revealed by ultraviolet absorption spectroscopy. *J Mol Biol* **39**(2), 265-77.
- Dayringer, H. E., Tramontano, A., Sprang, S. R. & Fletterick, R. J. (1986). Insight II. *Molec. Graphics* **4**, 82-87.
- Dennis, M. S., Herzka, A. & Lazarus, R. A. (1995). Potent and selective Kunitz domain inhibitors of plasma kallikrein designed by phage display. *J Biol Chem* **270**(43), 25411-7.
- Dennis, M. S. & Lazarus, R. A. (1994a). Kunitz domain inhibitors of tissue factor-factor VIIa. I. Potent inhibitors selected from libraries by phage display. *J Biol Chem* **269**(35), 22129-36.

- Dennis, M. S. & Lazarus, R. A. (1994b). Kunitz domain inhibitors of tissue factor-factor VIIa. II. Potent and specific inhibitors by competitive phage selection. *J Biol Chem* 269(35), 22137-44.
- Dunn, I. S. (1996). Phage display of proteins. *Curr Opin Biotechnol* 7(5), 547-53.
- Eerola, R., Saviranta, P., Lilja, H., Pettersson, K., Lovgren, T. & Karp, M. (1994). Expression of prostate specific antigen on the surface of a filamentous phage. *Biochem Biophys Res Commun* 200(3), 1346-52.
- Ellington, A. D. & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature* 346(6287), 818-22.
- Ellington, A. D. & Szostak, J. W. (1992). Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* 355(6363), 850-2.
- Engh, R. A., Huber, R., Bode, W. & Schulze, A. J. (1995). Divining the serpin inhibition mechanism: a suicide substrate 'springe'? *Trends Biotechnol* 13(12), 503-10.
- Erpel, T., Hwang, P., Craik, C. S., Fletterick, R. J. & McGrath, M. E. (1992). Physical map location of the new Escherichia coli gene eco, encoding the serine protease inhibitor ecotin. *J Bacteriol* 174(5), 1704.
- Evnin, L. B., Vasquez, J. R. & Craik, C. S. (1990). Substrate specificity of trypsin investigated by using a genetic selection. *Proc Natl Acad Sci U S A* 87(17), 6659-63.
- Fazioli, F. & Blasi, F. (1994). Urokinase-type plasminogen activator and its receptor: new targets for anti-metastatic therapy? *Trends Pharmacol Sci* 15(1), 25-9.
- Fehlhammer, H., Bode, W. & Huber, R. (1977). Crystal structure of bovine trypsinogen at 1.8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J Mol Biol* 111(4), 415-38.
- Fleminger, G., Wolf, T., Hadas, E. & Solomon, B. (1990). Eupergit C as a carrier for high-performance liquid chromatographic-based immunopurification of antigens and antibodies. *J Chromatogr* 510, 311-9.
- Francisco, J. A., Earhart, C. F. & Georgiou, G. (1992). Transport and anchoring of beta-lactamase to the external surface of Escherichia coli. *Proc Natl Acad Sci U S A* 89(7), 2713-7.
- Friesen, W. J. & Darby, M. K. (1997). Phage display of RNA binding zinc fingers from transcription factor IIIA. *J Biol Chem* 272(17), 10994-7.

- Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. & Gordon, E. M. (1994). Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J Med Chem* 37(9), 1233-51.
- Gardell, S. J., Craik, C. S., Hilvert, D., Urdea, M. S. & Rutter, W. J. (1985). Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis. *Nature* 317(6037), 551-5.
- Geysen, H. M., Meloen, R. H. & Barteling, S. J. (1984). Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc Natl Acad Sci U S A* 81(13), 3998-4002.
- Gill, S. C. & 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(2), 319-26.
- Gong, B. L., SA; Schultz, PG. (1992). A chromagenic assay for screening large antibody libraries. *Journal of the American Chemical Society* 114(N4), 1486-1487.
- Gordon, E. M., Barrett, R. W., Dower, W. J., Fodor, S. P. & Gallop, M. A. (1994). Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J Med Chem* 37(10), 1385-401.
- Graf, L., Craik, C. S., Patthy, A., Rocznik, S., Fletterick, R. J. & Rutter, W. J. (1987). Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry* 26(9), 2616-23.
- Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K. & Baker, D. (1995). A phage display system for studying the sequence determinants of protein folding. *Protein Sci* 4(6), 1108-17.
- Haber, E., Quertermous, T., Matsueda, G. R. & Runge, M. S. (1989). Innovative approaches to plasminogen activator therapy. *Science* 243(4887), 51-6.
- Hawkins, R. E., Russell, S. J. & Winter, G. (1992). Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J Mol Biol* 226(3), 889-96.
- Hedstrom, L., Perona, J. J. & Rutter, W. J. (1994). Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant. *Biochemistry* 33(29), 8757-63.
- Hedstrom, L., Szilagyi, L. & Rutter, W. J. (1992). Converting trypsin to chymotrypsin: the role of surface loops. *Science* 255(5049), 1249-53.
- Hermes, J. D., Blacklow, S. C. & Knowles, J. R. (1990). Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. *Proc Natl Acad Sci U S A* 87(2), 696-700.

- Hermes, J. D., Parekh, S. M., Blacklow, S. C., Koster, H. & Knowles, J. R. (1989). A reliable method for random mutagenesis: the generation of mutant libraries using spiked oligodeoxyribonucleotide primers. *Gene* **84**(1), 143-51.
- Hermodson, M. A., Ericsson, L. H., Neurath, H. & Walsh, K. A. (1973). Determination of the amino acid sequence of porcine trypsin by sequenator aalysis. *Biochemistry* **12**(17), 3146-53.
- Higaki, J. N., Evnin, L. B. & Craik, C. S. (1989). Introduction of a cysteine protease active site into trypsin. *Biochemistry* **28**(24), 9256-63.
- Higaki, J. N., Haymore, B. L., Chen, S., Fletterick, R. J. & Craik, C. S. (1990). Regulation of serine protease activity by an engineered metal switch. *Biochemistry* **29**(37), 8582-6.
- Hoogenboom, H. R. (1997). Designing and optimizing library selection strategies for generating high-affinity antibodies. *Trends Biotechnol* **15**(2), 62-70.
- Hoogenboom, H. R. & Winter, G. (1992). By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. *J Mol Biol* **227**(2), 381-8.
- Houghten, R. A. (1985). General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc Natl Acad Sci U S A* **82**(15), 5131-5.
- Huber, R. & Bode, W. (1978). Structural basis of the activation and action of trypsin. *Acc Chem Res* **11**, 114-22.
- Jamieson, A. C., Kim, S. H. & Wells, J. A. (1994). In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* **33**(19), 5689-95.
- Janda, K. D., Lo, C. H., Li, T., Barbas, C. r., Wirsching, P. & Lerner, R. A. (1994). Direct selection for a catalytic mechanism from combinatorial antibody libraries. *Proc Natl Acad Sci U S A* **91**(7), 2532-6.
- Jin, L., Fendly, B. M. & Wells, J. A. (1992). High resolution functional analysis of antibody-antigen interactions. *J Mol Biol* **226**(3), 851-65.
- Jin, L. & Wells, J. A. (1994). Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting. *Protein Sci* **3**(12), 2351-7.
- Jones, T. A. (1978). A Graphics Model Building and Refinement System for Macromolecules. *Journal of Applied Crystallography* **11**, 268-272.
- Kadonaga, J. T. & Knowles, J. R. (1985). A simple and efficient method for chemical mutagenesis of DNA. *Nucleic Acids Res* **13**(5), 1733-45.
- Kauffman, S. A. (1992). Applied molecular evolution. *J Theor Biol* **157**(1), 1-7.

- Knowles, J. R. (1987). Tinkering with enzymes: what are we learning? *Science* **236**(4806), 1252-8.
- Kohler, G. & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**(5517), 495-7.
- Kossiakoff, A. A., Chambers, J. L., Kay, L. M. & Stroud, R. M. (1977). Structure of bovine trypsinogen at 1.9 Å resolution. *Biochemistry* **16**(4), 654-64.
- Ku, J. & Schultz, P. G. (1995). Alternate protein frameworks for molecular recognition. *Proc Natl Acad Sci U S A* **92**(14), 6552-6.
- Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A* **82**(2), 488-92.
- Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol* **154**, 367-82.
- Lang, I. M., Chuang, T. L., Barbas, C. F. r. & Schleef, R. R. (1996). Purification of storage granule protein-23. A novel protein identified by phage display technology and interaction with type I plasminogen activator inhibitor. *J Biol Chem* **271**(47), 30126-35.
- Langley, K. E., Fowler, A. V. & Zabin, I. (1975a). Amino acid sequence of beta-galactosidase. IV. Sequence of an alpha-complementing cyanogen bromide peptide, residues 3 to 92. *J Biol Chem* **250**(7), 2587-92.
- Langley, K. E., Villarejo, M. R., Fowler, A. V., Zamenhof, P. J. & Zabin, I. (1975b). Molecular basis of beta-galactosidase alpha-complementation. *Proc Natl Acad Sci U S A* **72**(4), 1254-7.
- Langley, K. E. & Zabin, I. (1976). beta-Galactosidase alpha complementation: properties of the complemented enzyme and mechanism of the complementation reaction. *Biochemistry* **15**(22), 4866-75.
- Laskowski, M. J. & Kato, I. (1980). Protein inhibitors of proteinases. *Annu Rev Biochem* **49**(593), 593-626.
- Lasters, I., Van Herzele, N., Lijnen, H. R., Collen, D. & Jespers, L. (1997). Enzymatic properties of phage-displayed fragments of human plasminogen. *Eur J Biochem* **244**(3), 946-52.
- Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55**(3), 379-400.
- Lee, H. R., Seo, J. H., Kim, O. M., Lee, C. S., Suh, S. W., Hong, Y. M., Tanaka, K., Ichihara, A., Ha, D. B. & Chung, C. H. (1991). Molecular cloning of the

- ecotin gene in *Escherichia coli* [published erratum appears in *FEBS Lett* 1992 Jan 20;296(2):235]. *Febs Lett* 287(1-2), 53-6.
- Lesley, S. A., Patten, P. A. & Schultz, P. G. (1993). A genetic approach to the generation of antibodies with enhanced catalytic activities. *Proc Natl Acad Sci U S A* 90(4), 1160-5.
- Leung, D. W., Chen, E. Y. & Goeddel, D. V. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* 1, 11-15.
- Li, B., Tom, J. Y., Oare, D., Yen, R., Fairbrother, W. J., Wells, J. A. & Cunningham, B. C. (1995). Minimization of a polypeptide hormone. *Science* 270(5242), 1657-60.
- Liao, D. I., Breddam, K., Sweet, R. M., Bullock, T. & Remington, S. J. (1992). Refined atomic model of wheat serine carboxypeptidase II at 2.2-Å resolution. *Biochemistry* 31(40), 9796-812.
- Light, J. & Lerner, R. A. (1995). Random mutagenesis of staphylococcal nuclease and phage display selection. *Bioorg Med Chem* 3(7), 955-67.
- Little, M., Fuchs, P., Breitling, F. & Dubel, S. (1993). Bacterial surface presentation of proteins and peptides - an alternative to phage technology. *Trends Biotechnol.* 11(Jan), 3-5.
- Livnah, O., Stura, E. A., Johnson, D. L., Middleton, S. A., Mulcahy, L. S., Wrighton, N. C., Dower, W. J., Jolliffe, L. K. & Wilson, I. A. (1996). Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å [see comments]. *Science* 273(5274), 464-71.
- Lorsch, J. R. & Szostak, J. W. (1994). In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* 371(6492), 31-6.
- Lowman, H. B., Bass, S. H., Simpson, N. & Wells, J. A. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* 30(45), 10832-8.
- Lowman, H. B. & Wells, J. A. (1993). Affinity maturation of human growth hormone by monovalent phage display. *J Mol Biol* 234(3), 564-78.
- Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogini, E., Rothberg, I., Ryan, K. & Laskowski, M., Jr. (1997). Binding of amino acid side-chains to S1 cavities of serine proteinases. *J Mol Biol* 266(2), 441-61.
- Lundblad, R. L. (1991). *Chemical reagents for protein modification*. 2nd edit, CRC Press, Boca Raton, Florida.

