NATURAL AND ARTIFICIAL EVOLUTION IN THE SERINE PROTEASES OF *STREPTOMYCES GRISEUS*

by

Sachdev Singh Sidhu

B.Sc. (First Class Honours) Simon Fraser University, 1990

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in the Department of Chemistry

SIMON FRASER UNIVERSITY

June 1996

APPROVAL

Name:	Sachdev Singh Sidhu
Degree:	Doctor of Philosophy (Science)
Title of Thesis:	Natural and Artificial Evolution in the Serine Proteases of *Streptomyces griseus.*
Examining Comittee:	Chair: Dr. S. Holdcroft

Dr. T. J. Borgford
Associate Professor of Chemistry
Senior Supervisor

Dr. A. J. Bennet
Assistant Professor of Chemistry
Committee Member

Dr. B. M. Pinto
Professor of Chemistry
Committee Member

Dr. J. K. Scott
Assistant Professor of Chemistry
Internal Examiner

Dr. D. A. Agard
Professor of Biochemistry and Biophysics
University of California, San Francisco
External Examiner

Date Approved: JUNE 14, 1996

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay:

NATURAL AND ARTIFICIAL EVOLUTION IN THE SERINE PROTEASES OF STREPTOMYCES GRISEUS

Author:

(signature)

SACHDEV SIDHU

(name)

JULY 30, 1996

(date

ABSTRACT

The natural diversity of the *Streptomyces griseus* serine protease family was investigated; the insights gained were used to design enzymes with altered specificities. Genes encoding *S. griseus* protease E and two previously uncharacterized proteases (*S. griseus* proteases C and D, SGPC and SGPD) were isolated and sequenced. A shuttle vector was constructed to allow for genetic manipulations in *Escherichia coli* and protein secretion in *Bacillus subtilis*; the recombinant gene products were purified and characterized. The two newly characterized enzymes possessed domains unique among bacterial proteases: SGPC contained a carboxyl-terminal domain homologous to chitin binding domains while SGPD possessed an unusually long amino-terminal prepeptide with characteristics of mitochondrial import signals. Furthermore, recombinant SGPD existed as a very stable homodimer.

The *S. griseus* proteases are secreted in a promature form, and autocatalytic removal of the propeptide is essential for activity. *S. griseus* protease B (SGPB, a previously characterized enzyme) has primary specificity for large hydrophobic substrates. We genetically substituted the Leu residue at the promature junction P1 site of SGPB with various amino acids; substitution with poor SGPB substrates reduced or abolished active protease secretion in *E. coli.* Thus, it was demonstrated that the specificity of SGPB is constrained by the promature junction sequence; active protease production is dependent on the efficiency of autocatalytic processing. The link between substrate specificity and proteolytic activity was used to select enzymes with altered primary specificities: loss of proteolytic activity due to mutations at the promature junction P1 site could be restored by compensatory changes in enzyme specificity. Sequence alignments and crystallographic data were used to design an *E. coli* expression

library containing 29,952 possible SGPB mutants with variations at seven sites involved in conferring primary specificity. A visual screening strategy allowing for the rapid analysis of up to 10^6 individual variants was developed to detect active protease secretion. The expression library was screened in conjunction with various promature junction sequences; variants producing increased proteolytic activity were isolated. Variant enzymes with increased specificity towards various hydrophobic and polar substrates were isolated.

DEDICATION

To my parents.

ACKNOWLEDGMENTS

I thank my supervisory committee:
Dr. T. J. Borgford, Dr. A. J. Bennet, and Dr. B. M. Pinto.

I also thank Dr. G. B. Kalmar and Mr. L. G. Willis for contributions to the second chapter and Dr. C. A. J. Braun for contributions to the third chapter.

QUOTATIONS

A FOUR STAGE PH.D.

I Illusion

All I ever wanted
was simply just to be me.
All you need is the truth
and the truth will set you free.

(V. Morrison, 1991)

II Disillusion

Business men they drink my wine
ploughmen dig my earth.
None of them along the line
know what any of this is worth.

(B. Dylan, 1968)

III Resolution

I know that I might make you nervous.
But I tell you I can't live in service.
Like a doctor whose born for a purpose.
Rudy can't fail.

(The Clash, 1979)

IV Absolution

Leader of the new school
uncool.
Never played the fool
just made the rules.

(Public Enemy, 1988)

TABLE OF CONTENTS

Page

CHAPTER 2: NATURAL EVOLUTION IN THE SERINE PROTEASES OF *STREPTOMYCES GRISEUS*

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 3

ABBREVIATIONS

α-LP	α-lytic protease of *Lysobacter enzymogenes*
amp	ampicillin
bp	base pair(s)
cfu	colony forming unit(s)
CIP	calf intestinal phosphatase
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
ddH_2O	deionized, distilled water
dGTP	2'-deoxyguanosine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediaminetetraacetic acid
FPLC	fast protein liquid chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
kbp	kilobase pair(s)
kDA	kilodalton(s)
OD	optical density
OD_{600}	optical density at 600 nm
OMTKY3	turkey ovomucoid inhibitor third domain
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pI	isoelectric point
pmP1	the promature junction P1 site
S1 family	chymotrypsin family
S2 family	α-lytic protease family
S8 family	subtilisin family

SA clan	chymotrypsin clan
SB clan	subtilisin clan
SDS	sodium dodecyl sulfate
SGPA	*Streptomyces griseus* protease A
SGPB	*Streptomyces griseus* protease B
SGPC	*Streptomyces griseus* protease C
SGPD	*Streptomyces griseus* protease D
SGPE	*Streptomyces griseus* protease E
SGT	*Streptomyces griseus* trypsin
sprA	the gene encoding SGPA
sprB	the gene encoding SGPB
sprC	the gene encoding SGPC
sprD	the gene encoding SGPD
sprE	the gene encoding SGPE
Suc-AAPX-pNA	*N*-succinyl-Ala-Ala-Pro-X-*p*-nitroanilide
Tris	tris(hydroxymethyl)methylamine
UMF	univeral M13 forward primer
UMR	univeral M13 reverse primer
YT	yeast extract/tryptone
YTC	yeat extract/tryptone/$CaCl_2$

CHAPTER 1: AN INTRODUCTION TO SERINE PROTEASES

The serine peptidase (or serine protease) enzyme class includes all proteolytic enzymes with catalytic activity dependent on a serine residue. The members are numerous and widespread; the class is among the best studied in terms of mechanism and structure. This chapter provides an introduction to serine protease classification and gives a brief overview of previous studies on catalytic mechanism and substrate specificity.

1 Serine protease families and clans

Recently, a classification scheme grouping serine proteases into families and clans (or superfamilies) has been proposed (Rawlings & Barrett, 1994). Enzymes shown to be evolutionarily related on the basis of sequence comparisons are grouped together into families. Families in turn are grouped together into clans; a clan contains families which are believed to have descended from a common ancestor but have diverged to such an extent that the relationships between families are uncertain. Thus, all members of a given clan are believed to have arisen from a common ancestor through divergent evolution. In contrast, members from separate clans represent distinct evolutionary lines which have converged upon a similar catalytic mechanism.

The serine proteases are grouped into over 20 families which can be further grouped into six clans. The work in this thesis involves the chymotrypsin clan (SA clan) and, to a lesser extent, the substilisin clan (SB clan); therefore, only these two clans will be discussed.

1.1 Chymotrypsin clan (SA)

The SA clan includes ten families of endopeptidases (proteases which preferentially cleave the inner regions of peptide chains). The catalytic activity of clan members is dependent on not only the catalytic serine but also two other invariant residues (an aspartate and a histidine) which together form a "catalytic triad" (see section 2). Furthermore, the order of catalytic residues in the SA clan primary structure is invariably His/Asp/Ser. The two SA clan families which concern this thesis are described below.

1.1.1 Chymotrypsin family (S1)

In terms of both sequence and structure information, the S1 family is the best characterized peptidase family. The members are almost exclusively eukaryotic, but there are a few exceptions (including *Streptomyces griseus* trypsin (Awad et al., 1972; Kim et al., 1991)). In animals, many important functions (including digestion, hormone activation, and blood clotting) are dependent upon S1 family members (Neurath, 1984).

The enzymes are translated in a pre-pro-mature form. The prepeptide directs secretion from the cell and thus, S1 family members are inherently targeted to the secretory pathway (Rawlings & Barrett, 1994). The propeptide (which may be as small as two amino acids) functions to hold the enzyme in an inactive state. Intermolecular cleavage at the promature junction allows the new amino terminus to form a salt bridge with a particular aspartate residue; the end result is the ordering of the catalytic site into an active conformation (Read & James, 1988).

In some cases, the propeptide remains attached to the mature enzyme through disulfide bonds as an amino-terminal domain adapting particular enzymes to specialized tasks. However, the core catalytic unit is a polypeptide of approximately 220 amino acids (Rawlings & Barrett, 1994). Aside from two exceptions (granzyme A (Simon & Kramer, 1994) and tryptase (Schwartz, 1994)), the enzymes are monomers.

1.1.2 α-lytic protease family (S2)

S2 family members are exclusively prokaryotic. The S2 protease catalytic domain (approximately 190 amino acids) is smaller than that of S1 family members, and sequence homology between the two families is slight. However, members from the two families are largely superimposable at the tertiary structure level; most sequence insertions and deletions occur in variable surface loops. Thus, it has been concluded that the S1 and S2 proteases arose from a common ancestor (Delbaere et al., 1975).

S2 family members are produced in a pre-pro-mature form analogous to that of the S1 proteases (Henderson et al., 1987; Silen et al., 1988). In prokaryotes, the prepeptide signals secretion from the cell; therefore, all known S2 proteases are extracellular enzymes. The S2 propeptide is always extensive (rivaling the mature enzyme in length) and its function represents a fundamental difference between the S1 and S2 enzymes. The S1 propeptide plays a passive role in protein folding, merely blocking the amino terminus. In contrast, The S2 propeptide participates in the folding mechanism as a catalyst, lowering the energy barrier from the folded to the unfolded state (Silen & Agard, 1989). Enzyme activation involves a presumably autocatalytic

cleavage at the promature junction, but it has not been conclusively proven whether the event is intermolecular or intramolecular.

1.2 Subtilisin clan (SB) and family (S8)

The subtilisin clan contains only one family. It is the second largest family of serine peptidases, and contains both prokaryotic and eukaryotic members. Bacterial subtilisins are extracellular enzymes with broad substrate specificities well suited to non-specific protein degradation. Eukaryotic subtilisins, on the other hand, form a distinct subfamily (the kexin subfamily) responsible for proprotein processing (Seidah & Chretien, 1994); these enzymes preferentially cleave peptide bonds following paired basic amino acids.

Crystallographic studies (Poulos et al., 1976; Bode et al., 1986) have identified a catalytic triad containing the same residues as that of the chymotrypsin clan. Triads from the two clans are virtually superimposable in three-dimensional space; however, the order of catalytic residues in the subtilisin clan primary structure (Asp/His/Ser) differs from that in the chymotrypsin clan (His/Asp/Ser). Furthermore, the tertiary structures of the two clans share no similarities outside the catalytic triad. Thus, the presence of an identical catalytic triad in two distinct clans is a classic example of convergent evolution (Fersht, 1985).

2 Catalytic mechanism

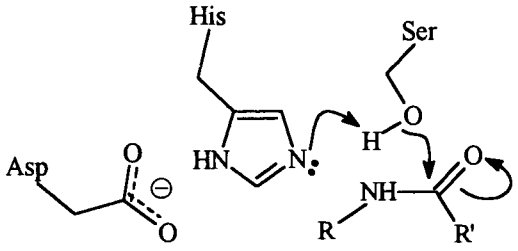
The chymotrypsin and subtilisin clans have converged on a common catalytic mechanism dependent upon the $Asp_C/His_C/Ser_C$ catalytic triad (the subscript "c" denotes a member of the catalytic triad). The mechanism has been well studied and reviewed (Kraut, 1977; Polgar, 1989).

Productive substrate binding allows nucleophilic attack by the side-chain hydroxl of Ser_C on the scissile bond carbonyl carbon; the nucleophilicity of Ser_C is greatly enhanced by His_C which accepts a proton from the attacking hydroxyl (Fig. 1-1A). The result is a negatively charged tetrahedral intermediate (Fig. 1-1B) stabilized by the "oxyanion hole", a region of the enzyme which provides a binding site for the oxyanion. The positively charged His_C ring is stabilized by the negatively charged Asp_C side-chain. Dissociation of the amine (facilitated by proton donation from His_C to the amine nitrogen) produces an acyl-enzyme complex (Fig. 1-1C).

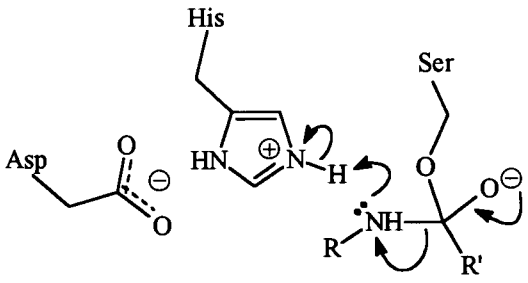
Deacylation occurs by a process analogous to the one described above except that the nucleophile is a water molecule and the leaving group is Ser_C. His_C abstracts a proton from the attacking water; a negatively charged tetrahedral intermediate is again formed (Fig. 1-1D). Proton donation from His_C facilitates the departure of Ser_C; the hydrolyzed substrate is released and the enzyme is regenerated (Fig 1-1E).

Mutational analysis of subtlisin BPN' (isolated from *Bacillus amyloliquefaciens*) has confirmed that the catalytic triad plays an essential role in enzymatic rate enhancement (Carter & Wells, 1987; Carter & Wells, 1988). Replacement of Ser_C, His_C, or Asp_C by alanine reduced the catalytic turnover rate constant (k_{cat}) by a factor of 2×10^6, 2×10^6, or 3×10^4, respectively. Simultaneous alanine replacement of all three residues caused

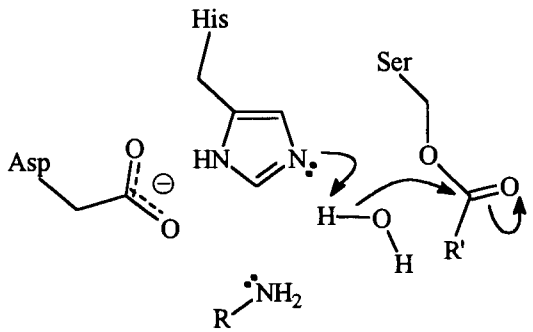
Fig. 1-1. The role of the catalytic triad in peptide hydrolysis. Details of the mechanism are discussed in the text.