- Lundblad, R. L. (1995). *Techniques in protein modification*, CRC Press, Boca Raton, Florida.
- Maenaka, K., Furuta, M., Tsumoto, K., Watanabe, K., Ueda, Y. & Kumagai, I. (1996). A stable phage-display system using a phagemid vector: phage display of hen egg-white lysozyme (HEL), *Escherichia coli* alkaline, phosphatase, and anti-HEL monoclonal antibody, HyHEL10. *Biochem Biophys Res Commun* **218**(3), 682-7.
- Main, A. R. (1964). Affinity and phosphorylation constants for the inhibition of esterases by organophosphates. *Science* **144**, 992.
- Makowski, L. (1993). Structural constraints on the display of foreign peptides on filamentous bacteriophages. *GGGene* **128**(1), 5-11.
- Markland, W., Ley, A. C. & Ladner, R. C. (1996a). Iterative optimization of high-affinity protease inhibitors using phage display. 2. Plasma kallikrein and thrombin. *Biochemistry* **35**(24), 8058-67.
- Markland, W., Ley, A. C., Lee, S. W. & Ladner, R. C. (1996b). Iterative optimization of high-affinity proteases inhibitors using phage display. 1. Plasmin. *Biochemistry* **35**(24), 8045-57.
- Markland, W., Roberts, B. L., Saxena, M. J., Guterman, S. K. & Ladner, R. C. (1991). Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage M13. *Gene* **109**(1), 13-9.
- Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D. & Winter, G. (1991). By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J Mol Biol* **222**(3), 581-97.
- Martin, F., Toniatti, C., Salvati, A. L., Ciliberto, G., Cortese, R. & Sollazzo, M. (1996). Coupling protein design and in vitro selection strategies: improving specificity and affinity of a designed beta-protein IL-6 antagonist. *J Mol Biol* **255**(1), 86-97.
- Mattheakis, L. C., Bhatt, R. R. & Dower, W. J. (1994). An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc Natl Acad Sci U S A* **91**(19), 9022-6.
- Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. & et al. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **77**(5), 761-71.
- Matthews, D. J. & Wells, J. A. (1993). Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**(5111), 1113-7.

- McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**(6301), 552-4.
- McCafferty, J., Jackson, R. H. & Chiswell, D. J. (1991). Phage-enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. *Protein Eng* **4**(8), 955-61.
- McConnell, S. J., Uveges, A. J. & Spinella, D. G. (1995). Comparison of plate versus liquid amplification of M13 phage display libraries. *Biotechniques* **18**(5), 803-4, 806.
- McCoy, J. M. & Khorana, H. G. (1983). Introduction and characterization of amber mutations in the bacteriorhodopsin gene. *J Biol Chem* **258**(13), 8456-61.
- McGrath, M. E., Erpel, T., Browner, M. F. & Fletterick, R. J. (1991a). Expression of the protease inhibitor ecotin and its co-crystallization with trypsin. *J Mol Biol* **222**(2), 139-42.
- McGrath, M. E., Erpel, T., Bystroff, C. & Fletterick, R. J. (1994). Macromolecular chelation as an improved mechanism of protease inhibition: structure of the ecotin-trypsin complex. *Embo J* **13**(7), 1502-7.
- McGrath, M. E., Gillmor, S. A. & Fletterick, R. J. (1995). Ecotin: lessons on survival in a protease-filled world. *Protein Sci* **4**(2), 141-8.
- McGrath, M. E., Hines, W. M., Sakanari, J. A., Fletterick, R. J. & Craik, C. S. (1991b). The sequence and reactive site of ecotin. A general inhibitor of pancreatic serine proteases from *Escherichia coli*. *J Biol Chem* **266**(10), 6620-5.
- McGrath, M. E., Vasquez, J. R., Craik, C. S., Yang, A. S., Honig, B. & Fletterick, R. J. (1992). Perturbing the polar environment of Asp102 in trypsin: consequences of replacing conserved Ser214. *Biochemistry* **31**(12), 3059-64.
- McGregor, D. (1996). Selection of proteins and peptides from libraries displayed on filamentous bacteriophage. *Mol Biotechnol* **6**(2), 155-62.
- Miller, C. G., Green, L. & Schultz, R. (1982). *Salmonella typhimurium* mutations affecting utilization of L-leucine beta-naphthylamide. *Mol Gen Genet* **186**(2), 228-34.
- Morrison, J. F. (1969). Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim Biophys Acta* **185**(2), 269-86.
- Munir, K. M., French, D. C., Dube, D. K. & Loeb, L. A. (1992). Permissible amino acid substitutions within the putative nucleoside binding site of herpes simplex virus type 1 encoded thymidine kinase established by random sequence mutagenesis [corrected] [published erratum appears in J Biol Chem 1992 Jul 25;267(21):15258]. *J Biol Chem* **267**(10), 6584-9.

- Munir, K. M., French, D. C. & Loeb, L. A. (1993). Thymidine kinase mutants obtained by random sequence selection. *Proc Natl Acad Sci U S A* **90**(9), 4012-6.
- Needels, M. C., Jones, D. G., Tate, E. H., Heinkel, G. L., Kochersperger, L. M., Dower, W. J., Barrett, R. W. & Gallop, M. A. (1993). Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc Natl Acad Sci U S A* **90**(22), 10700-4.
- Nelson, F. K., Friedman, S. M. & Smith, G. P. (1981). Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III. *Virology* **108**(2), 338-50.
- Neurath, H. (1985). Proteolytic enzymes, past and present. *Fed Proc* **44**(14), 2907-13.
- Nord, K., Nilsson, J., Nilsson, B., Uhlen, M. & Nygren, P. A. (1995). A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng* **8**(6), 601-8.
- O'Neil, K. T. & Hoess, R. H. (1995). Phage display: protein engineering by directed evolution. *Curr Opin Struct Biol* **5**(4), 443-9.
- O'Neil, K. T., Hoess, R. H., Raleigh, D. P. & DeGrado, W. F. (1995). Thermodynamic genetics of the folding of the B1 immunoglobulin-binding domain from streptococcal protein G. *Proteins* **21**(1), 11-21.
- Ortiz, C., Tellier, C., Williams, H., Stolowich, N. J. & Scott, A. I. (1991). Diastereotopic covalent binding of the natural inhibitor leupeptin to trypsin: detection of two interconverting hemiacetals by solution and solid-state NMR spectroscopy. *Biochemistry* **30**(41), 10026-34.
- Pal, G., Sprengel, G., Patthy, A. & Graf, L. (1994). Alteration of the specificity of ecotin, an E. coli serine proteinase inhibitor, by site directed mutagenesis. *FEBS Lett* **342**(1), 57-60.
- Pal, G., Szilagyi, L. & Graf, L. (1996). Stable monomeric form of an originally dimeric serine proteinase inhibitor, ecotin, was constructed via site directed mutagenesis. *FEBS Lett* **385**(3), 165-70.
- Pannekoek, H., van Meijer, M., Schleef, R. R., Loskutoff, D. J. & Barbas, C. d. (1993). Functional display of human plasminogen-activator inhibitor 1 (PAI-1) on phages: novel perspectives for structure-function analysis by error-prone DNA synthesis. *Gene* **128**(1), 135-40.
- Parmley, S. F. & Smith, G. P. (1988). Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73**(2), 305-18.
- Parmley, S. F. & Smith, G. P. (1989). Filamentous fusion phage cloning vectors for the study of epitopes and design of vaccines. *Adv Exp Med Biol* **251**, 215-8.

- Patel, A. H., Ahsan, A., Suthar, B. P. & Schultz, R. M. (1983). Transition-state affinity chromatography of trypsin-like proteinases with dipeptidyl argininal ligands. *Biochim Biophys Acta* **748**(2), 321-30.
- Patel, A. H. & Schultz, R. M. (1982). Transition-state affinity purification of proteases; the preparation of an argininal affinity resin for the selective binding of trypsin-like proteases. *Biochem Biophys Res Commun* **104**(1), 181-6.
- Perona, J. J. & Craik, C. S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Science*.
- Perona, J. J., Evnin, L. B. & Craik, C. S. (1993a). A genetic selection elucidates structural determinants of arginine versus lysine specificity in trypsin. *Gene* **137**(1), 121-6.
- Perona, J. J., Tsu, C. A., Craik, C. S. & Fletterick, R. J. (1993b). Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J Mol Biol* **230**(3), 919-33.
- Peters, E. A., Schatz, P. J., Johnson, S. S. & Dower, W. J. (1994). Membrane insertion defects caused by positive charges in the early mature region of protein pIII of filamentous phage fd can be corrected by prlA suppressors. *J Bacteriol* **176**(14), 4296-305.
- Ponder, J. W. & Richards, F. M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* **193**(4), 775-91.
- Posner, B., Smiley, J., Lee, I. & Benkovic, S. (1994). Catalytic antibodies: perusing combinatorial libraries. *Trends Biochem Sci* **19**(4), 145-50.
- Qasim, M. A., Ganz, P. J., Saunders, C. W., Bateman, K. S., James, M. N. & Laskowski, M., Jr. (1997). Interscaffolding additivity. Association of P1 variants of eglin c and of turkey ovomucoid third domain with serine proteinases. *Biochemistry* **36**(7), 1598-607.
- Qiu, X., Culp, J. S., DiLella, A. G., Hellmig, B., Hoog, S. S., Janson, C. A., Smith, W. W. & Abdel-Meguid, S. S. (1996). Unique fold and active site in cytomegalovirus protease. *Nature* **383**(6597), 275-9.
- Qiu, X., Janson, C. A., Culp, J. S., Richardson, S. B., Debouck, C., Smith, W. W. & Abdel-Meguid, S. S. (1997). Crystal structure of varicella-zoster virus protease. *Proc Natl Acad Sci U S A* **94**(7), 2874-9.
- Rawlings, N. D. & Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochem J*, 205-18.
- Rebar, E. J. & Pabo, C. O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**(5147), 671-3.