His
Ser
Asp
HN
NH
R
R'

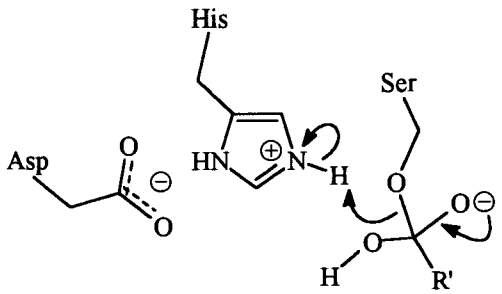
A


His
Ser
Asp
HN
NH
R
R'

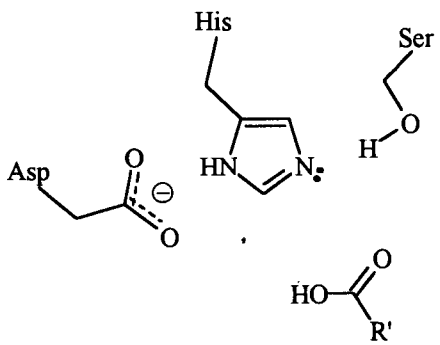
B


His
Ser
Asp
HN
R'
R
NH2

C


His
Ser
Asp
HN
R'

D


His
Ser
Asp
HN
HO
R'

E

only a 2 x 10^6 reduction in k_{cat}, demonstrating the cooperative nature of the catalytic triad. In all cases, only slight changes in K_m were observed; clearly, the catalytic triad is crucial to catalysis but does not contribute significantly to substrate binding. Notably, the triple mutant exhibited a k_{cat} value approximately 10^3-fold above the non-enzymatic rate; thus, interactions independent of the catalytic triad (such as the oxyanion hole) also contribute to catalysis.

An analogous study of rat trypsin yielded similar results (Corey & Craik, 1992).

3 Protease substrate specificity

In describing protease substrate specificity, the nomenclature of Schecter and Berger (1967) is the standard. Substrate amino acids are represented by P*n*,...,P2,P1,P1',P2',...,P*n*'; where P1-P1' is the scissile bond. The corresponding binding sites on the protease are represented by S*n*,...,S2,S1,S1',...,S*n*'.

While enzymes of the chymotrypsin and subtilisin clan exhibit preferences for substrate amino acids extending across several sites on either side of the scissile bond, the P1 residue is usually the main specificity determinant; this preference for particular side-chains immediately preceding the scissile bond is known as "primary specificity". Not surprisingly, many substrate specificity studies have focused on the interactions between the P1 substrate residue and the S1 binding pocket.

Traditionally, serine endopeptidases have been grouped into three main classes on the basis of primary specificity , with each class being named after an archetypical pancreatic digestive enzyme (Stroud, 1974). Trypsin-like

and chymotrypsin-like enzymes cleave bonds following basic or large hydrophobic residues respectively, while elastase-like enzymes exhibit primary specificity for small hydrophobic residues. Despite the discovery of new specificities (including many enzymes with specificity for acidic residues (Nienaber et al., 1993)) the classification scheme remains useful because it extends to the structural level; the specificity of each class is due to common structural features.

The crystal structures of chymotrypsin (Matthews et al., 1967), trypsin (Ruhlmann et al., 1973), and elastase (Watson et al., 1970) reveal S1 pockets which complement the primary specificity in each case. The S1 pocket of chymotrypsin is deep and hydrophobic, allowing for favourable interactions with large, hydrophobic P1 side-chains. The S1 pocket of trypsin is similar but contains an aspartic acid residue (Asp189, chymotrypsin numbering (Greer, 1990)) at the base; the pocket accommodates long arginine or lysine side-chains while the Asp189 side-chain provides a negative counter-charge for the positively charged substrate. Finally, the S1 pocket of elastase is shallow and hydrophobic due to the substitution of large aliphatic side-chains for small side-chains at corresponding positions in chymotrypsin; thus, elastase prefers small P1 side-chains.

3.1 Mutational studies of primary specificity

The structural basis of substrate specificity in the subtilisin, chymotrypsin and α-lytic protease families has been investigated using site-directed mutagenesis; a recent review highlights contributions made to our general understanding of enzyme structure and function (Perona & Craik, 1995).

3.1.1 Subtilisin

Subtilisin BPN' exhibits a broad primary specificity with a preference for large, hydrophobic residues; k_{cat}/K_m varies linearly with substrate hydrophobicity (Estell et al., 1986). Molecular modeling suggested that side-chains at position 166 (a Gly residue in the wild-type) could extend into the S1 pocket; mutants with substitutions at position 166 were constructed and kinetically characterized. Increasing the volume of the uncharged side-chain at position 166 reduced k_{cat}/K_m against large P1 substrates by up to 5,000-fold; in contrast, k_{cat}/K_m against small P1 substrates was increased by up to 10-fold. A combined optimal binding volume for the S1 and P1 side-chains was estimated to be 160 $Å^3$. If this optimal volume is exceeded, k_{cat}/K_m is reduced due to steric repulsions which outweigh the gain in hydrophobicity.

The primary specificity of subtilisin BPN' towards charged substrates has been markedly increased by engineering electrostatic complementarity between enzyme and substrate (Wells et al., 1987). Molecular modeling suggested that charged side-chains at positions 156 and 166 could stabilize oppositely charged P1 side-chains through salt bridge formation. Kinetic analysis of mutant enzymes confirmed the prediction; the introduction of a charged side-chain at either position increased k_{cat}/K_m towards oppositely charged P1 side-chains and combined effects at the two sites were roughly additive. A mutant subtilisin BPN' containing the substitutions E156Q and G166K exhibited a 1,900-fold increase in k_{cat}/K_m against a substrate with Glu at P1.

Sequence comparisons of closely related enzymes can greatly aid protein engineering studies. The tertiary structure of *Bacillus licheniformis* subtilisin Carlsberg (Bode et al., 1986) is very similar to that of subtilisin BPN'

(McPhalen & James, 1988), but the primary specificity of the enzymes towards various substrates (as measured by k_{cat}/K_m) various up to 50-fold. The sequences differ at 86 of 275 positions, but only three substitutions are within 7 Å of the S1 pocket . These three residues in subtlisin BPN' were replaced with the corresponding residues in subtilisin Carlsberg; the substrate specificity profile of the subtilisin BPN' triple mutant (E156S/G169A/Y217L) approached that of subtilsin Carlsberg (Wells et al., 1987). Similarly, subtilisin BPN' variants with increased specificity for peptide bonds following dibasic residues have been engineered using the kexin family as a guide in selecting potential mutations (Ballinger et al., 1995).

3.1.2 Mammalian trypsin

The S1 pockets of chymotrypsin, trypsin, and elastase complement the respective primary specificities; thus, it was believed that primary specificity in these enzymes is determined by a limited number of amino acids in close proximity to the P1 side-chain (Stroud, 1974). Indeed, the mutational analyses of subtilisin (section 3.1.1) supported this view, and it seemed likely that specificity in the chymotrypsin family could be predictably modulated by engineering substitutions at key sites close to the S1 pocket. However, studies with rat trypsin have proved the supposition overly simplistic; structural elements far removed in three-dimensional space impact greatly on the interactions between the S1 binding pocket and the P1 substrate residue.

Trypsin exhibits a marked preference for basic P1 side-chains. Crystallographic studies suggested that Asp189 at the base of the S1 pocket provides a counter-ion for positively charged substrate side-chains and thus is a major primary specificity determinant (Ruhlmann et al., 1973); mutational

studies have confirmed the prediction. A double mutant (D189G/G226D) was designed to move the negative counter-ion to a nearby position at the base of the pocket (Ruhlmann et al., 1973); the variant exhibited a 10-fold increase in K_m and a 10^3-fold decrease in k_{cat} against substrates with basic P1 side-chains. Crystallographic analysis revealed that only one carboxylate oxygen of Asp226 is available for interactions with substrate (Perona et al., 1993). Thus, the position of the negative charge is critical for productive interactions with substrate; non-optimal interactions result in reduced substrate binding affinities (increased K_m) and reductions in k_{cat} due to misalignment of the scissile bond with respect to the catalytic machinery. Replacement of Asp189 with Lys produced a poor, non-specific protease rather than an enzyme with primary specificity towards acidic residues (Graf et al., 1987). The result was attributed to a non-appropriate protein environment for the Lys189 side-chain. In another study, an expression library containing variant trypsins with all combinations of the natural amino acids at positions 189 and 190 was screened using a genetic selection for those members exhibiting trypsin-like activity (Evnin et al., 1990; Perona et al., 1993). The absence of a negative charge at the base of the S1 pocket reduced k_{cat}/K_m against substrates with Lys or Arg at P1 by a factor of at least 10^5. An Asp or Glu residue at position 189 or 190 partially restored activity, but the nature and position of the negative charge was again found to be critical for optimal activity. The above studies taken together not only demonstrate the importance of Asp189 but also the requirement for precise positioning of the charged side-chain and the need for proper intramolecular contacts. In other words, trypsin has evolved to position a negative charge at the base of the S1 pocket in a precise orientation conducive to efficient catalysis, and the protein environment plays a crucial role in maintaining the correct conformation.

Attempts to convert trypsin into an enzyme with chymotrypsin-like specificity have revealed that primary specificity is dependent upon structural elements far removed from the S1 pocket. Since the S1 pockets of trypsin and chymotrypsin are very similar, the conversion appeared simple. However, the replacement of Asp189 with Ser (the corresponding residue in chymotrypsin) produced a poor, nonspecific protease (Graf et al., 1988). Several other residues in contact with substrate were mutated to the analogous chymotrypsin sequence (D189S/Q192M/I138T together with insertion of Thr219), but chymotrypsin-like specificity was still not achieved. The exchange of two surface loops with the analogous loops of chymotrypsin, together with the S1 pocket mutations described above, produced an enzyme with chymotrypsin-like acylation and deacylation rates but greatly increased K_m against a substrate with Phe at the P1 position (Hedstrom et al., 1992). An additional mutation (Y172W) to the chymotrypsin sequence reduced K_m and finally resulted in a mutant trypsin with 2-15% chymotrypsin activity (Hedstrom et al., 1994). It was speculated that the mutated surface loops confer structural flexibility or cause subtle alterations in the S1 pocket conformation, but their contribution to substrate specificity remains unclear. Nevertheless, it was clearly demonstrated that primary specificity in the chymotrypsin family depends not only on residues directly in contact with substrate but also on structural regions distant from the S1 pocket.

3.1.3 The α-lytic protease

The α-lytic protease of *Lysobacter enzymogenes* (α-LP) exhibits elastase-like specificity, preferentially cleaving bonds following small hydrophobic residues (Bauer et al., 1981). In contrast, *Streptomyces griseus* protease B (SGPB, a closely related enzyme) exhibits primary specificity for large, hydrophobic amino acids (Bauer, 1978). The side-chains of residues 192 and 213 (chymotrypsin numbering, (James et al., 1978)) extend into the S1 pocket; the presence of small side-chains at these positions produces a large, hydrophobic pocket in SGPB (Read et al., 1983) while replacement with Met side-chains greatly reduces the size of the pocket in α-LP (Fujinaga et al., 1985).

Mutation of the α-LP sequence from Met to Ala at either site (*i.e.* M192A or M213A) increased primary specificity for large hydrophobic residues while maintaining high activity against substrates with Ala at P1 (Bone et al., 1989). The variant enzymes were affected in both k_{cat} and K_m but the changes could not be correlated with the size or hydrophobicity of the P1 side-chain. Furthermore, each mutation provided an equal change in side-chain volume but conferred substantially different effects. Crystallographic studies suggested that the mutated S1 pockets had gained considerable structural plasticity; alternate side-chain and main-chain conformations accommodate different P1 side-chains (Bone et al., 1989; Bone et al., 1991).

The primary specificity of α-LP has been redesigned using a relatively simple computational method (Bone & Agard, 1991). The method uses a side-chain rotamer library to assess the viability of particular side-chain substitutions; five to six different rotamer conformations are allowed for each residue. Several residues can be varied simultaneously and all different rotamer combinations are tested using a molecular dynamics force field with

a solvation term (Wilson et al., 1993). Both Leu and Ile are poor substrates for α-LP; the program selected a single mutation (M192V) predicted to enhance primary specificity for Leu while discriminating against Ile. Kinetic analysis confirmed the prediction; the mutant protease exhibited a dramatic increase in k_{cat}/K_m against a substrate with Leu at P1 and significantly less activity against an analogous Ile substrate (Wilson et al., 1991).

Recently, a surface loop far removed from the S1 pocket was subjected to alanine scanning mutagenesis (Mace et al., 1995). Each of the 19 amino acids within the loop was individually mutated to alanine and the effects on primary specificity were assessed. Mutations at sites as far as 21 Å from the catalytic serine significantly altered k_{cat} and K_m. The position of mutations within the loop did not affect specificity in a predictable manner; the effects were again attributed to alterations in active site plasticity.

The primary specificity of α-LP has also been probed through random mutagenesis of four residues in the S1 pocket (Graham et al., 1993). An expression library containing the mutation M192A in conjunction with all possible amino acid combinations at positions 192a, 192b, 213, and 217a was screened using synthetic substrates. Qualitative characterization of active variants revealed that the majority retained a specificity profile similar to the parent. However, a distinct subclass with increased primary specificity for His was identified; all members of the subclass contained the mutation M213H.

CHAPTER 2: NATURAL EVOLUTION AMONG THE SERINE PROTEASES OF *STREPTOMYCES GRISEUS*

1 INTRODUCTION

The dependence of many important biological processes (including digestion, hormone activation, and blood clotting) on proteases of the chymotrypsin clan (SA clan) (Neurath, 1984) (Rawlings & Barrett, 1994) underscores the need to better understand how these enzymes have become adapted to diverse activities. Prokaryotic and eukaryotic members of the SA clan are believed to have descended from a common ancestor but are sufficiently divergent to be placed into distinct groups; the chymotrypsin family (S1) and the α-lytic protease family (S2) contain the majority of eukaryotic and prokaryotic SA clan members, respectively. The gram positive soil bacterium *Streptomyces griseus* is unique in that it produces members of not only the prokaryotic S2 family (*S. griseus* proteases A, B, and E; SGPA, SGPB and SGPE) but also a member of the almost exclusively eukaryotic S1 family (*S. griseus* trypsin, SGT). Thus, *S. griseus* provides an ideal opportunity for studying the evolution of structural and functional diversity among serine proteases of the SA clan.