- Reidhaar, O. J. & Sauer, R. T. (1988). Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* **241**(4861), 53-7.
- Rickles, R. J., Botfield, M. C., Zhou, X. M., Henry, P. A., Brugge, J. S. & Zoller, M. J. (1995). Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci U S A* **92**(24), 10909-13.
- Ring, C. S., Kneller, D. G., Langridge, R. & Cohen, F. E. (1992). Taxonomy and conformational analysis of loops in proteins [published erratum appears in *J Mol Biol* 1992 Oct 5;227(3):977]. *J Mol Biol* **224**(3), 685-99.
- Roberts, B. L., Markland, W., Ley, A. C., Kent, R. B., White, D. W., Guterman, S. K. & Ladner, R. C. (1992a). Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc Natl Acad Sci U S A* **89**(6), 2429-33.
- Roberts, B. L., Markland, W., Siranosian, K., Saxena, M. J., Guterman, S. K. & Ladner, R. C. (1992b). Protease inhibitor display M13 phage: selection of high-affinity neutrophil elastase inhibitors. *Gene* **121**(1), 9-15.
- Rothman, J. H. & Stevens, T. H. (1986). Protein sorting in yeast: mutants defective in vacuole biogenesis mislocalize vacuolar proteins into the late secretory pathway. *Cell* **47**(6), 1041-51.
- Rydel, T. J., Ravichandran, K. G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C. & Fenton, J. W. d. (1990). The structure of a complex of recombinant hirudin and human alpha-thrombin. *Science* **249**(4966), 277-80.
- Saggio, I., Gloaguen, I. & Laufer, R. (1995). Functional phage display of ciliary neurotrophic factor. *Gene* **152**(1), 35-9.
- Sambrook, J., Fritsch, E. F. & Maniatis, T., Eds. (1989). *Molecular cloning: a laboratory manual*. 2nd edit. Plainview, New York: Cold Spring Harbor Laboratory Press.
- Schatz, P. J. (1993). Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme - a 13 residue consensus peptide specifies biotinylation in escherichia-coli. *Bio-technol.* **11**(Oct), 1138-1143.
- Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157.
- Schmitt, S., Goetschel, J. D., Rommelt, E. M. & Adler, C. (1990). One-step purification of trypsin and alpha-chymotrypsin by affinity chromatography on Eupergit-aprotinin, a novel carrier for purification of serine proteases. *J Chromatogr* **510**, 239-42.

- Schmitz, R., Baumann, G. & Gram, H. (1996). Catalytic specificity of phosphotyrosine kinases Blk, Lyn, c-Src and Syk as assessed by phage display. *J Mol Biol* **260**(5), 664-77.
- Schultz, R. M., Varma-Nelson, P., Ortiz, R., Kozlowski, K. A., Orawski, A. T., Pagast, P. & Frankfater, A. (1989). Active and inactive forms of the transition-state analog protease inhibitor leupeptin: explanation of the observed slow binding of leupeptin to cathepsin B and papain. *J Biol Chem* **264**(3), 1497-507.
- Schultz, S. C. & Richards, J. H. (1986). Site-saturation studies of beta-lactamase: production and characterization of mutant beta-lactamases with all possible amino acid substitutions at residue 71. *Proc Natl Acad Sci U S A* **83**(6), 1588-92.
- Schulze, A. J., Huber, R., Bode, W. & Engh, R. A. (1994). Structural aspects of serpin inhibition. *Febs Lett* **344**(2-3), 117-24.
- Scott, J. K. & Smith, G. P. (1990). Searching for peptide ligands with an epitope library. *Science* **249**(4967), 386-90.
- Seong, I. S., Lee, H. R., Seol, J. H., Park, S. K., Lee, C. S., Suh, S. W., Hong, Y. M., Kang, M. S., Ha, D. B. & Chung, C. H. (1994). The P1 reactive site methionine residue of ecotin is not crucial for its specificity on target proteases. A potent inhibitor of pancreatic serine proteases from *Escherichia coli*. *J Biol Chem* **269**(34), 21915-8.
- Seymour, J. L., Lindquist, R. N., Dennis, M. S., Moffat, B., Yansura, D., Reilly, D., Wessinger, M. E. & Lazarus, R. A. (1994). Ecotin is a potent anticoagulant and reversible tight-binding inhibitor of factor Xa. *Biochemistry* **33**(13), 3949-58.
- Shaw, E., Mares-Guia, M. & Cohen, W. (1965). Evidence for an active-center histidine in trypsin through the use of a specific reagent, 1-chloro-3-tosylamido-7-amino-2-heptanone, the chloromethyl ketone derived from Na-tosyl-L-lysine. *Biochemistry* **4**, 2219.
- Shaw, W. V. (1987). Protein engineering. The design, synthesis and characterization of factitious proteins. *Biochem J* **246**(1), 1-17.
- Shieh, H. S., Kurumbail, R. G., Stevens, A. M., Stegeman, R. A., Sturman, E. J., Pak, J. Y., Wittwer, A. J., Palmier, M. O., Wiegand, R. C., Holwerda, B. C. & Stallings, W. C. (1996). Three-dimensional structure of human cytomegalovirus protease. *Nature* **383**(6597), 279-82.
- Shin, D. H., Hwang, K. Y., Kim, K. K., Lee, H. R., Lee, C. S., Chung, C. H. & Suh, S. W. (1993). Crystallization and preliminary X-ray crystallographic analysis of the protease inhibitor ecotin. *J Mol Biol* **229**(4), 1157-8.
- Shin, D. H., Lee, C. S., Chung, C. H. & Suh, S. W. (1996a). Molecular symmetry of the ClpP component of the ATP-dependent Clp protease, an *Escherichia coli* homolog of 20 S proteasome. *J Mol Biol* **262**(2), 71-6.

- Shin, D. H., Song, H. K., Seong, I. S., Lee, C. S., Chung, C. H. & Suh, S. W. (1996b). Crystal structure analyses of uncomplexed ecotin in two crystal forms: implications for its function and stability. *Protein Sci* **5**(11), 2236-47.
- Shortle, D. & Lin, B. (1985). Genetic analysis of staphylococcal nuclease: identification of three intragenic "global" suppressors of nuclease-minus mutations. *Genetics* **110**(4), 539-55.
- Sloane, D. L., Leung, R., Craik, C. S. & Sigal, E. (1991). A primary determinant for lipoygenase positional specificity. *Nature* **354**(6349), 149-52.
- Smiley, J. A. & Benkovic, S. J. (1994). Selection of catalytic antibodies for a biosynthetic reaction from a combinatorial cDNA library by complementation of an auxotrophic *Escherichia coli*: antibodies for orotate decarboxylation. *Proc Natl Acad Sci U S A* **91**(18), 8319-23.
- Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**(4705), 1315-7.
- Smith, G. P. (1988a). Filamentous phage assembly: morphogenetically defective mutants that do not kill the host. *Virology* **167**(1), 156-65.
- Smith, G. P. (1988b). Filamentous phages as cloning vectors. *Biotechnology* **10**(61), 61-83.
- Smith, G. P. & Scott, J. K. (1993). Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol* **217**(228), 228-57.
- Smith, J. W., Tachias, K. & Madison, E. L. (1995). Protein loop grafting to construct a variant of tissue-type plasminogen activator that binds platelet integrin α IIb β 3. *J Biol Chem* **270**(51), 30486-90.
- Soumillion, P., Jespers, L., Bouchet, M., Marchand, B. J., Sartiaux, P. & Fastrez, J. (1994a). Phage display of enzymes and in vitro selection for catalytic activity. *Appl Biochem Biotechnol* **47**(2-3), 175-89; discussion 189-90.
- Soumillion, P., Jespers, L., Bouchet, M., Marchand-Brynaert, J., Winter, G. & Fastrez, J. (1994b). Selection of beta-lactamase on filamentous bacteriophage by catalytic activity. *J Mol Biol* **237**(4), 415-22.
- Spraggon, G., Phillips, C., Nowak, U. K., Ponting, C. P., Saunders, D., Dobson, C. M., Stuart, D. I. & Jones, E. Y. (1995). The crystal structure of the catalytic domain of human urokinase-type plasminogen activator. *Structure* **3**(7), 681-91.
- Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer, M. J., Xuong, N. H., Hamlin, R., Rutter, W. J. & Craik, C. S. (1987). The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. *Science* **237**(4817), 905-9.