Previous studies of *S. griseus* proteases proceeded along similar lines. Hydrolytic activities toward specific synthetic substrates were first detected in Pronase (a commercial preparation of *S. griseus* culture supernatants). The enzymes were then isolated on the basis of their activity toward that substrate; primary structure determinations provided the information necessary to design oligonucleotides that could be used as either radiolabeled probes or polymerase chain reaction (PCR) primers for the isolation of the

corresponding genes. This approach had resulted in the isolation of four *S. griseus* SA clan members (Awad et al., 1972; Yoshida et al., 1988): SGPA, SGPB, SGPE, and SGT. The genes encoding SGPA, SGPB (Henderson et al., 1987), and SGT (Kim et al., 1991) had been cloned and sequenced.

Due to inherent limitations, it seemed unlikely that the approach outlined above had determined the full breadth and divergence of the *S. griseus* SA clan proteases. Intracellular proteases would be absent from Pronase while the presence of non-constitutively expressed extracellular proteases would be dependent on growth conditions. Furthermore, a Pronase enzyme with a narrow substrate range may very well escape detection with commonly used synthetic substrates; such had been the case with SGPE (Yoshida et al., 1988).

We adopted a strategy for cloning protease genes which was independent of the gene product but depended rather on similarities at the DNA level. The degree of homology between the genes encoding the known *S. griseus* S1 family proteases suggested that members of the family should cross-hybridize with one another. *S. griseus* genomic DNA was probed with a portion of the gene encoding SGPB; genes encoding SGPA, SGPB, SGPE and two hitherto uncharacterized proteases were detected. This chapter describes the cloning and sequencing of the genes encoding SGPE (Sidhu et al., 1993) and the two new enzymes which we designated *S. griseus* proteases C and D (SGPC and SGPD) (Sidhu et al., 1994; Sidhu et al., 1995). In addition, we also constructed an *Escherichia coli/Bacillus subtilis* shuttle vector (pEB11) allowing for genetic manipulations in *E. coli* and recombinant protein secretion in *B. subtilis*. Purification and characterization of the recombinant protease gene products is reported.

2 MATERIALS AND METHODS

2.1 Enzymes and reagents

T7 DNA polymerase was purchased from Pharmacia and calf intestinal alkaline phosphatase (CIP) was from Boehringer Mannheim Biochemicals. Vent DNA polymerase (New England Biolabs) was used for all polymerase chain reactions (PCR). All other enzymes for digesting or modifying DNA were purchased from New England Biolabs, Inc. or Bethesda Research Laboratories, Inc.. Enzymes were used in accordance with the recommendations of the supplier.

Ampicillin, kanamycin, Suc-AAPF-pNA, and Suc-AAPL-pNA were from Sigma; Suc-AAPX-pNA (where X = A, E, I, M, or V) were from Bachem. [α-^{32}P]dATP (~3,000 Ci/mmol) was from Amersham. Zeta-probe and Immobilon-p membranes were purchased from Biorad and Millipore, respectively. The AP-1 cation exchange column was from Waters; all other protein chromatography supplies were from Pharmacia. X-ray film was from Kodak. All chemicals and reagents were of the highest grade commercially available.

2.2 Strains and plasmids

S. griseus IMRU3499 was obtained from the Waksman Institute of Microbiology, Rutgers University, Piscataway, N.J. *E. coli* DH5α/P3 was a gift from Dr. R. Kay, Terry Fox Laboratories, Vancouver, Canada. The plasmid pPT30 and *B. subtilis* DB104 (Kawamura & Doi, 1984) were gifts from Professor A.R. Fersht, Cambridge University, U.K..

2.3 Media and Growth conditions

Growth of *S. griseus* mycelium for the isolation of DNA was as described previously (Hopwood et al., 1985). *E. coli* was transformed as described previously (Sambrook et al., 1989) and grown on YT medium (Miller, 1972) containing 200 μg/mL of ampicillin (YT/amp medium). Protoplasts of *B. subtilis* were prepared by lysozyme treatment, transformed with plasmid DNA, and selected for resistance to kanamycin as described previously (Chang & Cohen, 1979). *B. subtilis* transformants were grown on YT medium containing 50 μg/mL of kanamycin and 1.5% w/v skim milk powder.

For protease expression in *B. subtilis*, a special medium (YTC broth) was prepared as follows. Activated charcoal (30 g/L) was added to a solution containing 35 g/L tryptone, 2 g/L yeast extract, and 5 g/L NaCl. This mixture was stirred for 15 minutes and then centrifuged for 20 minutes at 10,000 x g. The supernatant was decanted and filtered through Whatman No. 1 filter paper using a buchner funnel. The charcoal treated media was then autoclaved, and kanamycin (20 μg/mL) and $CaCl_2$ (10 mM) were added from sterile stock solutions prior to inoculation. Cultures were grown with vigorous shaking at 30 °C.

2.4 Oligonucleotides

The following oligonucleotides (listed from 5' to 3') were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer:

BF1: GGTACCTGCGCAACCCCCCGGACG

CF1: AGTACTGACCCCGCGGCCACC

CR1: GAATTCGGGACCGGGGCGG

DF1: AGTACTAGTGACGATGTACCG

DR1: GAATTCGCTCCGGCCGGTGTACCG

EF1: AGCTCCGGCGGTTCGT

EF2: CCCGGGGCCGACACCCCGCCGGCC

ER1: CAGTTGCAGTGGCAG

ER2: CCGCACTGCCACATGACTAGATCT

P206: GGCTCCACCACCCAGGT

P207: GGCGGCACGACGTACTT

P252: AACTACCCGACCTTCAT

P332: TCCGGCCGTCACGAAGTA

SF1: CGGATCCGATATCGGAAACAAGCGAGTG

SF2: CTCTAGACCCGGGGGCCTGGGCAGAG

UMF: GTAAAACGACGGCCAGT

UMR: AACAGCTATGACCATG

2.5 Isolation of DNA

Chromosomal DNA was isolated from *S. griseus* as described previously (Hopwood et al., 1985) . *E. coli* plasmid DNA was purified by an alkaline lysis procedure (Sambrook et al., 1989) and *B. subtilis* plasmid DNA was purified by a modified alkaline lysis procedure (Lovett & Amulos, 1989). DNA fragments and vectors were gel purified for each cloning step by established methods (Sambrook et al., 1989).

2.6 DNA sequencing

DNA was sequenced by a combination of manual and automated methods. Manual sequencing involved the use of [α-^{32}P]dATP in accordance with protocols provided in Pharmacia's "^{T7}Sequencing Kit". Severe compressions in the DNA sequence caused by regions of high secondary structure were alleviated by the addition of dimethylsulfoxide to the sequencing reaction and by the substitution of 7-deaza-dGTP and 7-deaza-dATP for dGTP and dATP. Automated sequencing involved the use of an Applied Biosystems 370A DNA Sequencer with reagents and protocols supplied by the manufacturer.

2.7 Southern blots

DNA was size fractionated on agarose gels. The gels were soaked in 0.5 M NaOH, 1.5 M NaCl for 40 min to denature DNA and then in 1.0 M Tris, 1.5 M NaCl (pH 8.0) for 40 min to neutralize the solution. Southern blot transfer of DNA from agarose gels to Zeta-probe filters was performed as described previously (Sambrook et al., 1989).

Alternatively, size fractionated DNA was centrifugally transferred (Wilkins & Snell, 1987) from agarose gels to Zeta-probe membranes at 1000 rpm in an International Equipment Co. HN-S centrifuge equipped with a microtiter plate rotor.

In all Southern Blot procedures, DNA was covalently cross-linked to Zeta-probe membranes using a Stratagene UV Stratalinker 2400 in accordance with the manufacturer's instructions. Probe hybridization is described in section 2.9.

2.8 Colony blots

A previously described colony hybridization method (Vogelli & Kaytes, 1987) was used to select clones from *S. griseus* genomic libraries. Libraries were plated on nitrocellulose discs overlaid on YT/amp plates and incubated for 12 h at 37 °C. The nitrocellulose discs were removed and overlaid with Zeta-probe discs; the discs were separated, each disc was placed colony side up on a fresh YT/amp plate and incubated 5 h at 37 °C. The nitrocellulose master plates were stored at 4 °C; the Zeta-probe plates were used for colony hybridization.

Each Zeta-probe disc was wet with 2X SSC/SDS (0.30 M NaCl, 30 mM sodium citrate, 5% SDS, pH 7.0) for 2 min, microwaved (600 Watts) for 2 min, and allowed to dry completely. The discs were washed for 2 h at 42 °C with shaking (wash buffer: 50 mM Tris, 1 M NaCl, 1 mM EDTA, 0.1% SDS, pH 8.0) and then hybridized with radioactive probe as described in section 2.9.

2.9 Hybridization

^{32}P-labeled hybridization probes were prepared according to protocols provided in New England Biolabs' "Random Priming System I". Filters were hybridized with labeled probe for 18 h at 42 °C (Hybridization buffer: 5% SDS, 50% formamide, 1.0 mg/mL BSA in a 0.50 M sodium phosphate buffer at pH 7.0) as described previously (Vogelli & Kaytes, 1987). Filters were subsequently washed at 55 °C (wash buffer: 0.15 M NaCl, 15 mM sodium citrate, 0.3-1.0% SDS, pH 7.0) and hybridization was detected by autoradiography.

2.10 Characterization of *sprE* (the gene encoding SGPE)

2.10.1 Construction of a genomic library

S. griseus IMRU3499 chromosomal DNA was digested to completion with *Bgl*II and size fractionated on a 1.0% agarose gel. DNA fragments ranging in size from 0.5 to 12 kilobase pairs (kbp) were isolated and cloned into pUC18 (Norrander et al., 1983) which had been linearized with *Bam*HI and treated with CIP. The *S. griseus Bgl*II fragments (0.2 μg) and linearized pUC18 (0.2 μg) were ligated in a final volume of 10 μL and the ligation mixture was used to transform *E. coli* DH5α/P3.

2.10.2 Cloning of *sprE*

Two oligonucleotides (EF1 and ER1) having sequences complementary to the previously reported partial DNA sequence of SGPE (Svendsen et al., 1991) were used as PCR primers; a 234 bp DNA fragment (E-mat), encoding amino acids 44 to 121 of mature SGPE, was amplified from *Bgl*II-digested *S. griseus* genomic DNA. E-mat was radiolabeled and used as a hybridization probe to select clones from the *S. griseus* genomic library (section 2.10.1) for further analysis.

Plasmid DNA was purified from individual selected clones and digested with appropriate restriction endonucleases. The digested plasmid DNA was electrophoresed on a 1.0% agarose gel and centrifugally transferred to Zeta-probe membranes. Filters were hybridized with the E-mat probe; four plasmids, each containing a 4.7 kbp insert, were found to hybridize strongly

with the probe. One of these plasmids, designated pDS-E, was chosen for further characterization.

2.10.3 Sequence analysis of *sprE*

pDS-E subclones used to determine the sequence of *sprE* are summarized in Table 2-1. The sequencing strategy is diagrammed in Fig. 2-1.

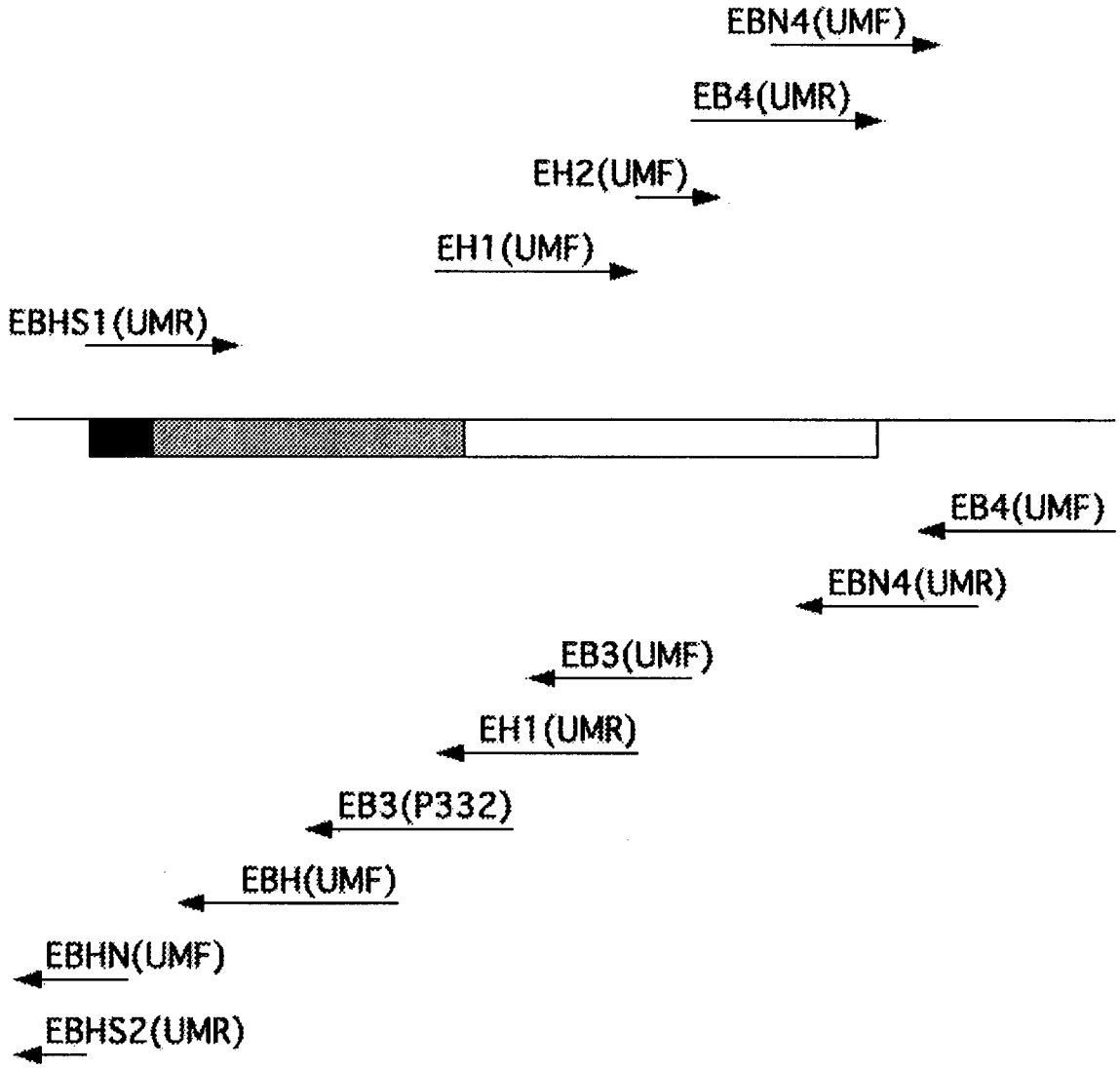
pDS-E restriction fragments were selected for sequencing on the basis of Southern blot hybridization with the E-mat probe; all plasmids were transformed into *E. coli* DH5α/P3. pDS-E was digested with *Bam*HI and two strongly hybridizing fragments were ligated into pUC18 digested with *Bam*HI and treated with CIP; the resulting plasmids containing 1.2 kbp and 0.56 kbp inserts were designated pDS-EB3 and pDS-EB4, respectively. pDS-E was digested with *Hinc*II and the 0.27 kbp and 0.12 kbp fragments were ligated into pUC18 digested with *Hinc*II and treated CIP; the resulting plasmids were designated pDS-EH1 and pDS-EH2, respectively. pDS-EB3 was digested with *Hinc*II and religated to produce the plasmid pDS-EBH; pDS-EBH digested with *Hinc*II and *Nae*I was religated to produce pDS-EBHN. pDS-EBH was digested with *Sst*I and religated to produce pDS-EBHS1; the liberated 0.38 kbp fragment was ligated into pUC18 digested with *Sst*I and treated with CIP to produce pDS-EBHS2