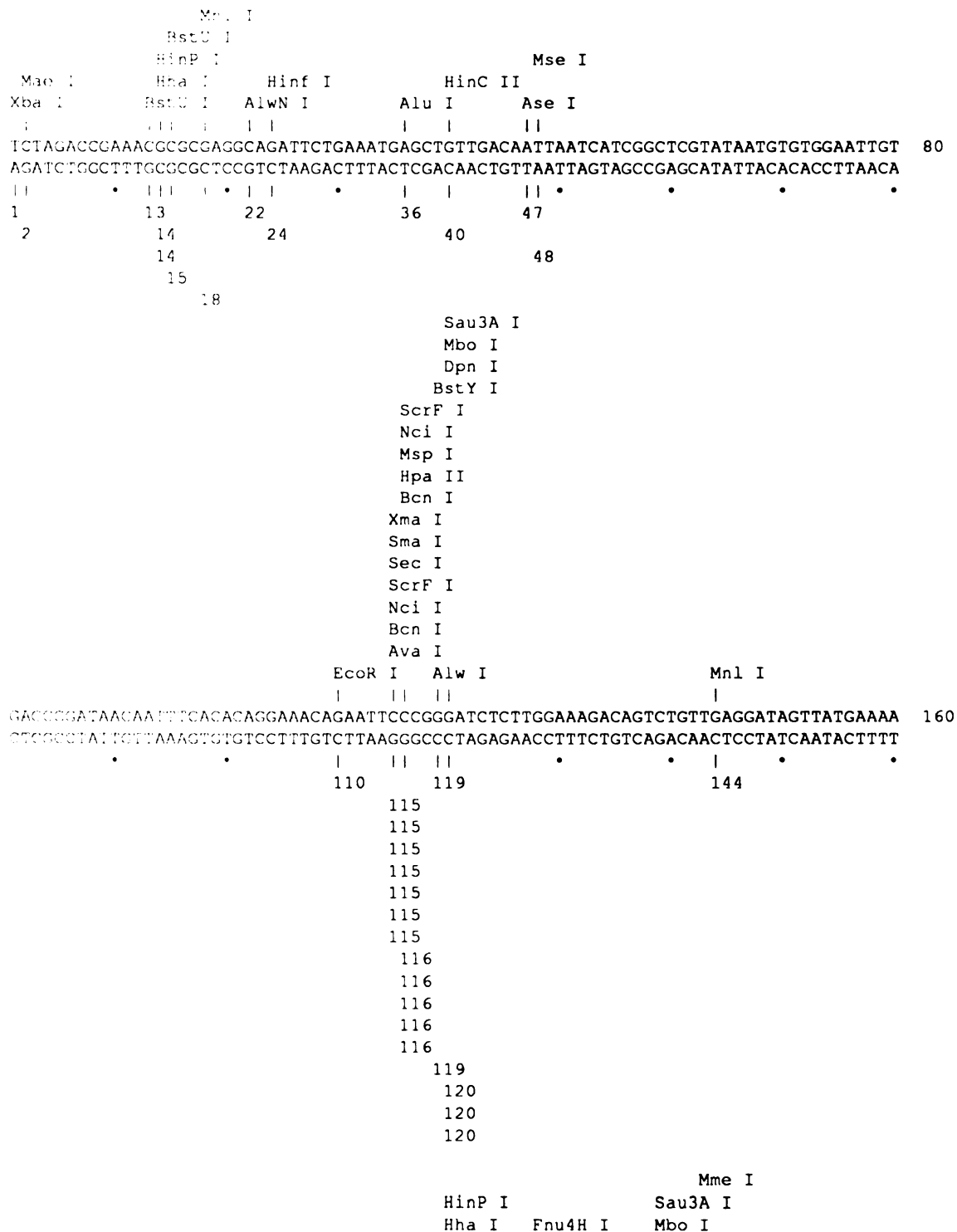
- Stemmer, W. P. (1994a). Rapid evolution of a protein in vitro by DNA shuffling [see comments]. *Nature* **370**(6488), 389-91.
- Stemmer, W. P. C. (1994b). DNA Shuffling by Random Fragmentation and Reassembly: *In vitro* Recombination for Molecular Evolution. *Proc. Natl. Acad. Sci. USA* in press.
- Stratton-Thomas, J. R., Min, H. Y., Kaufman, S. E., Chiu, C. Y., Mullenbach, G. T. & Rosenberg, S. (1995). Yeast expression and phagemid display of the human urokinase plasminogen activator epidermal growth factor-like domain. *Protein Eng* **8**(5), 463-70.
- Stubbs, M. T. & Bode, W. (1995). The clot thickens: clues provided by thrombin structure [published erratum appears in *Trends Biochem Sci* 1995 Mar;20(3):131]. *Trends Biochem Sci* **20**(1), 23-8.
- Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-A resolution. *Biochemistry* **13**(20), 4212-28.
- Szostak, J. W. (1992). In vitro genetics. *Trends Biochem Sci* **17**(3), 89-93.
- Takeuchi, Y., Satow, Y., Nakamura, K. T. & Mitsui, Y. (1991). Refined crystal structure of the complex of subtilisin BPN' and *Streptomyces* subtilisin inhibitor at 1.8 Å resolution. *J Mol Biol* **221**(1), 309-25.
- Tanaka, A. S., Sampaio, C. A., Fritz, H. & Auerswald, E. A. (1995). Functional display and expression of chicken cystatin using a phagemid system. *Biochem Biophys Res Commun* **214**(2), 389-95.
- Tang, Y., Jiang, N., Parakh, C. & Hilvert, D. (1996). Selection of linkers for a catalytic single-chain antibody using phage display technology. *J Biol Chem* **271**(26), 15682-6.
- Terskikh, A. V., Le Doussal, J. M., Cramer, R., Fisch, I., Mach, J. P. & Kajava, A. V. (1997). "Peptabody": a new type of high avidity binding protein. *Proc Natl Acad Sci U S A* **94**(5), 1663-8.
- Tong, L., Qian, C., Massariol, M. J., Bonneau, P. R., Cordingley, M. G. & Lagace, L. (1996). A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature* **383**(6597), 272-5.
- Tsu, C. A. & Craik, C. S. (1996). Substrate recognition by recombinant serine collagenase 1 from *Uca pugnator*. *J Biol Chem* **271**(19), 11563-70.
- Tsu, C. A., Perona, J. J., Schellenberger, V., Turck, C. W. & Craik, C. S. (1994). The substrate specificity of *Uca pugnator* collagenolytic serine protease 1 correlates with the bovine type I collagen cleavage sites. *J Biol Chem* **269**(30), 19565-72.

- Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**(4968), 505-10.
- Ulmer, J. S., Lindquist, R. N., Dennis, M. S. & Lazarus, R. A. (1995). Ecotin is a potent inhibitor of the contact system proteases factor XIIa and plasma kallikrein. *FEBS Lett* **365**(2-3), 159-63.
- Ultsch, M. H., Somers, W., Kossiakoff, A. A. & de Vos, A. M. (1994). The crystal structure of affinity-matured human growth hormone at 2 Å resolution. *J Mol Biol* **236**(1), 286-99.
- van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W. & Bode, W. (1995). Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *Embo J* **14**(21), 5149-57.
- van de Locht, A., Stubbs, M. T., Bode, W., Friedrich, T., Bollschweiler, C., Hoffken, W. & Huber, R. (1996). The ornithodorin-thrombin crystal structure, a key to the TAP enigma? *Embo J* **15**(22), 6011-7.
- van Meijer, M., Roelofs, Y., Neels, J., Horrevoets, A. J., van Zonneveld, A. J. & Pannekoek, H. (1996). Selective screening of a large phage display library of plasminogen activator inhibitor 1 mutants to localize interaction sites with either thrombin or the variable region 1 of tissue-type plasminogen activator. *J Biol Chem* **271**(13), 7423-8.
- Vanwetswinkel, S., Touillaux, R., Fastrez, J. & Marchand-Brynaert, J. (1995). Bifunctional activity labels for selection of filamentous bacteriophages displaying enzymes. *Bioorg Med Chem* **3**(7), 907-15.
- Vasquez, J. R., Evnin, L. B., Higaki, J. N. & Craik, C. S. (1989). An expression system for trypsin. *J Cell Biochem* **39**(3), 265-76.
- Wang, C. I., Yang, Q. & Craik, C. S. (1995). Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin. *J Biol Chem* **270**(20), 12250-6.
- Wang, C. I., Yang, Q. & Craik, C. S. (1996). Phage display of proteases and macromolecular inhibitors. *Methods Enzymol* **267**, 52-68.
- Wells, J. A. (1990). Additivity of mutational effects in proteins. *Biochemistry* **29**(37), 8509-17.
- Wells, J. A. (1996). Binding in the growth hormone receptor complex. *Proc Natl Acad Sci U S A* **93**(1), 1-6.
- Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. (1987). Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering. *Proc Natl Acad Sci U S A* **84**(15), 5167-71.

- Wells, J. A. & Estell, D. A. (1988). Subtilisin--an enzyme designed to be engineered. *Trends Biochem Sci* **13**(8), 291-7.
- Wells, J. A., Vasser, M. & Powers, D. B. (1985). Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **34**(2-3), 315-23.
- Widersten, M. & Mannervik, B. (1995). Glutathione transferases with novel active sites isolated by phage display from a library of random mutants. *J Mol Biol* **250**(2), 115-22.
- Wieczorek, M. & Laskowski, M. J. (1983). Formation of covalent hybrids from amino-terminal and carboxy-terminal fragments of two ovomucoid third domains. *Biochemistry* **22**(11), 2630-6.
- Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R. & Holbrook, J. J. (1988). A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science* **242**(4885), 1541-4.
- Willett, W. S., Brinen, L. S., Fletterick, R. J. & Craik, C. S. (1996). Delocalizing trypsin specificity with metal activation. *Biochemistry* **35**(19), 5992-8.
- Williams, J. W. & Morrison, J. F. (1979). The kinetics of reversible tight-binding inhibition. *Methods Enzymol* **63**, 437-67.
- Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu Rev Immunol* **12**(433), 433-55.
- Wrighton, N. C., Farrell, F. X., Chang, R., Kashyap, A. K., Barbone, F. P., Mulcahy, L. S., Johnson, D. L., Barrett, R. W., Jolliffe, L. K. & Dower, W. J. (1996). Small peptides as potent mimetics of the protein hormone erythropoietin [see comments]. *Science* **273**(5274), 458-64.
- Yang, Q., Wang, C. I. & Craik, C. S. (1995). Altering the function of enzymes and macromolecular inhibitors by phage display. In *Combinatorial libraries: synthesis, screening and application potential* (Cortese, R., ed.), pp. 205-221. Walter de Gruyter, Berlin.
- Yang, S. Q. & Craik, C. S. (1997). Ecotin : a serine protease inhibitor with two distinct and interacting binding sites.
- Zacher, A. N. d., Stock, C. A., Golden, J. W. d. & Smith, G. P. (1980). A new filamentous phage cloning vector: fd-tet. *Gene* **9**(1-2), 127-40.
- Zinder, N. D. & Boeke, J. D. (1982). The filamentous phage (Ff) as vectors for recombinant DNA--a review. *Gene* **19**(1), 1-10.

Appendix One: Restriction Map of Trypsin Phage Display Vector Xba I / Hind III Fragment of pMTsn::III

Positions of Restriction Endonucleases sites (unique sites underlined>



1292 1297 1312 1348

HinP I
Hha I
Hae II
Eco47 III Nde I Mae III Sec I
Dsa I

AATGTCGCCCTTTTGTCTTTAGCGCTGGTAAACCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCGGTGGT
1440
TTACAGCGGGAAAACAGAAATCGCGACCATTGGTATACTTAAAGATAACTAACACTGTTTTATTGAATAAGGCACCA
1381 1394 1415 1434
1381 1434
1382
1382

Mae II Ple I
Hinf I

GTGTTTGGCTTTCCTTTTATATGTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGTC
1520
CAGAAACGCGAAAGAAAATATACAACGGTGGAATACATACATAAAAGATGCAAACGATTGTATGACGCATTATTCCTCAG
1489 1516
1516

Alu I
Hind III
Msc I
1521
1524
1525

Unique Restriction Sites

Enzyme	Site	Use	Site position (Fragment length)	Fragment order
Aha II	gr/cgyc	1	1(600) 2	601(929) 1
Alw I	ggatc	4/5	1(118) 2	119(1411) 1
Apal I	g/tgac	1	1(544) 2	545(985) 1
Ase I	at/taat	1	1(46) 2	47(1483) 1
Ava II	g/gwcc	1	1(891) 1	892(638) 2
Bbv II	gaagac	2/6	1(452) 2	453(1077) 1
Bcl I	t/gatca	1	1(478) 2	479(1051) 1
Bgl I	gacnnnn/nggc	1	1(519) 2	520(1010) 1
BstX I	ccannnnn/ntgg	1	1(284) 2	285(1245) 1
BstY I	r/gatcy	1	1(118) 2	119(1411) 1
Cla I	at/cgat	1	1(1197) 1	1198(332) 2
Dra III	cacnnn/qtg	1	1(557) 2	558(972) 1
Drd I	gacnnnn/ngtc	1	1(690) 2	691(839) 1
Dsa I	c/crygg	1	1(1433) 1	1434(96) 2
Eae I	y/ggccc	1	1(517) 2	518(1012) 1
Esp I	gc/tgagc	1	1(590) 2	591(939) 1
Gau I	ctgga	16/14	1(621) 2	622(908) 1
HinC III	qly/rac	1	1(39) 2	40(1490) 1
Hind III	a/agcctt	1	1(1523) 1	1524(6) 2
Xmo I	tccrac	20/18	1(222) 2	223(1307) 1
Msc I	tgg/cga	1	1(517) 2	518(1012) 1
Nde I	ca/tatg	1	1(1393) 1	1394(136) 2
PaeR7 I	c/tcgag	1	1(511) 2	512(1018) 1