Table 2-1. Subclones used in the sequencing of *sprE*

	UMF site[b]		UMR site[b]		
pDS subclone[a]	pUC18	insert	pUC18	insert	sequence[c]
EB3	*Bam*HI	*Bam*HI (916)	*Bam*HI	*Bam*HI (5')	916-693 (UMF) 673-393 (332)
EB4	*Bam*HI	*Bam*HI (1477)	*Bam*HI	*Bam*HI (916)	916-1173 (UMR) 1482-1218 (UMF)
EBH	*Hinc*II	*Hinc*II (468)	*Bam*HI	*Bam*HI (5')	517-218 (UMF)
EBHN	*Hinc*II	*Nae*I (153)	*Bam*HI	*Bam*HI (5')	153-1 (UMF)
EBHS1	*Hinc*II	*Hinc*II (468)	*Sst*I	*Sst*I (97)	97-303 (UMR)
EBHS2	*Bam*HI	*Bam*HI (5')	*Sst*I	*Sst*I (97)	97-1 (UMR)
EBN4	*Sma*I	*Nae*I (1023)	*Sma*I	*Nae*I (1299)	1023-1249 (UMF) 1299-1053 (UMR)
EH1	*Hinc*II	*Hinc*II (570)	*Hinc*II	*Hinc*II (840)	570-840 (UMF) 840-570 (UMR)
EH2	*Hinc*II	*Hinc*II (840)	*Hinc*II	*Hinc*II (957)	840-957 (UMF)

[a] Construction of subclones is described in section 2.10.3. [b] Description of ligation site closest to the universal M13 forward (UMF site) or reverse (UMR site) priming site of pUC18. Numbers in parantheses indicate the positions of the restriction sites within the known sequence of *sprE*; "5'"indicates a site upstream of the known sequence (Figure 2-5).
[c] Nucleotides of known *sprE* sequence obtained from each subclone; parantheses refer to the sequencing primer used in each case.

Fig. 2-1. Strategy for the sequencing of *sprE*. The pre-pro-mature open reading frame is shown as follows: preregion, *solid box;* proregion, *shaded box;* and mature region, *open box.* Arrows above or below the open reading frame represent sequence obtained from the coding or non-coding strand, respectively; the length of each arrow is proportional to the amount of sequence obtained. Each arrow is labeled with the subclone and sequencing primer (in parantheses) used to obtain the sequence. Subclones are described in detail in Table 2-1 and section 2.10.3.


EBN4(UMF)
EB4(UMR)
EH2(UMF)
EH1(UMF)
EBHS1(UMR)
EB4(UMF)
EBN4(UMR)
EB3(UMF)
EH1(UMR)
EB3(P332)
EBH(UMF)
EBHN(UMF)
EBHS2(UMR)

25b

2.11 Characterization of *sprC* (the gene encoding SGPC)

2.11.1 Construction of a genomic library

S. griseus IMRU3499 chromosomal DNA was digested to completion with *Bam*HI and size fractionated on a 1.0% agarose gel. DNA fragments ranging in size from 3.6 to 5.2 kilobase pairs (kbp) were isolated and cloned into pUC18 which had been linearized with *Bam*HI and treated with CIP. The *S. griseus Bam*HI fragments (0.2 μg) and linearized pUC18 (0.2 μg) were ligated in a final volume of 10 μL and the ligation mixture was used to transform *E. coli* DH5α/P3.

2.11.2 Cloning of *sprC*

A DNA fragment encoding amino acids 9-185 of mature SGPB (B-mat) (Henderson et al., 1987) was radiolabeled and used as a hybridization probe to select clones from the genomic library (section 2.11.1).

Plasmid DNA purified from individual selected clones was digested with appropriate restriction endonucleases, electrophoresed on a 1.0% agarose gel, and centrifugally transferred to Zeta-probe membranes. Filters were hybridized with the B-mat probe; three plasmids, each containing a 4.8 kbp insert, were found to hybridize strongly with the probe. One of these plasmids was designated pDS-C and chosen for further characterization.

2.11.3 Sequence analysis of *sprC*

pDS-C subclones used to determine the sequence of *sprC* are summarized in Table 2-2. The sequencing strategy is diagrammed in Fig. 2-2.

pDS-C restriction fragments were selected for sequencing on the basis of Southern blot hybridization with the B-mat probe. All fragments were ligated into pUC18 digest with *Sma*I and treated with CIP; all plasmids were transformed into *E. coli* DH5α/P3. pDS-C was digested with *Pvu*II and a strongly hybridizing 1.1 kbp fragment was ligated into pUC18; the resulting plasmid was designated pDS-CP. pDS-CP was digested with *Sma*I and the 0.40 kbp fragment was ligated into pUC18 to produce the plasmid pDS-CPS2. pDS-C was digested with *Nae*I and the strongly hybridizing 0.84 kbp fragment was ligated into pUC18 to produce the plasmid pDS-CN2. pDS-C was digested with *Sma*I and two strongly hybridizing fragments were ligated into pUC18; the resulting plasmids containing 0.62 kbp and 0.47 kbp inserts were designated pDS-CS3 and pDS-CS4, respectively.

2.12 Characterization of *sprD* (the gene encoding SGPD)

2.12.1 Construction of a genomic library

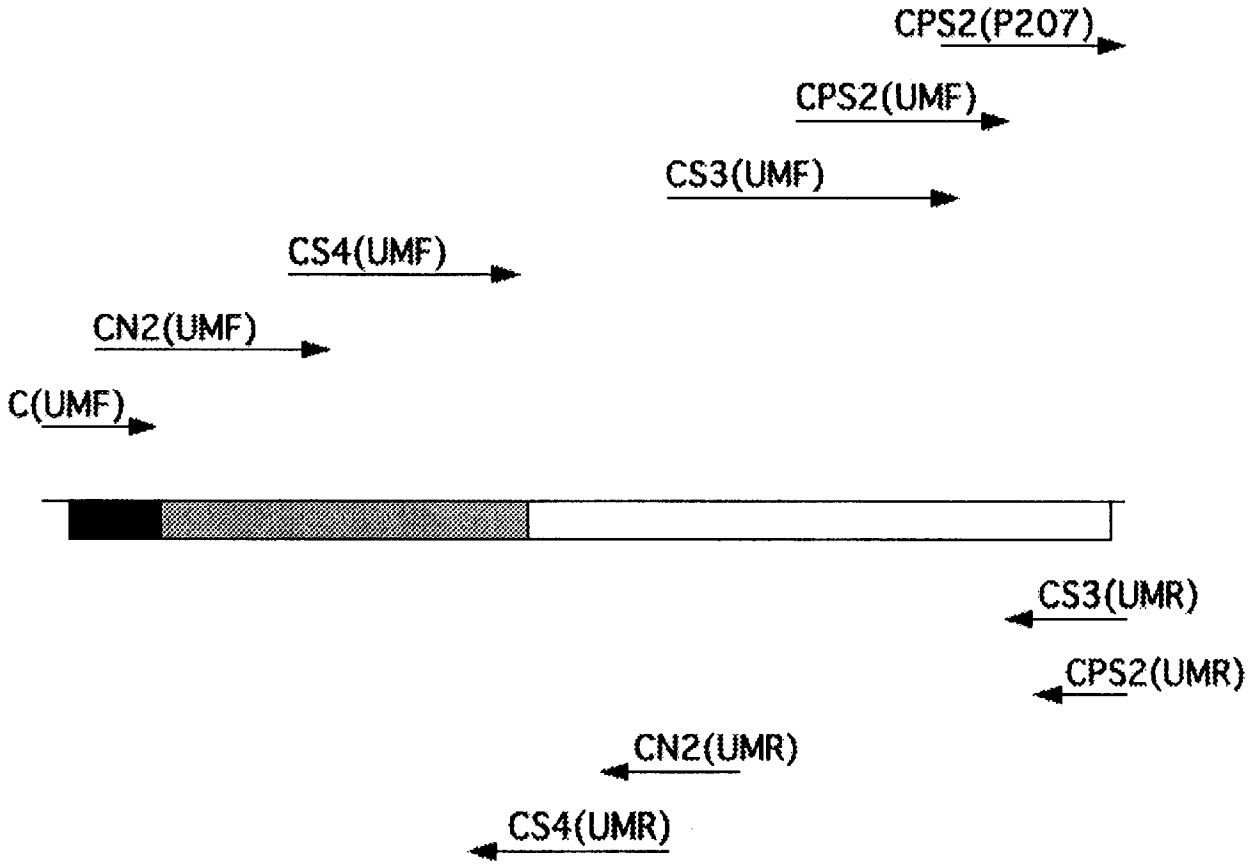
A genomic library was constructed as described in section 2.11.1 except that DNA fragments ranging in size from 2.0 to 2.6 kbp were used as inserts.

Table 2-2. Subclones used in the sequencing of *sprC*

	UMF site[b]		UMR site[b]		
pDS subclone[a]	pUC18	insert	pUC18	insert	sequence[c]
C	*Bam*HI	*Bam*HI (1)	*Bam*HI	*Bam*HI (3')	1-150 (UMF)
CN2	*Sma*I	*Nae*I (73)	*Sma*I	*Nae*I (913)	73-379 (UMF) 913-731 (UMR)
CPS2	*Sma*I	*Pvu*II (994)	*Sma*I	*Sma*I (1421)	994-1276 (UMF) 1421-1303 (UMR) 1183-1421 (207)
CS3	*Sma*I	*Sma*I (822)	*Sma*I	*Sma*I (1421)	822-1208 (UMF) 1421-1265 (UMR)
CS4	*Sma*I	*Sma*I (323)	*Sma*I	*Sma*I (822)	323-630 (UMF)

[a] Construction of subclones is described in section 2.11.3. [b] Description of ligation site closest to the universal M13 forward (UMF site) or reverse (UMR site) priming site of pUC18. Numbers in parantheses indicate the positions of the restriction sites within the known sequence of *sprC*; "3'" indicates a site downstream of the known sequence (Figure 2-6).
[c] Nucleotides of known *sprC* sequence obtained from each subclone; parantheses refer to the sequencing primer used in each case.

Fig. 2-2. Strategy for the sequencing of *sprC*. The pre-pro-mature open reading frame is shown as follows: preregion, *solid box*; proregion, *shaded box*; and mature region, *open box*. Arrows above or below the open reading frame represent sequence obtained from the coding or non-coding strand, respectively; the length of each arrow is proportional to the amount of sequence obtained. Each arrow is labeled with the subclone and sequencing primer (in parantheses) used to obtain the sequence. Subclones are described in detail in Table 2-2 and section 2.11.3.


CPS2(P207)
CPS2(UMF)
CS3(UMF)
CS4(UMF)
CN2(UMF)
C(UMF)
CS3(UMR)
CPS2(UMR)
CN2(UMR)
CS4(UMR)

29b

2.12.2 Cloning of *sprD*

The genomic library (section 2.12.1) was screened as described in section 2.11.2. Nine plasmids, each containing a 2.3 kbp insert, were isolated. One of these plasmids, designated pDS-D, was chosen for further characterization.

2.12.3 Sequence analysis of *sprD*

pDS-D subclones used to determine the sequence of *sprD* are summarized in Table 2-3. The sequencing strategy is diagrammed in Fig. 2-3.

pDS-D restriction fragments were selected for sequencing on the basis of Southern blot hybridization with the B-mat probe. All fragments were ligated into pUC18 digested with *Sma*I and treated with CIP; all plasmids were transformed into *E. coli* DH5α/P3. pDS-D was digested with *Hinc*II and two strongly hybridizing fragments were ligated into pUC18; the resulting plasmids containing 0.46 kbp and 0.35 kbp inserts were designated pDS-DH2 and pDS-DH3, respectively. pDS-D was digested with *NaeI* and religated to produce pDS-DN1; the liberated 1.05 kbp and .41 kbp fragments were ligated into pUC18 to produce the plasmids pDS-DN2 and pDS-DN3, respectively.

2.13 Recombinant protease secretion in *B. subtilis*

2.13.1 Construction of pEB11 (an *E. coli*/*B. subtilis* shuttle vector)

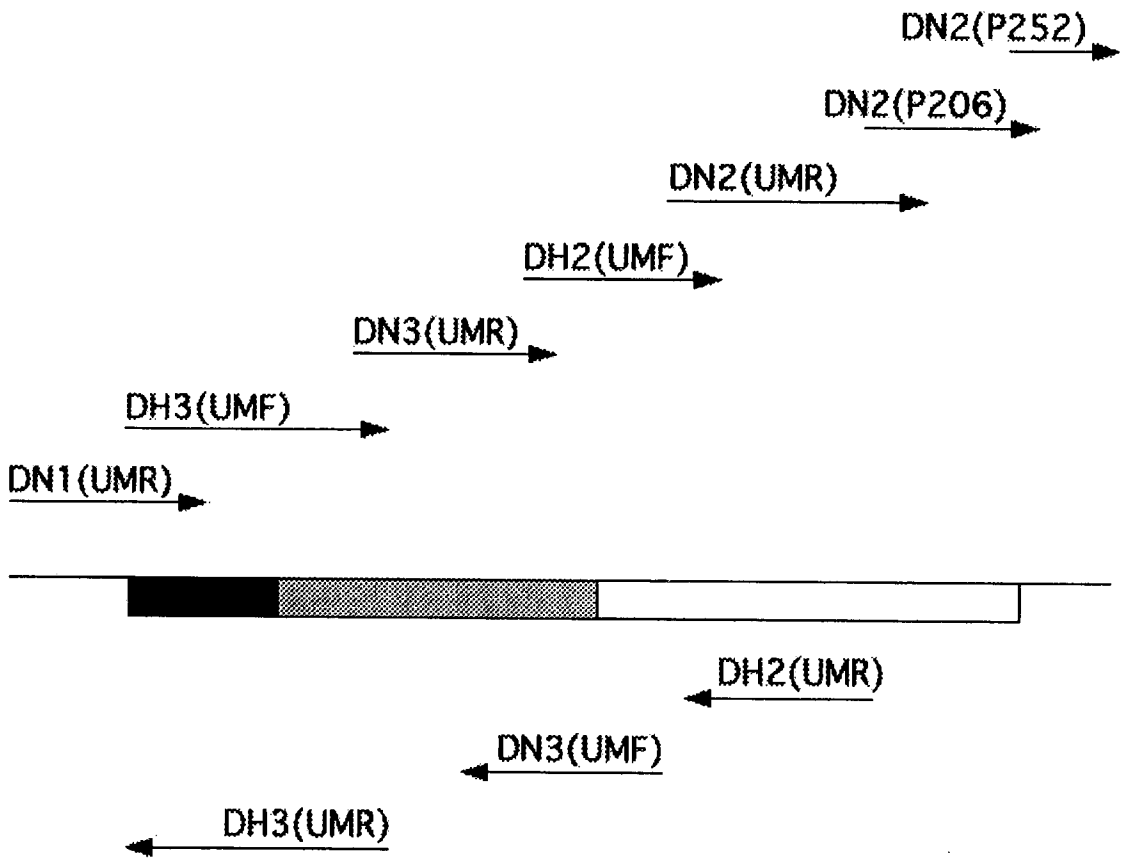
A 3.4 kbp fragment of pUB110 (digested with *Eco*RI and *Pvu*II and blunt-ended with T4 DNA polymerase) and a 2.3 kbp fragment of pUC18 (digested with PvuII and treated with CIP) were gel purified and ligated (0.2 μg

Table 2-3. Subclones used in the sequencing of *sprD*

	UMF site[b]		UMR site[b]		
pDS subclone[a]	pUC18	insert	pUC18	insert	sequence[c]
DH2	*Sma*I	*Hinc*II (680)	*Sma*I	*Hinc*II (1136)	680-938 (UMF) 1136-890 (UMR)
DH3	*Sma*I	*Hinc*II (155)	*Sma*I	*Hinc*II (500)	155-500 (UMF) 500-155 (UMR)
DN1	*Bam*HI	*Bam*HI (3')	*Bam*HI	*Bam*HI (1)	1-263 (UMR)
DN2	*Sma*I	*Nae*I (3')	*Sma*I	*Nae*I (866)	866-1210 (UMR) 1125-1353 (206) 1312-1457 (252)
DN3	*Sma*I	*Nae*I (866)	*Sma*I	*Nae*I (452)	452-722 (UMR) 866-598 (UMF)

[a] Construction of subclones is described in section 2.12.3. [b] Description of ligation site closest to the universal M13 forward (UMF site) or reverse (UMR site) priming site of pUC18. Numbers in parantheses indicate the positions of the restriction sites within the known sequence of *sprD*; "3'"indicates a site downstream of the known sequence (Fig. 2-8).
[c] Nucleotides of known *sprD* sequence obtained from each subclone; parantheses refer to the sequencing primer used in each case.

Fig. 2-3. Strategy for the sequencing of *sprD*. The pre-pro-mature open reading frame is shown as follows: preregion, *solid box*; proregion, *shaded box*; and mature region, *open box*. Arrows above or below the open reading frame represent sequence obtained from the coding or non-coding strand, respectively; the length of each arrow is proportional to the amount of sequence obtained. Each arrow is labeled with the subclone and sequencing primer (in parantheses) used to obtain the sequence. Subclones are described in detail in Table 2-3 and section 2.12.3.


DN2(P252)
DN2(P206)
DN2(UMR)
DH2(UMF)
DN3(UMR)
DH3(UMF)
DN1(UMR)
DH2(UMR)
DN3(UMF)
DH3(UMR)

32b

of each fragment) in a final volume of 10 μL. The ligation mixture was used to transform *E. coli* DH5α/P3. A recombinant plasmid (pEB1) was selected by restriction enzyme analysis such that both the ampicillin resistance gene of pUC18 and the kanamycin resistance gene of pUB110 had the same polarity.

Based on the sequence of the *Bacillus amyloliquefaciens* subtilisin gene (subtilisin BPN') (Wells et al., 1983) two oligonucleotides (SF1 and SR1) were synthesized and used as PCR primers with *Eco*RI-digested pPT30 (a plasmid carrying the subtilisin BPN' gene) as template. The 330 bp PCR-amplified fragment encompassing a region of the subtilisin gene extending from 210 bp upstream of the initiation codon to the last codon of the preregion was isolated. After an initial cloning step involving pUC18, the PCR generated fragment was blunt-end ligated into pEB1 digested with *Pvu*II and treated with CIP. The ligation mixture was used to transform *E. coli* DH5α/P3. A plasmid in which the subtilisin preregion has the same polarity as the antibiotic resistance genes was selected by restriction analysis; the plasmid was designated pEB11.

2.13.2 Construction of vectors for protease secretion

All ligations were performed in final volumes of 10 μL and were transformed into *E. coli* DH5α/P3.

2.13.2.1 Construction of pEB-B8 (SGPB secretion)

BF1 and UMR were used as PCR primers with pNC-B (a pUC19 derivative containing *sprB*) (Carson, 1994) as template; the amplified product (encoding promature SGPB) was digested with *Pst*I and ligated into pUC18

(digested with *Pst*I and *Sma*I and treated with CIP). Vectors with inserts of the correct size were identified by restriction analysis and were partially sequenced to verify accurate amplification and ligation. A plasmid containing a correctly amplified fragment of *sprB* was designated pDS-B8.

pDS-B8 was digested with *Fsp*I and *Pst*I, and the fragment encoding the promature portion of SGPB was gel-purified and ligated into pEB11 digested with *Sma*I and *Pst*I and treated with CIP. A vector containing the correct insert in the correct orientation was isolated on the basis of restriction enzyme analysis and was designated pEB-B8. pEB-B8 was transformed into *B. subtilis* DB104; secretion of active SGPB from these transformants was verified using a skim milk clearing assay (Wells et al., 1983).

2.13.2.2 Construction of pEB-E (SGPE secretion)

EF2 and ER2 were used as PCR primers with pDS-E as template. The amplified product (encoding promature SGPE) was digested with *Xba*I and ligated into pUC18 which had been digested with *Xba*I and *Sma*I and treated with CIP. Vectors with inserts of the correct size were identified by restriction analysis and were partially sequenced to verify accurate amplification and ligation. A plasmid containing a correctly amplified fragment of *sprE* was designated pDS-E2.

pDS-E2 was digested with *Pst*I and *Sma*I; the fragment encoding promature SGPE was gel-purified and ligated into pEB11 (digested with *Pst*I and *Sma*I and treated with CIP). A vector containing the correct insert in the correct orientation was isolated on the basis of restriction enzyme analysis and designated pEB-E. *B. subtilis* DB104 was transformed with pEB-E;

secretion of active SGPE from these transformants was verified using a skim milk clearing assay (Wells et al., 1983).

2.13.2.3 Construction of pEB-C8 and pEB-D8 (SGPC and SGPD secretion)

CF1 and CR1 were used as PCR primers with the plasmid pDS-C as template to amplify a DNA fragment encoding promature SGPC. Similarly, a DNA fragment encoding promature SGPD was amplified from the plasmid pDS-D using DF1 and DR1 as PCR primers. Each amplified product was digested with *Eco*RI and ligated into pUC18 digested with *Eco*RI and *Sma*I and treated with CIP. Vectors with inserts of the correct size were identified by restriction analysis and were partially sequenced to verify accurate amplification and ligation. Plasmids containing correctly amplified fragments of *sprC* or *sprD* were designated pDS-C8 or pDS-D8, respectively.

pDS-C8 or pDS-D8 digested with *Eco*RI and *Sca*I was treated with T4 DNA polymerase to produce blunt-ended fragments. The fragment encoding the promature portion of SGPC or SGPD was gel-purified and ligated into pEB11 that had been digested with *Sma*I and treated with CIP. Vectors containing the correct inserts in the correct orientation were isolated on the basis of restriction analyses. These vectors, designated pEB-C8 (containing *sprC* promature region) and pEB-D8 (containing *sprD* promature region), were transformed into *B. subtilis* DB104. Secretion of active protease was verified using a skim milk clearing assay (Wells et al., 1983).

2.14 Protein purification

2.14.1 Purification of SGPB, SGPC, and SGPE

B. subtilis DB104 harbouring pEB-B8, pEB-C8, or pEB-E were streaked on YT/milk plates containing 50 μg/mL kanamycin (section 2.3) and grown overnight at 30 °C. A single colony with a well-defined zone of clearing was used to inoculate 2 mL of YTC broth (section 2.3). The 2-mL culture was grown approximately 12 h at 30 °C with vigorous shaking and was then used to inoculate 200 mL of YTC broth in a 1-litre Erlenmeyer flask. After a further 12 hour growth period under the same conditions, the culture was used to inoculate 15 L of YTC broth in a Chemap Fermentor equipped with a model FZ3000 control unit and a 20-litre G-type fermentation vessel.

The fermentor culture was maintained at 30 °C with a stir rate of 200 rpm and an aeration rate of 9 L/min, and proteolytic activity was monitored (section 2.15.1). When proteolytic activity had plateaued (approximately 48 h), the culture was cooled to approximately 15 °C.

Bacteria were removed from the culture by ultrafiltration using a Millipore pellicon apparatus equipped with a HVMP membrane cassette (0.45 μm cutoff). The filtrate (containing active protease) was next concentrated to 1.0 L with a PTGC membrane cassette (10,000 nominal molecular weight limit). The retentate was centrifuged for 30 minutes at 10,000 x g to remove any additional precipitate. Sodium acetate (3.0 M, pH 4.8) was added to the concentrated retentate to a final concentration of 100 mM sodium acetate.

Acetone was added to the retentate with stirring to a final concentration of 30% (v/v). After stirring for 10 minutes, the mixture was centrifuged at 4,000 x g for 15 minutes, and the pellet was discarded. A second

volume of acetone was added to the supernatant to a final concentration of 70% (v/v), and the mixture was again stirred and centrifuged as above. The pellet from the second fractionation, which contained active protease, was resuspended in 150 mL of 100 mM sodium acetate (pH 4.8). Proteolytic activity was monitored during all fractionations.

To remove anionic contaminants, the sample was applied to a 60 x 3 cm Q-sepharose anion exchange column equilibrated with 10 mM sodium acetate, pH 4.8; the column was washed with the same buffer, and the flow-through was collected in 25 mL fractions. Fractions exhibiting proteolytic activity were pooled and dialyzed against 5 mM sodium acetate, 2 mM calcium chloride, pH 4.8 (buffer A) overnight at 4 ^{o}C.

The dialyzed sample was applied to a Waters AP-1 cation exchange column using a Pharmacia FPLC system. The column was washed with buffer A until the A_{280} baselined. The enzyme was then eluted in a linear gradient from 0 to 0.25 M NaCl in buffer A over 60 minutes. Fractions with proteolytic activity were analyzed by SDS-PAGE in 12% gels (section 2.17) and those fractions exhibiting a single 26 kDa band (preparation of SGPC) or a single 19 kDa band (preparation of SGPB and SGPE) were pooled.

2.14.2 Purification of SGPD

Cultures of *B. subtilis* DB104 harbouring pEB-D8 were grown and processed as described in section 2.14.1 up to the acetone fractionation stage.

Acetone was added to the retentate with stirring to a final concentration of 60% (v/v). After stirring for 10 minutes, the mixture was centrifuged at 4,000 x g for 15 minutes, and the pellet was discarded. A second volume of acetone was added to the supernatant to a final concentration of

75% (v/v), and the mixture was again stirred and centrifuged as above. The pellet from this second fractionation, which contained active protease, was resuspended in 150 mL of 100 mM sodium phosphate (pH 7.0). Proteolytic activity was monitored during all fractionations.

This sample was applied to a 60 x 3 cm S-sepharose cation exchange column equilibrated with 10 mM sodium phosphate, pH 7.0, in order to remove cationic contaminants. The column was washed with the same buffer and the flow through was collected in 25 mL fractions. Fractions with proteolytic activity were pooled and dialyzed against a 10 mM Tris buffer, pH 8.0 (buffer B) overnight at 4 ^{o}C.

The dialyzed sample was applied to a Mono-Q anion exchange column using a Pharmacia FPLC system. The column was washed with buffer B until A_{280} baselined. The enzyme was then eluted with a salt gradient from 0 to 0.25 M NaCl in buffer B in 60 minutes. Fractions exhibiting proteolytic activity were analyzed by SDS-PAGE in 12% gels (section 2.17) and those fractions exhibiting a single 34 kDa band were pooled.

2.15 Proteolytic activity assays

2.15.1 Monitoring proteolytic activity during enzyme purification

Enzyme activity was determined using the following assay mixture: 1.00 mL of 150 µM Suc-AAPX-pNA (where X = F for SGPB, SGPC, and SGPD; X = E for SGPE) in 50 mM Tris buffer, pH 8.0 + 2-20 µL of enzyme solution. The release of *p*-nitroaniline was detected spectrophotometrically at 412 nm.