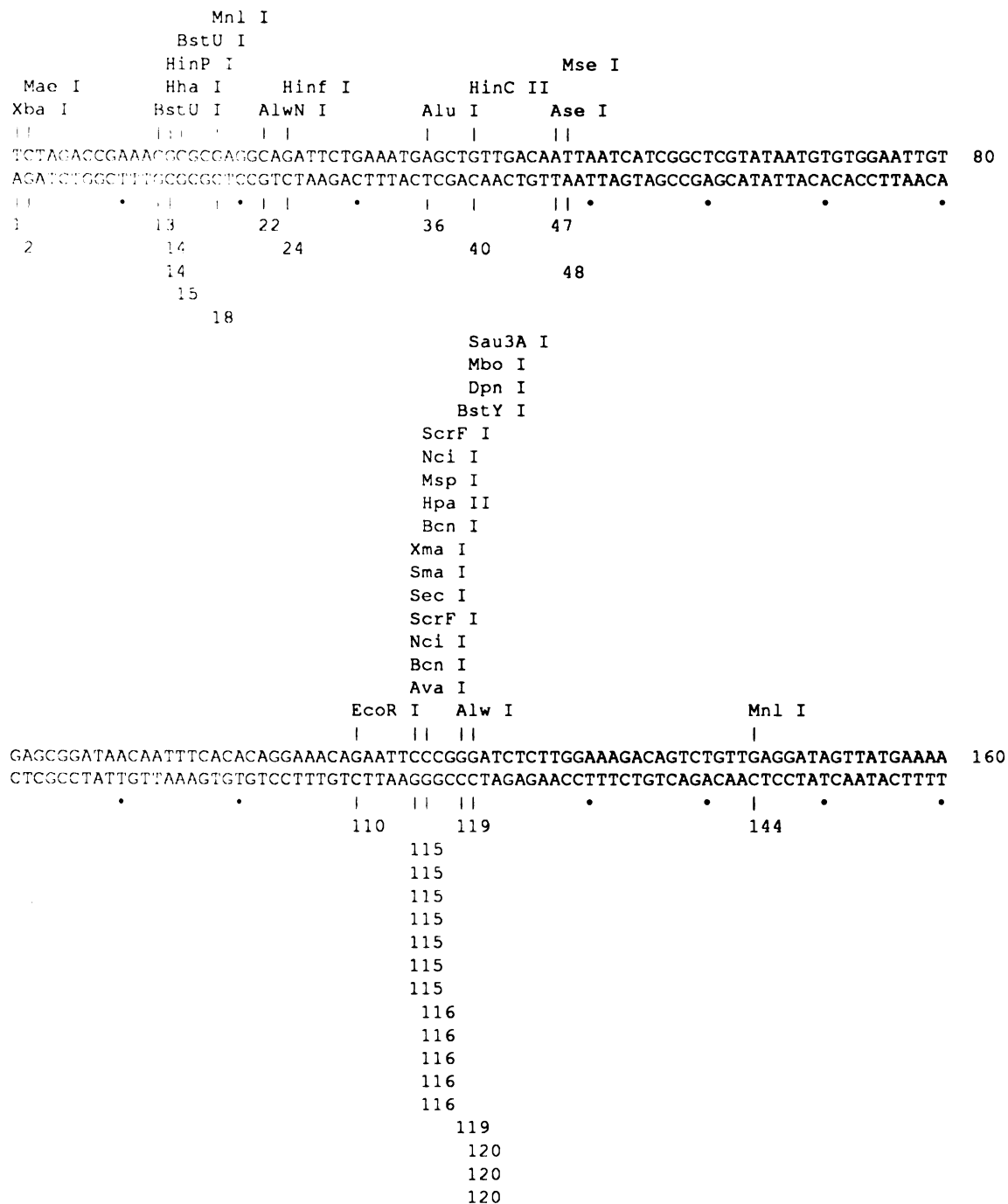
Pvu I	cgat/cg	1	1(217) 2	218(1312) 1
Rsa I	gt/aa	1	1(833) 1	834(696) 2
Sma I	ccc/ggg	1	1(114) 2	115(1415) 1
Ssp I	aat/att	1	1(1330) 1	1331(199) 2
Sty I	c/cwgg	1	1(838) 1	839(691) 2
Tth111 II	caarca 11/9	1	1(431) 2	432(1098) 1
Xba I	t/ctaga	1	1(0) 2	1(1529) 1
Xho I	c/tcga	1	1(511) 2	512(1018) 1
Xcm I	ccannnnn/nnntgg	1	1(737) 2	738(792) 1
Xma I	c/ccgg	1	1(114) 2	115(1415) 1
Xmn I	gaann/nnttc	1	1(1316) 1	1317(213) 2

No Sites found for the following Restriction Endonucleases

Aat II	gacat/c	Cfr10 I	r/ccggy	Nsp7524 I	r/catgy
Aac I	gt/mkac	Dra I	ttt/aaa	NspB II	cmg/ckg
Afl II	c/ttaag	Eag I	c/ggccg	NspH I	rcatg/y
Afl III	a/crygt	Eco57 I	ctgaag 16/14	Pml I	cac/gtg
Apa I	ggccc/c	EcoN I	cctnn/nnnagg	PpuM I	rg/gwccy
Asp718	g/gtacc	EcoO109 I	rg/gnccy	Pvu II	cag/ctg
Avr II	c/ctagg	EcoR V	gat/atc	Rsr II	cg/gwccg
BamH I	g/gatcc	Fsp I	tgc/gca	Sac I	gagct/c
Ban I	q/qyrcc	Gdi II	yggccg -5/-1	Sac II	ccgc/gg
Ban II	grqcy/c	HgiE II	accnnnnnnnggt	Sal I	g/tcgac
Bbe I	ggcgc/c	Hpa I	gtt/aac	Sca I	agt/act
Bgl II	a/gatct	Kpn I	ggtac/c	Sfi I	ggcnnnn/nggcc
BsaA I	yac/gtr	Mlu I	a/cgcgt	SnaB I	tac/gta
Bsm I	g/atgc 1/-1	Nae I	gcc/ggc	Spe I	a/ctagt
BspB I	t/catga	Nar I	gg/cgcc	Sph I	gcatg/c
BspM II	t/ccgga	Nco I	c/catgg	Spl I	c/gtacg
BssB II	g/ccgcg	Nhe I	g/ctagc	Stu I	agg/cct
BstB I	tt/cgaa	Not I	gc/ggccgc	Tth111 I	gacn/nngtc
BstE II	g/gtnacc	Nru I	tcg/cga	Xca I	gta/tac
Bsu36 I	cc/tnagg	Nsi I	atgca/t		

Appendix Two: Restriction Map of Trypsin Phage Display Vector Xba I / Hind III Fragment of pMTsn::VIII

Positions of Restriction Endonucleases sites (unique sites underlined>



Mme I
 HinP I Sau3A I
 Hha I Fnu4H I Mbo I
 Hae II Bbv I Dpn I
 Bsr I Eco47 III Pvu I Mnl I BspM I
 | | | | |
 AACTGGCGTTATCCCTTTCTCTGGTGCTGGCATTTCAGCGCTACCGCAGCATTGCGATCGTTGGAGGATACACCTGC 240
 TTGACCGCAATAGGGAAAGAGACCACGACCGTAAAAGGTCCGGATGGCGTCGTAAACGCTAGCAACCTCCTATGTGGACG
 | • • • | | • | • | •
 162 199 208 218 227 235
 199 208 219
 200 208 219
 200 219
 223
 Nla IV PflM I
 BstX I Mnl I Mnl I Bsr I
 | | | |
 CAAGAGAATTCTGTTCCTACCAAGTGTCCTGAAGTCTGGCTACCACTTCTGTGGAGGTCCCTCATCAATGACCAGTG 320
 GTTCTCTTAAGACAAGGGATGGTTCACAGGGACTTGAGACCGATGGTGAAGACACCTCCAAGGGAGTAGTTACTGGTCAC
 • • • • | • | • | •
 246 285 296 303 315
 298 315
 Alu I
 Fnu4H I
 Bbv I
 Fok I Bsr I HgiA I
 Pst I SfaN I BsmA I Bsp1286 I Mnl I
 | | | | |
 GGTGGTGCTGTCAGCTCACTGCTATAAGTCCCGCATCCAAGTGAGACTGGGAGAGCACAACATCAATGTCCTTGAGGGCA 400
 CCACCACAGACGTCGAGTGACGATATTCAGGGCGTAGGTTCACTCTGACCTCTCGTGTTGTAGTTACAGGAACCTCCCGT
 • • • • | • | • | •
 329 353 363 373 394
 331 354 366 373
 331
 333
 Sau3A
 Mbo I
 Dpn I
 Bcl I
 Fnu4H I Tth111 II Mbo II
 Bbv I Dpn I SfaN I Taq I Bbv II Nla III
 | | | | | |
 ATGAGCAGTTTGTCAATGCTGCCAAGATCATCAAGCATCCCAACTTCGATAGGAAGACCCTGAACAATGACATCATGCTG 480
 TACTCGTCAAACAGTTACGACGGTTCTAGTAGTTCGTAGGGTTGAAGCTATCCTTCTGGGACTTGTACTGTAGTAGCAG
 • • | • | • | • | • | •
 418 426 435 446 453 474
 418 432 453 479
 426 436 480
 426 480
 480
 480
 Msc I
 Hae I
 Eae I
 Pst I
 BspM I
 Mbo II Xho I Bgl I Alu I HgiA I
 Ear I PaeR7 I Mbo II Bsp1286 I Dde I
 Alu I Ava I Hae III Ear I AlwN I ApaL I Dra III
 | | | | | |
 ATCAAGCTCTCTTCCCTCTGAAACTCAATGCTCGAGTGCCACTGTGGCTCTTCCCAGCTCCTGTGCACCTGCAGGGCAC 560
 TAGTTCTGAGAGAAGCGGACACTTTGAGTTACGAGCTACCGGTGACACCGAGAAGGGTCGAGGACACGTGGACGTCCCGT
 | | • • | | | • | • | |
 485 512 519 530 537 545 558
 489 512 531 545 560
 490 512 520 538 545

513 518 518 518 549 551
 518 518 518
 ScrF I
 EcoR II
 PflM I
 Bsr I SfaN I
 Dde I Hga I Gsu I BstN I
 Esp I Aha II BspM I AlwN I Fok I
 Mnl I
 TCAGTGCCTCATCTCTGGGTGGGGCAACACGCTCAGCAGTGGCGTCAATGAACCAGACCTGCTCCAGTGCCTGGATGCCC 640
 AGTCACGGAGTAGAGACCGACCCCGTTGTGCGAGTCTGTCACCGCAGTTACTTGGTCTGGACGAGGTCACGGACCTACGGG
 567 591 601 617 625 633
 592 602 622 630
 624 634
 624 630
 630
 ScrF I Sau3A I
 EcoR II
 BstN I Mbo I Nla III Mnl I
 Fnu4H I
 Bbv I Alu I Mnl I Sec I Dpn I Drd I Mae I
 CACTGCTGCCCCAAGCTGACTGTGAAGCCTCCTACCCCTGGAAGATCACTGACAACATGGTCTGTGTTGGCTTCTTAGAG 720
 GTGACGACGGGGTTCGACTGACACTTCGGAGGATGGGACCTTTCTAGTGACTGTTGTACCAGACACAACCGAAGGATCTC
 645 654 668 675 684 691 715
 645 676 684 696 718
 676 684
 Mae III
 Hph I
 Xcm I Ple I
 Sec I Hinf I
 ScrF I
 EcoR II Sau96 I
 Mnl I Hinf I BstN I Hae III Alu I BsmA I
 GGAGGCAAGGATTCTGCCAGGGTGACTCTGGTGGCCCTGTGGTCTGTAATGGAGAGCTGCAGGGCATTGTCTCTCTGGGG 800
 CCTCCGTTCTTAAGGACGGTCCCACTGAGACCACCGGACACCAGACATTACCTCTCGACGTCCCCTAACAGAGGACCCC
 722 730 738 754 776 790
 738 754 777 794
 738 777 794
 738 745 778 794
 738 745 794
 742
 743
 ScrF I
 EcoR II Sty I
 BstN I Sec I Hinf I Fnu4H
 I
 Bsp1286 I Sec I Rsa I Bsr I Bbv I
 CTATGGCTGTGCCCTGCCAGATAACCCCTGGTGTGTACACCAAGGTCTGCAACTATGTGGACTGGATTTCAGGACACAATTG 880
 GATACCGACACGGGACGGTCTATTGGGACCACATGTGGTTCCAGACGTTGATACACCTGACCTAAGTCCTGTGTAAAC
 809 825 834 860 880
 826 839 864 880
 826 839
 826