2.15.2 Primary specificity comparison

The specific activities of SGPB, SGPC, and SGPD were compared in assays that made use of peptide substrates of the form Suc-AAPX-pNA (where X = A, E, F, I, L, M, or V). Assays were performed as described in section 2.15.1 except that the assay mixture contained 5% methanol and all substrates were at 100 μM. One unit of activity was defined as the amount of enzyme required to produce 1 mmol of *p*-nitroaniline in 1 h at 20 °C.

2.16 Gel filtration chromatography

Size exclusion chromatography was conducted on a Pharmacia FPLC system equipped with a Superose-12 gel filtration column. The column was equilibrated with 50 mM Tris, pH 8.0. Typically 100 μL samples (approximately 10 μg of protein) were loaded, and the column was eluted with 50 mM Tris, pH 8.0 at a flow rate of 0.3 mL/min. Proteins eluting from the column were detected by absorbance at 280 nm.

2.17 SDS-PAGE

SDS-PAGE was performed as described previously (Laemmli, 1970). Protein samples were prepared in one of two ways prior to electrophoresis: 1) "standard denaturing conditions" in which samples were heated for 10 min at 100 °C in 30 mM Tris, 10% glycerol, 4.4% β-mercaptoethanol, and 2% SDS, pH 6.8; or 2) "nondenaturing conditions" in which samples were prepared in 30 mM Tris, 10% glycerol, pH 6.8, without heating. Standard denaturing conditions were used unless otherwise indicated.

2.18 Amino-terminal sequence analysis

Samples were electrophoresed as described in section 2.17; proteins were transferred onto Immobilon-p membranes (composed of polyvinylidene diflouride) using a LKB Multipore transfer apparatus (transfer buffer: 0.2% SDS, 190 mM glycine, 25 mM Tris, pH 8.0). Membrane-bound protein samples were sequenced by the Microsequencing Center at the University of Victoria, Victoria, B.C., using an Applied Biosystems model 473 protein sequencer.

2.19 Homology searches

Homologous DNA and protein sequences were searched for and identified using the BLAST network service at the National Center for Biotechnology Information (NCBI).

2.20 Sequence alignment and phylogeny

Nucleotide sequence data corresponding to the mature regions of the proteases were examined using a DNA parsimony program (DNApars) (Felsenstein, 1989). Amino acid sequences were aligned, taking into consideration the known three-dimensional structures of SGPA (James et al., 1980), SGPB (Read et al., 1983), SGPE (Nienaber et al., 1993), and the α-lytic protease (Fujinaga et al., 1985) . Subsequently, the nucleotide sequences were aligned according to the correspondence of amino acid sequences. Confidence limits were also estimated using the bootstrap program (DNAboot) of Felsenstein (Felsenstein, 1989).

3 RESULTS

3.1 Southern blot analysis

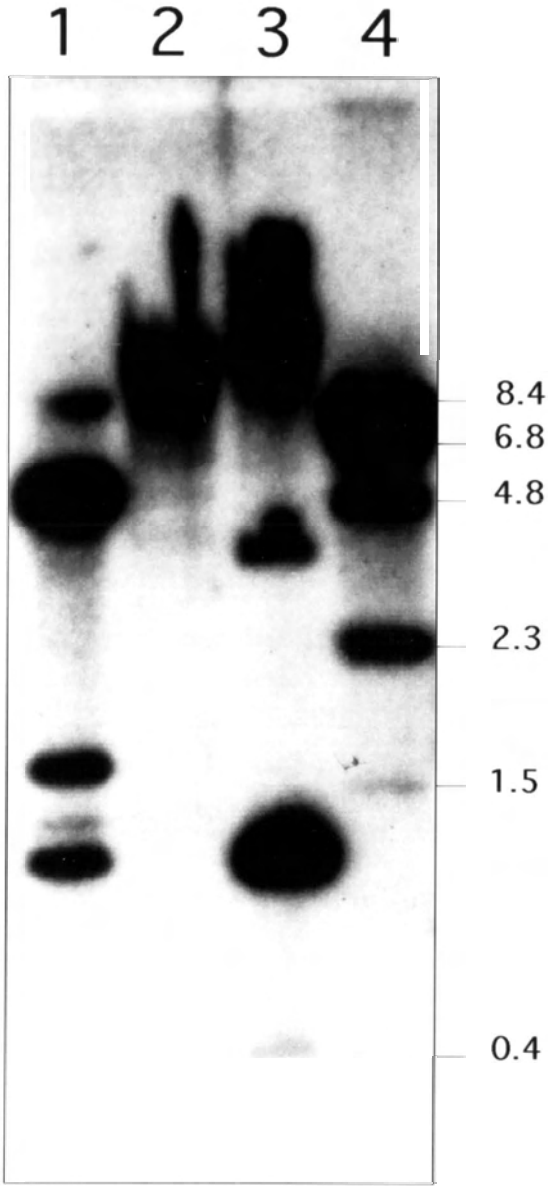
A ^{32}P labeled DNA fragment encoding amino acids 9 to 185 of mature SGPB (B-mat) was hybridized with *Bam*HI-digested, *S. griseus* genomic DNA; six unique genomic DNA fragments were detected (Fig. 2-4). The two largest fragments (8.4 and 6.8 kbp) could be attributed to the genes encoding SGPB and SGPA respectively (Henderson et al., 1987), while the two smallest fragments (1.5 and 0.4 kbp) corresponded in size to fragments containing 5' and 3' portions of *sprE* (Sidhu et al., 1993). However, two strongly hybridizing fragments (4.8 and 2.3 kbp) could not be correlated with the genes of any known *S. griseus* serine proteases.

3.2 Cloning and sequencing of protease genes

3.2.1 *sprE* (the gene encoding SGPE)

Taking advantage of the previously published partial gene sequence of *sprE* (Svendsen et al., 1991), we used the PCR to amplify a fragment of the gene encoding amino acids 44 to 121 of the mature protease (E-mat). The amplified fragment was then used as a probe to isolate plasmids containing *sprE* from a genomic DNA library prepared from *S. griseus* DNA digested to completion with *Bgl*II. From approximately 2×10^4 *E. coli* transformants screened by colony hybridization, 48 potential positive clones were selected for further analysis. Southern blot analysis of plasmid DNA isolated from the 48 clones revealed four that contained inserts which hybridized strongly with

Fig. 2-4. Southern blot analysis of *S. griseus* genomic DNA. DNA was digested with the restriction enzymes *Pvu*II (1), *Pvu*I (2), *Eco*RI (3), or *Bam*HI (4). The blot was hybridized with radiolabeled B-mat DNA, a probe prepared from the mature region of *sprB* as described in sections 2.7 and 2.9. Significant hybridization to six different genomic fragments was observed in *Bam*HI digests. The fragments at 8.4 and 6.8 kbp correspond to *sprB* and *sprA*, respectively (Henderson et al., 1987), and fragments at 1.5 and 0.4 kbp correspond to 5' and 3' portions of *sprE* (Fig. 2-5). The two strongly hybridizing fragments at 4.8 and 2.3 kbp correspond to *sprC* (Fig. 2-6) and *sprD* (Fig. 2-8), respectively.


1
2
3
4
8.4
6.8
4.8
2.3
1.5
0.4

the probe. Restriction analysis showed that each of the four plasmids contained an identical 4.7 kbp insert, and one of these plasmids (designated pDS-E) was chosen for sequencing .

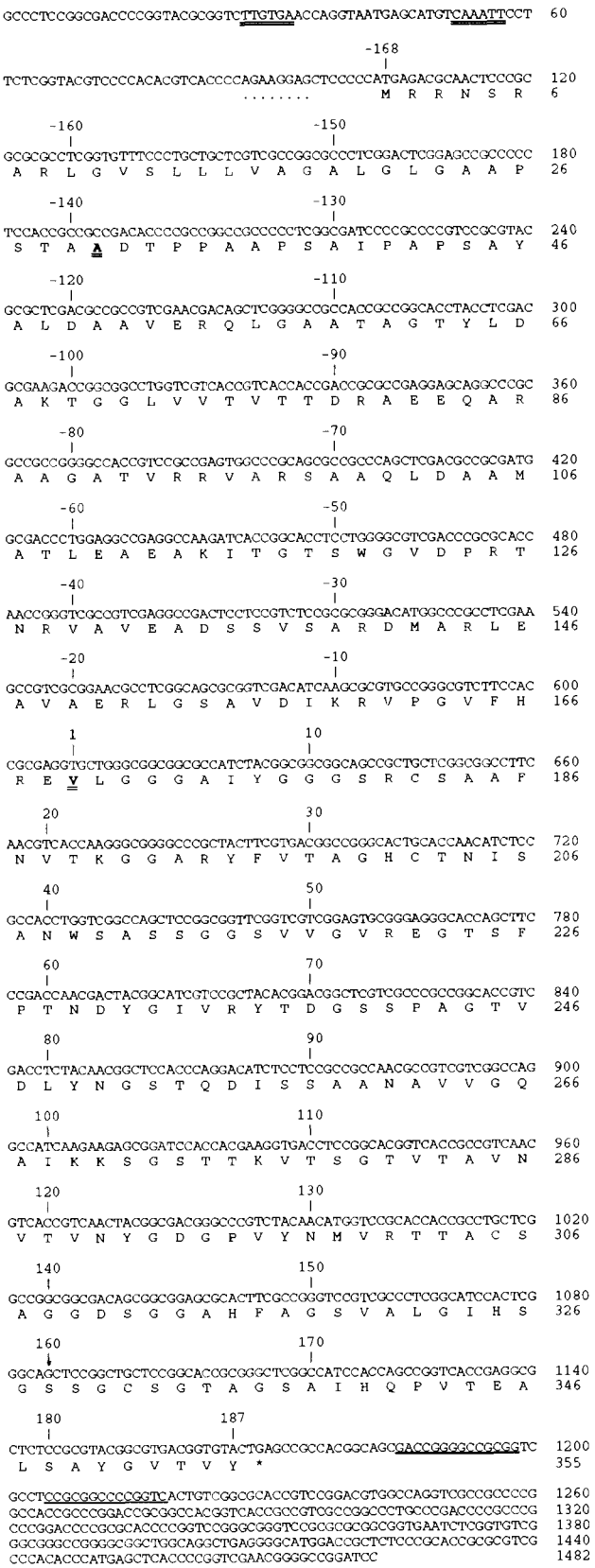
Sequence analysis confirmed that *sprE* was contained within the 4.7 kbp insert of pDS-E. The complete sequence of *sprE* revealed an enzyme organized into a pre-pro-mature precursor (Fig. 2-5) consisting of 187 amino acid mature SGPE preceded by a 168 amino acid extension. The preregion, determined by the method of von Heijne (von Heijne, 1986) using the computer program PC/Gene (IntelliGenetics, Inc., Mountain View, California), encompasses the first 29 amino acids of the extension and is typical of prokaryotic secretion signals. The remaining 139 residues constitute a propeptide.

The 5' untranslated region of *sprE* contains putative promoter and ribosome binding sites which were identified by comparison with other *Streptomyces* gene sequences (Henderson et al., 1987; Hutter & Eckhardt, 1988). The translation stop codon is followed by a pair of inverted repeated sequences capable of forming a stable hairpin loop. Such structures, believed to be involved in transcription termination, have been identified in other *Streptomyces* genes (Henderson et al., 1987; Hutter & Eckhardt, 1988).

3.2.2 *sprC* (the gene encoding SGPC)

The B-mat DNA probe was used to isolate plasmids containing the above mentioned (section 3.1) 4.8 kbp DNA fragment from a DNA library prepared from *S. griseus* genomic DNA digested to completion with *Bam*HI. Approximately 2 x 10^4 *E. coli* transformants were screened by colony hybridization, and 14 strongly hybridizing clones were selected for further

Fig. 2-5. DNA sequence of *sprE*, the gene encoding the protease SGPE. *sprE* encodes a 355-amino acid pre-pro-mature protein; the 187 carboxyl-terminal residues comprise the mature protein. The amino acid sequence deduced from the gene is shown in *one-letter code* beneath the nucleotide sequence. The numbering to the *right* of the sequences is relative to the first nucleotide in the known sequence and to the first amino acid coded by the gene. The numbering *above* the sequence is relative to the first amino acid in the mature protease. The first residues of the propeptide and mature protein are *double underlined in bold.* A termination codon is indicated by *asterisks,*while a putative ribosome binding site is indicated by *dots* below the nucleotide sequence. Putative promoter sites are *double underlined,* and a 3' inverted repeat sequence is *underlined.*