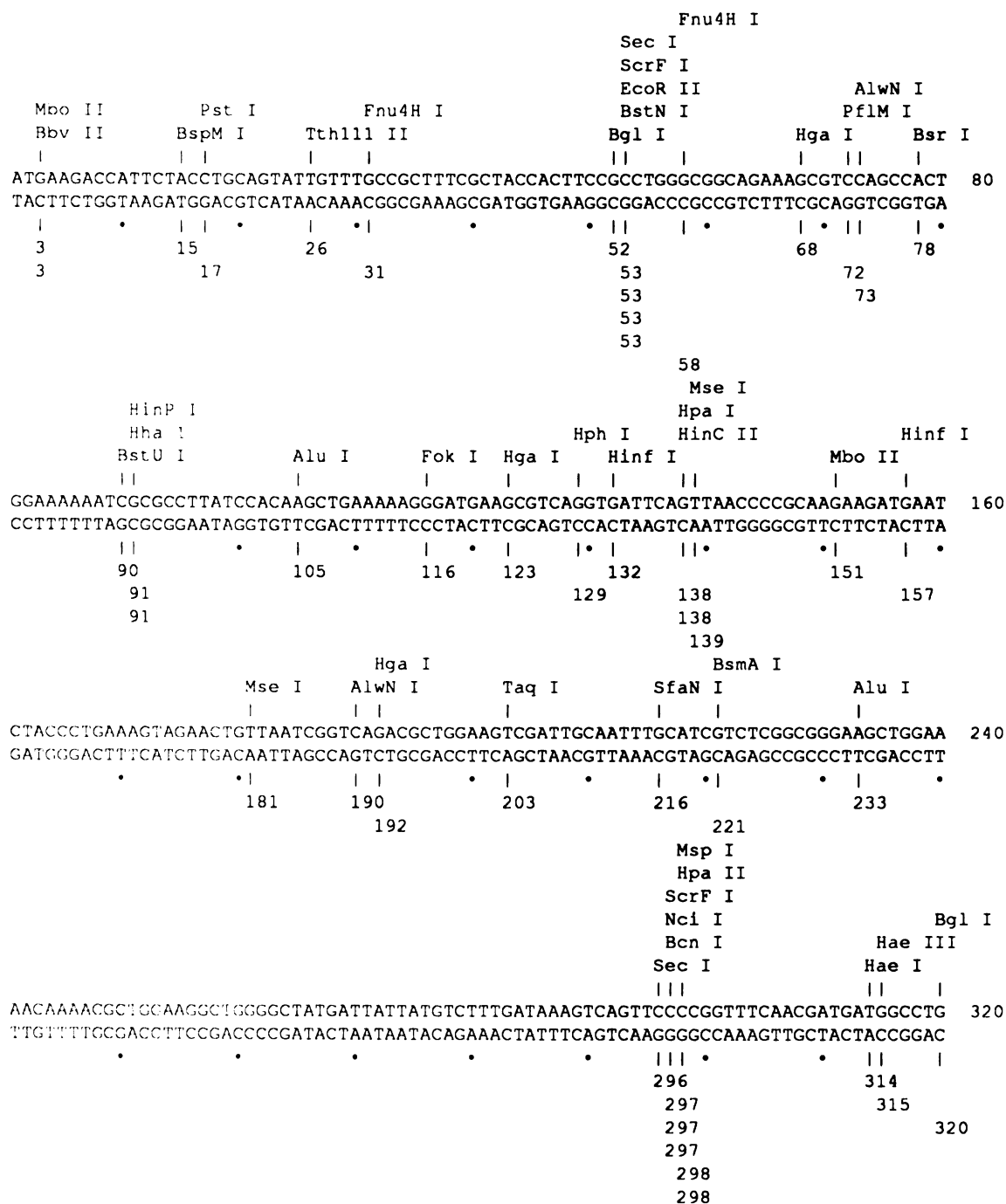
HinD III	a/agctt		1	1 (1046)	1	1047 (6)	2
Hpa II	c/cgg		1	1 (115)	2	116 (937)	1
Mme I	tcctac	20/18	1	1 (222)	2	223 (830)	1
Msc I	tgg/cca		1	1 (517)	2	518 (535)	1
Msp I	c/cgg		1	1 (115)	2	116 (937)	1
Nla IV	ggn/ncc		1	1 (297)	2	298 (755)	1
PaeR7 I	c/tcgag		1	1 (511)	2	512 (541)	1
Ple I	gagtc	4/5	1	1 (744)	1	745 (308)	2
Pvu I	cgat/cg		1	1 (217)	2	218 (835)	1
Rsa I	gt/ac		1	1 (833)	1	834 (219)	2
Sau96 I	g/gncc		1	1 (753)	1	754 (299)	2
Sma I	ccc/ggg		1	1 (114)	2	115 (938)	1
Sty I	c/cwgg		1	1 (838)	1	839 (214)	2
Tth111 II	caarca	11/9	1	1 (431)	2	432 (621)	1
Xba I	t/ctaga		1	1 (0)	2	1 (1052)	1
Xho I	c/tcgag		1	1 (511)	2	512 (541)	1
Xcm I	ccannnnn/nnnttg		1	1 (737)	1	738 (315)	2
Xma I	c/ccggg		1	1 (114)	2	115 (938)	1

No Sites found for the following Restriction Endonucleases

Aat II	gacgt/c		Cla I	at/cgat		Nsi I	atgca/t
Acc I	gt/mkac		Dra I	ttt/aaa		Nsp7524 I	r/catgy
Afl II	c/ttaag		Dsa I	c/crygg		NspB II	cmg/ckg
Afl III	a/crygt		Eag I	c/ggccg		NspH I	rcatg/y
Apa I	gggcc/c		Eco57 I	ctgaag	16/14	Pml I	cac/gtg
Asp718	g/gtacc		EcoN I	cctnn/nnnagg		PpuM I	rg/gwccy
Ava II	g/gwcc		EcoO109 I	rg/gnccy		Pvu II	cag/ctg
Avr II	c/ctagg		EcoR V	gat/atc		Rsr II	cg/gwccg
BamH I	g/gatcc		Fsp I	tgc/gca		Sac I	gagct/c
Ban I	g/gyrcc		Gdi II	yggccg	-5/-1	Sac II	ccgc/gg
Ban II	grgcy/c		HgiE II	accnnnnnnnggt		Sal I	g/tcgac
Boe I	ggcgc/c		Hpa I	gtt/aac		Sca I	agt/act
Bgl II	a/gatct		Kpn I	ggtac/c		Sfi I	ggccnnnn/ngggc
BsaA I	yac/qtr		Mae II	a/cgt		SnaB I	tac/gta
Bsm I	gaatgc	1/-1	Mlu I	a/cgcgt		Spe I	a/ctagt
BspH I	t/catqa		Nae I	gcc/ggc		Sph I	gcatg/c
BspM II	t/ccqga		Nar I	gg/cgcc		Spl I	c/gtacg
BssH II	g/cgcgc		Nco I	c/catgg		Ssp I	aat/att
BstB I	tt/cgaa		Nde I	ca/tatg		Stu I	agg/cct
BstE II	g/gtnacc		Nhe I	g/ctagc		Tth111 I	gacn/nngtc
Bsu36 I	cc/tnagg		Not I	gc/ggccgc		Xca I	gta/tac
Cfr10 I	r/ccagy		Nru I	tcg/cga		Xmn I	gaann/nnttc

Appendix Three: Restriction Map of Ecotin Phage Display Vector pBS eco-glII

Positions of Restriction Endonucleases sites (unique sites underlined)



Fok I
 Msp I
 Hpa II
 ScrF I
 Nci I
 Bcn I
 || |
 CCCGGATGGCAAGAAAGAGAAGAAATTTGTCACCGCGTATCTGGGCGATGCTGGAATGCTGCGTTACAACAGCAAGCTGC 400
 GGGCCTACCGTTCTTTCTCTCTTTAAACAGTGGCGCATAGACCGCTACGACCTTACGACGCAATGTTGTCGTTTCGACG
 || | . | . || | . | . | . | . || | .
 321 339 349 367 378 395
 321 350 374 383 396
 321 354 378 396
 322
 322
 324
 Sau3A I
 Mbo I
 Dpn I Xca I
 Pvu I Acc I
 || |
 CGATCGTGGTGTATACGCCAGACAATGTAGATGTGAAGTACCGCGTCTGGAAGGCGGAAGAGAAAATTGACAACGCGGTG 480
 GCTAGCACCACATATGCGGTCTGTTACATCTACACTTCATGGCGCAGACCTTCCGCTTCTCTTTAACTGTTGCGCCAC
 || | . | . | . || | . | . | . | .
 401 411 438 457 474
 402 411 442 457
 402 443
 402
 Sau96 I
 Mnl I
 I
 BstU I Ava II
 || |
 GTTCGCGGAGGCGGTCCATTCGTTTGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGG 560
 CAAGCGCCTCCGCCAGGTAAGCAAACACTTATAGTTCCGGTTAGCAGACTGGACGGAGTTGGAGGACAGTTACGACCGCC
 | | . | . | . || | . | | . | . | |
 484 493 516 530 541 557
 488 517 535 560
 493
 Mnl I
 Dde I
 Fnu4H I
 || ||
 CGGCTCTGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAGG 640
 GCCGAGACCACCACCAAGACCACCGCCGAGACTCCCACCACCGAGACTCCCACCGCCAAGACTCCCACCGCCGAGACTCC
 . | | . | | . | | . | | . | | .
 584 605 620 629 635
 590 607 622 637
 592
 Hph I
 Msp I
 Nla IV
 Mnl I Hpa II
 || ||
 GAGGCGGTTCCGGTGGTGGCTCTGGTCCGGTGATTTTGATTATGAAAAGATGGCAAACGCTAATAAGGGGGCTATGACC 720
 CTCCGCCAAGGCCACCACCGAGACCAAGGCCACTAAACTAATACTTTTCTACCGTTTGGCATTATTCCCCGATACTGG
 | | | . | | | . | . | . | .
 641 650 664 668 670
 646 650 668 670