GCCCTCCGGCGACCCCGGTACGCGGTCTTGTGAACCAGGTAATGAGCATGTCAAATTCCT 60
-168
TCTCGGTACGTCCCCACACGTCACCCCAGAAGGAGCTCCCCCATGAGACGCAACTCCCGC 120
M R R N S R 6
-160 -150
GCGCGCCTCGGTGTTTCCCTGCTGCTCGTCGCCGGCGCCCTCGGACTCGGAGCCGCCCCC 180
A R L G V S L L L V A G A L G L G A A P 26
-140 -130
TCCACCGCCGCCGACACCCCGCCGGCCGCCCCCTCGGCGATCCCCGCCCCGTCCGCGTAC 240
S T A A D T P P A A P S A I P A P S A Y 46
-120 -110
GCGCTCGACGCCGCCGTCGAACGACAGCTCGGGGCCGCCACCGCCGGCACCTACCTCGAC 300
A L D A A V E R Q L G A A T A G T Y L D 66
-100 -90
GCGAAGACCGGCGGCCTGGTCGTCACCGTCACCACCGACCGCGCCGAGGAGCAGGCCCGC 360
A K T G G L V V T V T T D R A E E Q A R 86
-80 -70
GCCGCCGGGGCCACCGTCCGCCGAGTGGCCCGCAGCGCCGCCCAGCTCGACGCCGCGATG 420
A A G A T V R R V A R S A A Q L D A A M 106
-60 -50
GCGACCCTGGAGGCCGAGGCCAAGATCACCGGCACCTCCTGGGGCGTCGACCCGCGCACC 480
A T L E A E A K I T G T S W G V D P R T 126
-40 -30
AACCGGGTCGCCGTCGAGGCCGACTCCTCCGTCTCCGCGCGGGACATGGCCCGCCTCGAA 540
N R V A V E A D S S V S A R D M A R L E 146
-20 -10
GCCGTCGCGGAACGCCTCGGCAGCGCGGTCGACATCAAGCGCGTGCCGGGCGTCTTCCAC 600
A V A E R L G S A V D I K R V P G V F H 166
1 10
CGCGAGGTGCTGGGCGGCGGCGCCATCTACGGCGGCGGCAGCCGCTGCTCGGCGGCCTTC 660
R E V L G G G A I Y G G G S R C S A A F 186
20 30
AACGTCACCAAGGGCGGGGCCCGCTACTTCGTGACGGCCGGGCACTGCACCAACATCTCC 720
N V T K G G A R Y F V T A G H C T N I S 206
40 50
GCCACCTGGTCGGCCAGCTCCGGCGGTTCGGTCGTCGGAGTGCGGGAGGGCACCAGCTTC 780
A N W S A S S G G S V V G V R E G T S F 226
60 70
CCGACCAACGACTACGGCATCGTCCGCTACACGGACGGCTCGTCGCCCGCCGGCACCGTC 840
P T N D Y G I V R Y T D G S S P A G T V 246
80 90
GACCTCTACAACGGCTCCACCCAGGACATCTCCTCCGCCGCCAACGCCGTCGTCGGCCAG 900
D L Y N G S T Q D I S S A A N A V V G Q 266
100 110
GCCATCAAGAAGAGCGGATCCACCACGAAGGTGACCTCCGGCACGGTCACCGCCGTCAAC 960
A I K K S G S T T K V T S G T V T A V N 286
120 130
GTCACCGTCAACTACGGCGACGGGCCCGTCTACAACATGGTCCGCACCACCGCCTGCTCG 1020
V T V N Y G D G P V Y N M V R T T A C S 306
140 150
GCCGGCGGCGACAGCGGCGGAGCGCACTTCGCCGGGTCCGTCGCCCTCGGCATCCACTCG 1080
A G G D S G G A H F A G S V A L G I H S 326
160 170
GGCAGCTCCGGCTGCTCCGGCACCGCGGGCTCGGCCATCCACCAGCCGGTCACCGAGGCG 1140
G S S G C S G T A G S A I H Q P V T E A 346
180 187
CTCTCCGCGTACGGCGTGACGGTGTACTGAGCCGCCACGGCAGCGACCGGGGCCGCGGTC 1200
L S A Y G V T V Y *
GCCTCCGCGGCCCCGGTCACTGTCGGCGCACCGTCCGGACGTGGCCAGGTCGCCGCCCCG 1260
GCCACCGCCCGGACCGCGGCCACGGTCACCGCCGTCGCCGGCCCTGCCCGACCCCGCCCG 1320
CCCGGACCCCGCGCACCCCGGTCCGGGCGGGTCCGCGCGCGGCGGTGAATCTCGGTGTCG 1380
GGCGGGCCCGGGGCGGCTGGCAGGCTGAGGGGCATGGACCGCTCTCCCGCACCGCGCGTCG 1440
CCCACACCCATGAGCTCACCCCGGTCGAACGGGGCCGGATCC 1482

analysis. Southern blot analysis of plasmid DNA isolated from these clones revealed three that contained a 4.8 kbp insert which hybridized strongly with the B-mat probe. One of these plasmids (designated pDS-C) was chosen for sequencing.

Sequence analysis of pDS-C revealed a gene which we designated *sprC* (Fig. 2-6). This gene contains an open reading frame encoding a polypeptide of 457 amino acids. The putative GTG initiation codon is preceded by a potential ribosome binding site but, since the coding region begins just 34 nucleotides from one end of the insert, the promoter is not contained within this DNA fragment. Comparison of the predicted polypeptide with the gene products of *sprA*, *sprB* (Henderson et al., 1987), *sprE* (section 3.2.1) and the α-lytic protease (α-LP) gene of *Lysobacter enzymogenes* (Silen et al., 1988) suggested a similar pre-pro-mature organization. The preregion, determined by the method of von Heijne (von Heijne, 1986) using the computer program PC/Gene (Intelligenetics, Inc., Mountain View, California), encompasses the first 40 amino acids of the polypeptide and is typical of prokaryotic secretion signals. The preregion is followed by a propeptide of 162 amino acids which in turn is followed by a 189 amino acid segment homologous (45% identity) to mature SGPB. The promature junction shown in Fig. 2-6 was initially assigned on the basis of sequence alignments with SGPB and α-LP and later confirmed by amino-terminal analysis of the purified, recombinant protease. The open reading frame terminates in a 66 amino acid extension which is not present in SGPB nor in any other known protease.

A computer search of the complete non-redundant DNA/protein database revealed that the sequence of the final 47 amino acids of the carboxyl-terminal extension is highly homologous to the carboxyl-terminal domain of chitinase A1 (57% identity) and the amino-terminal domain of chitinase D

Fig. 2-6. DNA sequence of *sprC*, the gene encoding the protease SGPC. *sprC* encodes a 457-amino acid pre-pro-mature protein; the 255 carboxyl-terminal residues comprise the mature protein. The amino acid sequence deduced from the gene is shown in *one-letter code* beneath the nucleotide sequence. The numbering to the *right* of the sequences is relative to the first nucleotide in the known sequence and to the first amino acid coded by the gene. The numbering *above* the sequence is relative to the first amino acid in the mature protease. The first residues of the propeptide and mature protein are *double underlined in bold.* A termination codon is indicated by *asterisks*. A putative ribosome binding site is indicated by *dots* below the nucleotide sequence. Mature SGPC consists of an amino-terminal protease domain (amino acids 1-189) connected to a smaller carboxyl-terminal domain (amino acids 209-255) by a proline/threonine-rich linker (amino acids 190-208).

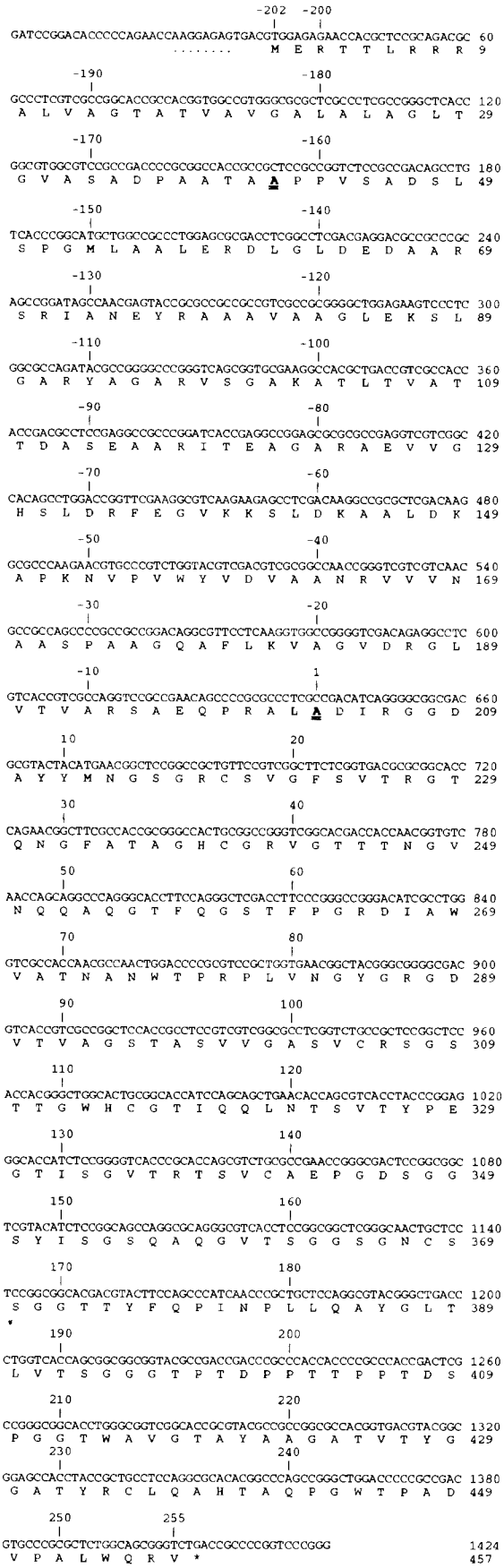
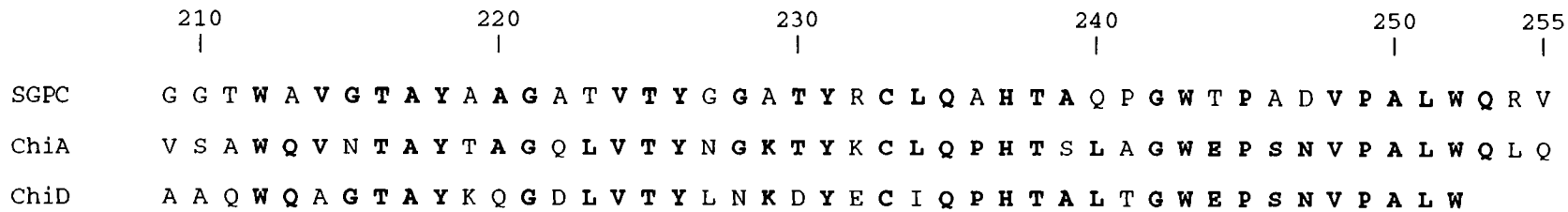

-202 -200
GATCCGGACACCCCCAGAACCAAGGAGAGTGACGTGGAGAGAACCACGCTCCGCAGACGC 60
M E R T T L R R R 9
-190 -180
GCCCTCGTCGCCGGCACCGCCACGGTGGCCGTGGGCGCGCTCGCCCTCGCCGGGCTCACC 120
A L V A G T A T V A V G A L A L A G L T 29
-170 -160
GGCGTGGCGTCCGCCGACCCCGCGGCCACCGCCGCTCCGCCGGTCTCCGCCGACAGCCTG 180
G V A S A D P A A T A A P P V S A D S L 49
-150 -140
TCACCCGGCATGCTGGCCGCCCTGGAGCGCGACCTCGGCCTCGACGAGGACGCCGCCCGC 240
S P G M L A A L E R D L G L D E D A A R 69
-130 -120
AGCCGGATAGCCAACGAGTACCGCGCCGCCGCCGTCGCCGCGGGGCTGGAGAAGTCCCTC 300
S R I A N E Y R A A A V A A G L E K S L 89
-110 -100
GGCGCCAGATACGCCGGGGCCCGGGTCAGCGGTGCGAAGGCCACGCTGACCGTCGCCACC 360
G A R Y A G A R V S G A K A T L T V A T 109
-90 -80
ACCGACGCCTCCGAGGCCGCCCGGATCACCGAGGCCGGAGCGCGCGCCGAGGTCGTCGGC 420
T D A S E A A R I T E A G A R A E V V G 129
-70 -60
CACAGCCTGGACCGGTTCGAAGGCGTCAAGAAGAGCCTCGACAAGGCCGCGCTCGACAAG 480
H S L D R F E G V K K S L D K A A L D K 149
-50 -40
GCGCCCAAGAACGTGCCCGTCTGGTACGTCGACGTCGCGGCCAACCGGGTCGTCGTCAAC 540
A P K N V P V W Y V D V A A N R V V V N 169
-30 -20
GCCGCCAGCCCCGCCGCCGGACAGGCGTTCCTCAAGGTGGCCGGGGTCGACAGAGGCCTC 600
A A S P A A G Q A F L K V A G V D R G L 189
-10 1
GTCACCGTCGCCAGGTCCGCCGAACAGCCCCGCGCCCTCGCCGACATCAGGGGCGGCGAC 660
V T V A R S A E Q P R A L A D I R G G D 209
10 20
GCGTACTACATGAACGGCTCCGGCCGCTGTTCCGTCGGCTTCTCGGTGACGCGCGGCACC 720
A Y Y M N G S G R C S V G F S V T R G T 229
30 40
CAGAACGGCTTCGCCACCGCGGGCCACTGCGGCCGGGTCGGCACGACCACCAACGGTGTC 780
Q N G F A T A G H C G R V G T T T N G V 249
50 60
AACCAGCAGGCCCAGGGCACCTTCCAGGGCTCGACCTTCCCGGGCCGGGACATCGCCTGG 840
N Q Q A Q G T F Q G S T F P G R D I A W 269
70 80
GTCGCCACCAACGCCAACTGGACCCCGCGTCCGCTGGTGAACGGCTACGGGCGGGGCGAC 900
V A T N A N W T P R P L V N G Y G R G D 289
90 100
GTCACCGTCGCCGGCTCCACCGCCTCCGTCGTCGGCGCCTCGGTCTGCCGCTCCGGCTCC 960
V T V A G S T A S V V G A S V C R S G S 309
110 120
ACCACGGGCTGGCACTGCGGCACCATCCAGCAGCTGAACACCAGCGTCACCTACCCGGAG 1020
T T G W H C G T I Q Q L N T S V T Y P E 329
130 140
GGCACCATCTCCGGGGTCACCCGCACCAGCGTCTGCGCCGAACCGGGCGACTCCGGCGGC 1080
G T I S G V T R T S V C A E P G D S G G 349
150 160
TCGTACATCTCCGGCAGCCAGGCGCAGGGCGTCACCTCCGGCGGCTCGGGCAACTGCTCC 1140
S Y I S G S Q A Q G V T S G G S G N C S 369
170 180
TCCGGCGGCACGACGTACTTCCAGCCCATCAACCCGCTGCTCCAGGCGTACGGGCTGACC 1200
S G G T T Y F Q P I N P L L Q A Y G L T 389
*
190 200
CTGGTCACCAGCGGCGGCGGTACGCCGACCGACCCGCCCACCACCCCGCCCACCGACTCG 1260
L V T S G G G T P T D P P T T P P T D S 409
210 220
CCGGGCGGCACCTGGGCGGTCGGCACCGCGTACGCCGCCGGCGCCACGGTGACGTACGGC 1320
P G G T W A V G T A Y A A G A T V T Y G 429
230 240
GGAGCCACCTACCGCTGCCTCCAGGCGCACACGGCCCAGCCGGGCTGGACCCCCGCCGAC 1380
G A T Y R C L Q A H T A Q P G W T P A D 449
250 255
GTGCCCGCGCTCTGGCAGCGGGTCTGACCGCCCCGGTCCCGGG 1424
V P A L W Q R V * 457

Fig. 2-7. Alignment of SGPC's carboxyl-terminal domain and the homologous domains of *Bacillus circulans* chitinases. Numbering *above* the sequences is relative to the amino-terminal end of mature SGPC. The sequence of the carboxyl-terminal SGPC domain appears at the *top* of the figure with sequences of the chitinase A1 (ChiA) and chitinase D (ChiD) domains (Watanabe et al., 1992) aligned *below*. The ChiA sequence is the putative chiting-binding domain at the carboxyl-terminal end of chitinase A1 (amino acids 653-699), while the ChiD sequence is the putative chitin-binding domain at the amino-terminal end of chitinase D (amino acids 31-77). Sequence identities are in *bold* text. Notably, the three sequences are colinear; gaps were not introduced in the process of alignment.