1297
1297

1302
1302

Ase I

Fnu4H I BstU I Mbo II

Bbv I Hae III Ear I

Alu I Gdi II HinP I

Pvu II Mse I Eae I Hha I

NspB II Hinf I BstU I Mnl I Hae II

||| || | || ||| | || ||

GTCCGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGCTATTGGGCGCTCTT

1440

CAGCCCTTTGGACAGCACGGTCGACGTAATTACTTAGCCGGTTGCGCGCCCTCTCCGCCAAACGCATAAACCCGCGAGAA

• ||| || • | || • ||| • | • • || || •

1380 1393 1404 1414 1432

1380 1388 1397 1405 1433

1381 1397 1405 1433

1382 1398 1436

1382 1406 1437

1387

Fnu4H I

Bbv I Fnu4H I

Ple I HinP I Fnu4H I

Mnl I Hinf I Hha I Bbv I Alu I

| | | | | |

CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTTCGGTTCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATA

1520

GGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTAT

• • | | • • | | • •

1447 1458 1466 1481 1499

1458 1466 1481

1463 1484

1463

Nla IV

ScrF I

Nla III

EcoR II

NspH I

BstN I

Nsp7524 I Hae III Hae III

Hinf I Afl III Hae I Hae I

| | || || | |

CGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA

1600

GGCAATAGGTGTCITAGTCCCTATTGCGTCCTTTCTGTACACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTTTT

• | • • || • || | | • •

1533 1558 1571 1582

1558 1572 1583

1558 1585

1559 1585

1585

1588

BstU I

Fnu4H I Hga I

Hae III Nla IV SfaN I Taq I Mnl I

|| | | | |

GGCGCGGTTGCTGGCGGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGC

1680

GGCGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTAGCTGCGAGTTCAGTCTCCACCG

• • | • • | • | | • | •

1601 1627 1646 1658 1673

1602 1660

1604 1660

ScrF I
 ScrF I EcoR II HinP I
 EcoR II BstN I Mnl I
 BstN I Sec I Alu I Hha I Mme I Fnu4H I
 | | | | | | |
 GAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCCTGCCG
 1760
 CTTTGGGCTGTCTGATATTTCTATGGTCCGCAAAGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGC
 . . | . | | . | | . | . | .
 1706 1718 1725 1736 1748 1757
 1706 1719 1730
 1706 1719 1736
 1719
 HinP I
 Hha I
 Msp I Hae II Alu I Dde I
 Hpa II | | | |
 CTTACGGGATAGCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCAGCTGTAGGTATCTCAGTTC
 1840
 GAATGGGCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTATCGAGTGCACATCCATAGAGTCAAG
 | . . . | | . | . . | .
 1765 1802 1815 1833
 1765 1803
 1803
 HinP I
 Hha I
 HgiA I Fnu4H I Mae III
 Bsp1286 I Bbv I Msp I
 Alu I ApaL I NspB II Hpa II
 | | | | | |
 GGTGTAAGTCTGTTGGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTATCCGGTAACT
 1920
 CCACATCCAGCAAGCGAGCTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGA
 . . | . . | | | . | | .
 1861 1872 1898 1912
 1872 1900 1912
 1872 1900 1915
 1903
 1903
 Msp I
 Hpa II AlwN I
 ScrF I Fnu4H I
 Mme I Fnu4H I
 Pst I Nci I Bbv I Bsr I
 Hinf I Bcn I Bsr I Bbv I Mae III Mnl I
 . | | | | | |
 ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
 2000
 TAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCTAATCGTCTCGCTCC
 | . | | . . . | | | . | . | .
 1929 1937 1961 1968 1978 1997
 1929 1937 1965 1974
 1932 1965
 1937 1968
 1938 1969
 1938

Hae III Hae I Mae I Hinf I
 Hha I
 TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAGGACAGTATTTGGTATCTGCGC
 2080
 ATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCTATAAACCATAGACGCG
 . . . || . . . | . . . | .
 2034 2053 2077
 2035 2077
 Msp I
 Hpa II
 Sau3A I
 Mbo I
 Dpn I HgiE II
 Bsr I
 Eco57 I Mae III Alu I Alw I Tth111 II NspB II
 TCTGCTGAAGCCAGTTACGTTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAACACCACCGCTGGTAGCGGTGGTT
 2160
 AGACGACTTCGGTCAATGGAAGCCTTTTCTCAACCATCGAGAACTAGGCCGTTTGTGTTGGTGGCGACCATCGCCACCAA
 | . . . | . . . | . . .
 2085 2094 2118 2125 2132 2143
 2091 2125 2139
 2125
 2128
 2128
 Sau3A I
 Mbo I
 Dpn I
 Fnu4H I
 Bbv I Hinf I
 Tth111 II Hha I BstY I BstY I Mbo I
 Tth111 II BstU I Alw I Mbo II Dpn I Hga I
 TTTTCTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
 2240
 AAAAAAAGAGCTTCGTCGTCTAATGCGCGTCTTTTTTCTAGAGTTCTTCTAGGAACTAGAAAAGATGCCCCAGACTG
 || . . . | . . . | . . .
 2165 2185 2199 2208 2219 2238
 2171 2186 2199 2210 2219
 2174 2186 2200 2211 2219
 2174 2200 2211 2211
 2200 2211
 2211
 Hph I
 Mbo II Sau3A I
 Sau3A I Mbo I
 Mbo I Dpn I
 Dpn I Alw I
 BstY I BstY I Mse I
 Dde I Mae II Nla III BspH I Alw I Mae I Dra I
 GCTCAGTGAACGAAACTCAGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAA
 2320
 CGAGTCACCTTGCTTTTGGAGTGCAATCCCTAAAACAGTACTCTAATAGTTTTTCTAGAAAGTGATCTAGGAAAATTT
 || . . . | . . . | . . .
 2242 2261 2278 2296 2306 2315
 2264 2279 2296 2308 2316
 2297 2309
 2297 2309
 2297 2309
 2299 2309
 2302

Mse I Mse I Nla IV
 Dra I Mae III Mse I Ban I
 T TAAAAATGAACI T TAAAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
 2400
 AATTTTACTTCAAAAATTAGTTAGATTTCATATATACTCATTGTAACCAGACTGTCAATGGTTACGAATTAGTCACTCC
 2321 2334 2377 2388 2397
 2335 2399
 2399
 Sau3A I
 Mbo I
 Dpn I Ple I
 Dde I Fok I Hinf I Mnl I
 CACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAG
 2480
 GTGGATACAGTGGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTC
 2408 2431 2446 2478
 2414 2446
 2414
 2414
 Bsr I Nla IV
 Sau96 I Msp I
 Nla IV Fnu4H I BsmA I Hpa II
 Hae III Bbv I BstU I Cfr10 I
 Hph I Gsu I
 GGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCA
 2560
 CCGAATCGTAGAGCGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGT
 2493 2502 2515 2529 2536
 2493 2502 2518 2531
 2493 2532
 2497 2532
 2534
 Alu I
 ScrF I
 Hinf I
 Msp I
 Hpa II
 Bcn I
 GCGAGCGCGAAGGGGCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAG
 2640
 CCGTGGGCTTGGGGGCTGGCGTCTTACCAGGACGTTGAAATAGGCGAGGTAGGTGAGTAAATTAACAACGGCCCTTC
 2565 2572 2589 2608 2615 2622 2633
 2566 2579 2589 2612 2623 2633
 2566 2573 2633
 2579 2633
 2633
 2639

Mae II
 HinP I
 Mse I Hha I
 Mae I Bsr I Fsp I SfaN I Mae III
 | | | | | |
 CTAGAGTAAGTACCTCCG CAGTTAATAGTTTGGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACACGCTCGTCG
 2720
 GATGTCATTGATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCAGAGCAGC
 | . . | . || | . . | . | .
 2641 2658 2671 2698 2708
 2662 2672
 2672
 2677
 Sau3A I
 Mbo I
 Dpn I
 Alw I
 Nla IV
 Msp I Sau3A I Nla III
 Hpa II Mbo I
 Alu I Dpn I Mae III Nla III
 | | | | |
 TTGGATGGGCTCATTGAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGT
 2800
 AAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCA
 . . | . | . | . | . | . | .
 2739 2755 2766 2780
 2743 2755 2770
 2743 2755 2773
 2745 2773
 2773
 2773
 2773
 Sau3A I
 Mbo I
 Pvu I Fnu4H I
 Xba I Hae III
 Sal96 I Gdi II Fnu4H I
 Alu I Ava II Dpn I Eae I Nla III Bbv I
 | | | | |
 TAGGTCCTTCGGTCCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCGAGTTATCACTCATGGTTATGCCAGCACTGCATA
 2880
 ATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTAT
 | . | | | . | | | . | . | .
 2802 2811 2819 2839 2858 2868
 2811 2839 2868
 2814 2840
 2818 2841
 2819
 2819
 Hph I
 Fok I Bsr I Rsa I
 Nla III SfaN I Mae III Sca I Dde I
 | | | | |
 ATTCGCTTACTGTGATGCGATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
 2960
 TAAGAGAATGACAGTACGCTAGGCATTCTACGAAAAGACACTGACCACCTCATGAGTTGGTTCAGTAAGACTCTTATCACA
 . | . | . | . | . | .
 2894 2908 2919 2929 2948
 2899 2922 2930
 2925

Hga I
 Aha II
 Msp I
 Hpa II
 ScrF I
 Nci I
 Bcn I
 Fnu4H I
 |
 ATGCGCGGACCGAGTTGGTCTTGTCCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCAT
 3040
 TACGCGGTGGCTCAACGAGAAACGGGCGCGAGTTATGCCCTATTATGGCGCGGTGTATCGTCTTGAATTTTCACGAGTA
 |
 2963 2984 3008 3026 3033
 2984 3009 3027
 2984 3009 3033
 2985
 2985
 2987
 2988
 Bsr I
 Sau3A I
 Mbo I
 Dpn I
 BstY I
 Mao II
 Xma I
 Mbo II
 |
 CAITGGCAAAACGTTCTTCCGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTG
 3120
 GTAACCTTTTGGCAAGAAGCCCGCTTTTGGAGAGTTCTAGAATGGCGACAACCTCTAGGTCAAGCTACATTGGGTGAGCAC
 |
 3046 3054 3076 3084 3093 3102 3118
 3050 3076 3094 3107 3118
 3077 3094 3118
 3077 3094
 3077 3094
 3097
 Eco57 I
 Mbo II
 Sau3A I
 Mbo I
 Dpn I
 SfaN I
 Hph I
 Hph I
 Fnu4H I
 |
 CACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
 3200
 GTGGGTGACTAGAAAGTCGTAGAAAATGAAAGTGGTCGCAAGACCCACTCGTTTTTGTCTTCCGTTTTACGGCGTTTT
 |
 3130 3138 3151 3166 3192
 3130
 3130
 3132
 3133
 Mbo II
 Ear I
 Ssp I
 |
 AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATCTCTTCTTTTCAATATTATTGAAGCATTATCAGGGTTA
 3280
 TTCCCTTATCCCGCTGTGCCTTTACAACCTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAAT
 |
 3240 3253
 3241

Nla III
 BspH I
 BsmA I
 . .
 TTGTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
 3360
 AACAGASTACTCGGCTATGATATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTC

 3283 3335
 3286 3340
 3287 3341
 3341
 Fnu4H I
 Bbv I
 HinP I
 Hha I
 BstU I
 BstU I
 Hga I
 Fnu4H I
 Hha I
 Mae III
 Mae III
 TCGGACGTGACGGCGGCTCTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCC
 3440
 AGCGTCTAATCTGCTGAAATCGCGCGGTAATTCGCGCGCGCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGG

 3369 3382 3394 3411 3423
 3371 3385 3394 3415
 3372 3385 3395 3416
 3372 3390 3416
 3396 3418
 3418
 Msp I
 Hpa II
 Nae I
 Cfr10 I
 Mae I
 HinP I
 Hha I
 Hae II
 Hae II
 Mbo II
 Mae II
 Alu I
 AGCGGCTAGCGCGGCTCTGCTTCTTCGCTTCTTCCTTCTCGCCACGTTTCGCGGCTTTCCCGTCAAGCTCTAAA
 3520
 TCGCGTGAATCGCGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGGCAGTTCGAGATTT

 3441 3449 3469 3489 3512
 3442 3450 3495
 3442 3450 3495
 3447 3496
 3496
 Taq I
 Nla IV
 Bsp1286 I
 Ban II
 Nla IV
 Mnl I
 Nla IV
 Ban I
 Hph I
 Mae II
 Dra III
 BsaA I
 TCGGAGGCTCGCTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTTAC
 3600
 AGCGCCCGAGGGAATCCCAAGGCTAAATCAGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTG

 3525 3538 3559 3588 3598
 3525 3559 3598
 3526 3563 3599
 3565

Ple I
 Mae II
 Hae III
 Sau96 I
 3680
 3606
 3607
 3641
 3642
 3644
 3648
 3661
 3670
 3670
 Bsr I
 Hae III
 Mse I
 3760
 3684
 3749
 3759
 BstU I
 Mse I
 Alu I
 Mse I
 Mse I
 Mse I
 Ssp I
 Bgl I
 3840
 3769
 3776
 3787
 3791
 3799
 3806
 3810
 3832
 HinP I
 Hha I
 Fsp I
 Fnu4H I
 Bbv I
 3920
 3841
 3841
 3843
 3844
 3844
 3863
 3864
 3864
 3864
 3873
 3874
 3876
 3877
 3878
 Mbo II
 Ear I
 Mnl I
 Hae III
 Sau96 I
 3893
 3893
 3894
 3908
 3914
 3914
 BstU I
 HinP I
 Hha I
 BstU I
 Bsr I
 Hae III
 Gdi II
 Eae I
 3926
 3935
 3941
 3941
 3941
 3941
 3952
 3955
 3961
 3975
 3975
 3976
 3978
 3985
 3985
 3985
 3986
 3987
 3987
 3988
 Ple I
 Hinf I
 3998
 3998

Sau3A I
 Mbo I
 Dpn I
 Nla IV
 BstY I
 BamH I
 Spe I Alw I SfaN I
 Cla I Dde I Nde I
 Alw I Taq I Mnl I
 BstU I Fnu4H I
 BstX I Hae III
 Alu I Sec I Gdi II
 Sac I Sac II Eag I
 HqiA I NspB II Eae I
 Bsp1286 I Not I Xba I Mae I
 Ban II Dsa I Fnu4H I Mae I
 TCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCATCGATGCTTAGGAGGTCAT
 4080
 AGTGAATATCCCGCTTAACCTCGAGGTGGCGCCACCGCGCGGAGATCTTGATCACCTAGGTAGCTACGAATCCTCCAGTA
 • • • • • • • • • • • • •
 4019 4027 4035 4044 4055 4062 4073
 4019 4035 4043 4050 4061 4068 4078
 4019 4027 4036 4049 4056 4064
 4019 4027 4036 4055
 4020 4027 4036 4055
 4024 4037 4055
 4028 4038 4056
 4056
 4056

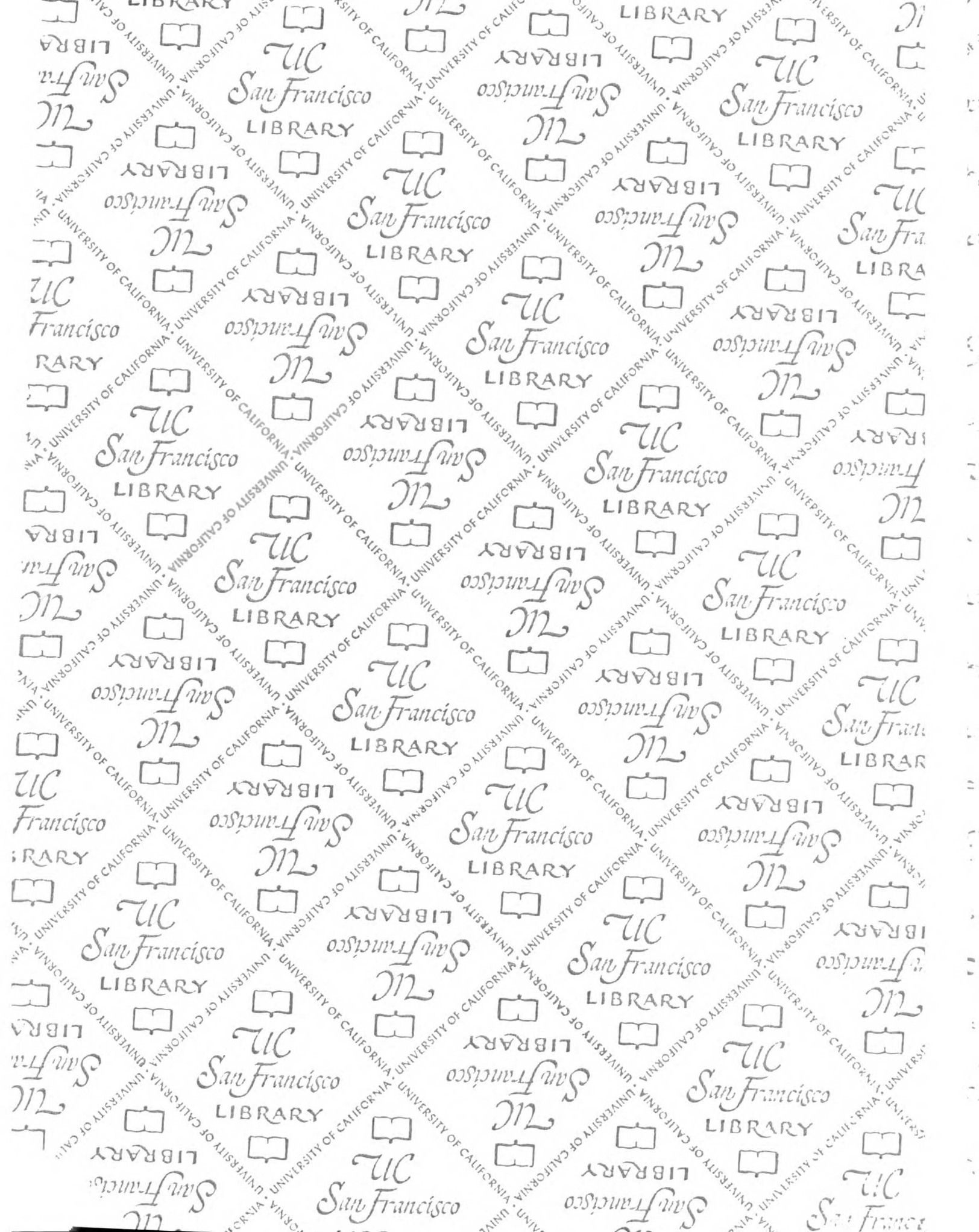
Unique Restriction Sites

Enzyme	Site	Use	Site position (Fragment length)	Fragment order
Afl III	c/ttaag	1	1121(4080)	1
Afl III	a/crygt	1	1558(4080)	1
Aha II	qr/cgyc	1	2987(4080)	1
Apa I	gggcc/c	1	1154(4080)	1
Asp718	g/gtacc	1	1160(4080)	1
Ava I	c/ycgrg	1	1145(4080)	1
BamH I	g/gatcc	1	4055(4080)	1
Bbv II	qaagac	2/6	3(4080)	1
BsaA I	yac/qtr	1	3598(4080)	1
Bsm I	gaatgc	1/-1	374(4080)	1
BstX I	ccanncnn/atgg	1	4024(4080)	1
Dra III	ccanncn/gtg	1	3598(4080)	1
Eag I	c/ggccg	1	4036(4080)	1
Eco47 III	agc/gct	1	982(4080)	1
EcoO109 I	rg/gnccy	1	1153(4080)	1
Gsu I	ctggag	16/14	2536(4080)	1
HqiE II	accnnnnnnnggt	1	2139(4080)	1
HinD III	a/agctt	1	1124(4080)	1
Hpa I	gtt/aac	1	138(4080)	1
Kpn I	ggtac/c	1	1160(4080)	1
Nae I	gcc/ggc	1	3495(4080)	1
Not I	gc/ggccgc	1	4035(4080)	1
Nsp7524 I	r/catgy	1	1558(4080)	1
NspH I	rcatg/y	1	1558(4080)	1
PaeRV I	c/tcgag	1	1145(4080)	1
PfIM I	ccanncnn/atgg	1	72(4080)	1
Pst I	ctgga/g	1	17(4080)	1
Sac I	gagct/c	1	4019(4080)	1
Sac II	ccgc/gg	1	4027(4080)	1
Sal I	g/tcgac	1	1139(4080)	1
Sca I	agt/act	1	2929(4080)	1
Spe I	a/ctaqt	1	4049(4080)	1
Xba I	t/ctaga	1	4043(4080)	1
Xca I	gta/tac	1	411(4080)	1

Xho I c/tcgag 1 1145 (4080) 1

No Sites found for the following Restriction Endonucleases

Aat II	gacgt/c	EcoR V	gat/atc	Rsr II	cg/gwccg
Avr II	c/ctagg	Esp I	gc/tnagc	Sfi I	ggccnnnn/nggcc
Bbe I	ggcgc/c	Mlu I	a/cgcgt	Sma I	ccc/ggg
Bcl I	t/gatca	Msc I	tgg/cca	SnaB I	tac/gta
Bgl II	a/gatct	Nar I	gg/cgcc	Sph I	gcatg/c
BspM II	t/ccgga	Nco I	c/catgg	Spl I	c/gtacg
BstB I	tt/cgaa	Nhe I	g/ctagc	Stu I	agg/cct
BstE II	g/gtnacc	Nru I	tcg/cga	Sty I	c/cwwgg
Bsu36 I	cc/tnagg	Nsi I	atgca/t	Tth111 I	gacn/nngtc
EcoN I	cctnn/nnnagg	Pml I	cac/gtg	Xcm I	ccannnnn/nnntgg
EcoR I	g/aattc	PpuM I	rg/gwccy	Xma I	c/ccggg



For reference

Not to be taken
from the room.

LIBRARY