210
220
230
240
250
255
SGPC G G T W A V G T A Y A A G A T V T Y G G A T Y R C L Q A H T A Q P G W T P A D V P A L W Q R V
ChiA V S A W Q V N T A Y T A G Q L V T Y N G K T Y K C L Q P H T S L A G W E P S N V P A L W Q L Q
ChiD A A Q W Q A G T A Y K Q G D L V T Y L N K D Y E C I Q P H T A L T G W E P S N V P A L W

(49% identity) of *Bacillus circulans* (Fig. 2-7). The intervening sequence which connects the small domain to the protease domain is rich in threonines and prolines; the homologous domains of chitinases A1 and D are connected to larger chitinolytic domains by similar linker regions (Watanabe et al., 1992).

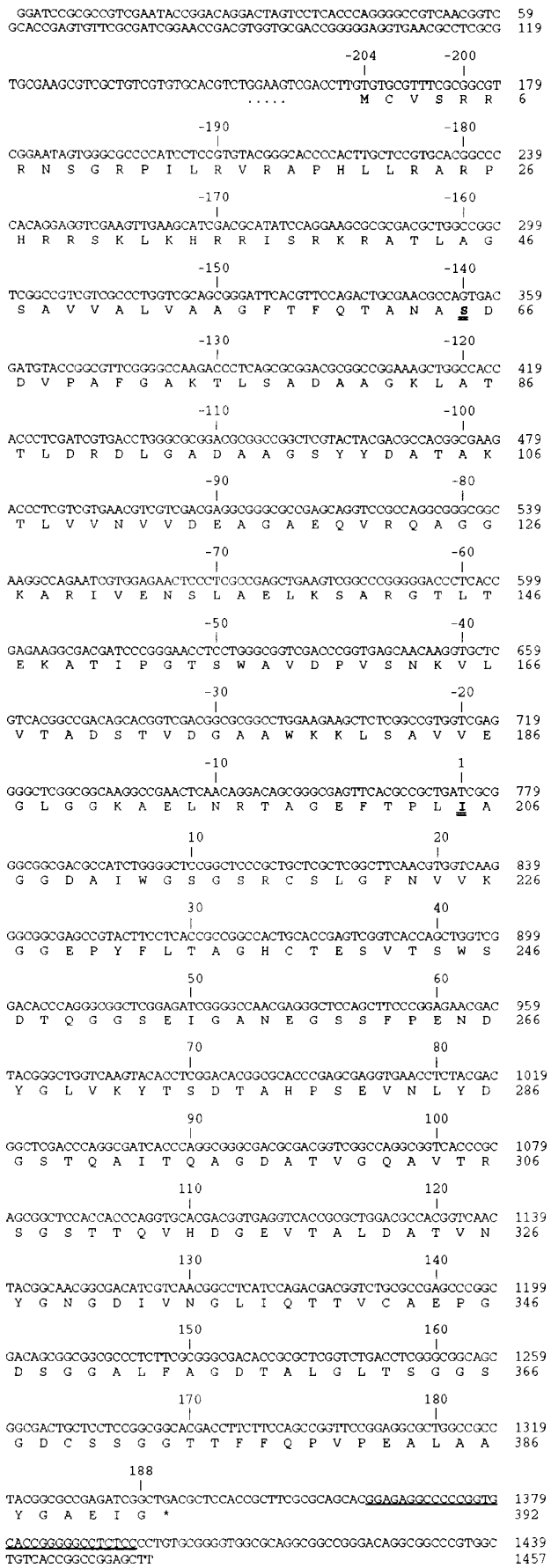
3.2.3 *sprD* (the gene encoding SGPD)

The B-mat DNA probe was used to isolate plasmids containing the above mentioned (section 3.1) 2.3 kbp DNA fragment from a DNA library prepared from *S. griseus* genomic DNA digested to completion with *Bam*HI. Approximately 1 x 10^4 *E. coli* transformants were screened by colony hybridization, and strongly hybridizing clones were selected for further analysis. Southern blot analysis of plasmid DNA isolated from these clones revealed nine that contained a 2.3 kbp insert which hybridized strongly with the B-mat probe. One of these plasmids was designated pDS-D and was chosen for sequence analysis.

The 2.3 kbp insert of pDS-D contains a gene, designated *sprD*, which encodes a polpypeptide of 392 amino acids (Fig. 2-8). The organization of the polypeptide is analogous to that of SGPA, SGPB (Henderson et al., 1987), SGPE (section 3.2.1) and α-LP (Silen et al., 1988). On the basis of sequence alignments, we concluded that *sprD* encodes a pre-pro-mature form of a previously uncharacterized serine protease which we designated *S. griseus* protease D (SGPD).

While the prepro and promature junctions shown in Fig. 2-8 were assigned on the basis of sequence alignments, subsequent amino-terminal analysis of the purified, recombinant protease confirmed the promature

Fig. 2-8. DNA sequence of *sprD*, the gene encoding the protease SGPD. *sprD* encodes a 392-amino acid pre-pro-mature protein; the 188 carboxyl-terminal residues comprise the mature protein. The amino acid sequence deduced from the gene is shown in *one-letter code* beneath the nucleotide sequence. The numbering to the *right* of the sequences is relative to the first nucleotide in the known sequence and to the first amino acid coded by the gene. The numbering *above* the sequence is relative to the first amino acid in the mature protease. The first residues of the propeptide and mature protein are *double underlined in bold.* A putative ribosome binding site is indicated by *dots* below the nucleotide sequence. A termination codon is indicated by *asterisks,*while inverted repeated sequences following the termination codon are *underlined.*

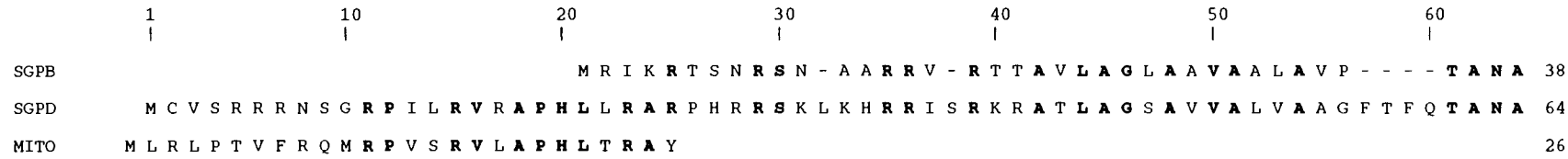

GGATCCGCGCCGTCGAATACCGGACAGGACTAGTCCTCACCCAGGGGCCGTCAACGGTC 59
GCACCGAGTGTTCGCGATCGGAACCGACGTGGTGCGACCGGGGGAGGTGAACGCCTCGCG 119
-204 -200
TGCGAAGCGTCGCTGTCGTGTGCACGTCTGGAAGTCGACCTTGTGTGCGTTTCGCGGCGT 179
..... M C V S R R 6
-190 -180
CGGAATAGTGGGCGCCCCATCCTCCGTGTACGGGCACCCCACTTGCTCCGTGCACGGCCC 239
R N S G R P I L R V R A P H L L R A R P 26
-170 -160
CACAGGAGGTCGAAGTTGAAGCATCGACGCATATCCAGGAAGCGCGCGACGCTGGCCGGC 299
H R R S K L K H R R I S R K R A T L A G 46
-150 -140
TCGGCCGTCGTCGCCCTGGTCGCAGCGGGATTCACGTTCCAGACTGCGAACGCCAGTGAC 359
S A V V A L V A A G F T F Q T A N A S D 66
-130 -120
GATGTACCGGCGTTCGGGGCCAAGACCCTCAGCGCGGACGCGGCCGGAAAGCTGGCCACC 419
D V P A F G A K T L S A D A A G K L A T 86
-110 -100
ACCCTCGATCGTGACCTGGGCGCGGACGCGGCCGGCTCGTACTACGACGCCACGGCGAAG 479
T L D R D L G A D A A G S Y Y D A T A K 106
-90 -80
ACCCTCGTCGTGAACGTCGTCGACGAGGCGGGCGCCGAGCAGGTCCGCCAGGCGGGCGGC 539
T L V V N V V D E A G A E Q V R Q A G G 126
-70 -60
AAGGCCAGAATCGTGGAGAACTCCCTCGCCGAGCTGAAGTCGGCCCGGGGGACCCTCACC 599
K A R I V E N S L A E L K S A R G T L T 146
-50 -40
GAGAAGGCGACGATCCCGGGAACCTCCTGGGCGGTCGACCCGGTGAGCAACAAGGTGCTC 659
E K A T I P G T S W A V D P V S N K V L 166
-30 -20
GTCACGGCCGACAGCACGGTCGACGGCGCGGCCTGGAAGAAGCTCTCGGCCGTGGTCGAG 719
V T A D S T V D G A A W K K L S A V V E 186
-10 1
GGGCTCGGCGGCAAGGCCGAACTCAACAGGACAGCGGGCGAGTTCACGCCGCTGATCGCG 779
G L G G K A E L N R T A G E F T P L I A 206
10 20
GGCGGCGACGCCATCTGGGGCTCCGGCTCCCGCTGCTCGCTCGGCTTCAACGTGGTCAAG 839
G G D A I W G S G S R C S L G F N V V K 226
30 40
GGCGGCGAGCCGTACTTCCTCACCGCCGGCCACTGCACCGAGTCGGTCACCAGCTGGTCG 899
G G E P Y F L T A G H C T E S V T S W S 246
50 60
GACACCCAGGGCGGCTCGGAGATCGGGGCCAACGAGGGCTCCAGCTTCCCGGAGAACGAC 959
D T Q G G S E I G A N E G S S F P E N D 266
70 80
TACGGGCTGGTCAAGTACACCTCGGACACGGCGCACCCGAGCGAGGTGAACCTCTACGAC 1019
Y G L V K Y T S D T A H P S E V N L Y D 286
90 100
GGCTCGACCCAGGCGATCACCCAGGCGGGCGACGCGACGGTCGGCCAGGCGGTCACCCGC 1079
G S T Q A I T Q A G D A T V G Q A V T R 306
110 120
AGCGGCTCCACCACCCAGGTGCACGACGGTGAGGTCACCGCGCTGGACGCCACGGTCAAC 1139
S G S T T Q V H D G E V T A L D A T V N 326
130 140
TACGGCAACGGCGACATCGTCAACGGCCTCATCCAGACGACGGTCTGCGCCGAGCCCGGC 1199
Y G N G D I V N G L I Q T T V C A E P G 346
150 160
GACAGCGGCGGCGCCCTCTTCGCGGGCGACACCGCGCTCGGTCTGACCTCGGGCGGCAGC 1259
D S G G A L F A G D T A L G L T S G G S 366
170 180
GGCGACTGCTCCTCCGGCGGCACGACCTTCTTCCAGCCGGTTCCGGAGGCGCTGGCCGCC 1319
G D C S S G G T T F F Q P V P E A L A A 386
188
TACGGCGCCGAGATCGGCTGACGCTCCACCGCTTCGCGCAGCACGGAGAGGCCCCCGGTG 1379
Y G A E I G * 392
CACCGGGGGCCTCTCCCCTGTGCGGGGTGGCGCAGGCGGCCGGGACAGGCGGCCCGTGGC 1439
TGTCACCGGCCGGAGCTT 1457

junction position. Mature SGPD, encompassing the final 188 amino acids of the open reading frame, is preceded by a leader peptide of 214 amino acids. The amino terminal 64 residues of this leader peptide constitute a prepeptide while the remaining 140 residues form the propeptide.

The preregion of SGPD is significantly longer than the preregions of the other proteases which range in length from 29 to 40 amino acids. The 40-residue carboxyl-terminal segment of the SGPD preregion is characteristic of bacterial secretion signals (Randall & Hardy, 1989) and shares significant homology with the preregion of SGPB (Fig. 2-9). The 24 amino-terminal residues form an amino-terminal extension not present in the other proteases. Interestingly, a computer search of the complete non-redundant DNA/protein database revealed that this region shares significant homology with the mitochondrial signal sequence of hsp60 (Jindal et al., 1989; Venner & Gupta, 1990) (Fig. 2-9). Moreover, a predictive method (Gavel & von Heijne, 1990) revealed that residues 1 to 44 comprise a sequence which is consistent with mitochondrial import signals, and residues 24 to 40 have the potential of forming an amphipathic α-helix with one face highly positively charged, a motif considered essential for translocation of proteins into mitochondria(Lemire et al., 1989; Hartl & Neupert, 1990).

The 5' untranslated region of *sprD* contains a putative ribosome binding site which was identified by comparison with other *Streptomyces* gene sequences (Henderson et al., 1987; Hutter & Eckhardt, 1988). The translation stop codon is followed by a pair of inverted repeated sequences capable of forming a stable hairpin loop. Such structures, believed to be involved in transcription termination, have been identified in other *Streptomyces* genes (Henderson et al., 1987; Hutter & Eckhardt, 1988).

Fig. 2-9. Homology of the SGPD prepeptide with bacterial and mitochondrial signal sequences. The prepeptide of SGPD is shown in alignment with the presequence of SGPB and the mitochondrial import sequence of human hsp60 (Jindal et al., 1989; Venner & Gupta, 1990). Sequence identities are in *bold* text; each sequences begins at the amino terminus. Numbering to the *right* is relative to the amino terminus of each sequence; numbering *above* the sequences refers to SGPD. The unusually long prepeptide of SGPD can be divided into two domains on the basis of homologies with the prokaryotic secretion and mitochondrial import signals. A predictive method (Gavel & von Heijne, 1990) revealed that residues 1-44 comprise a sequence that is consistent with mitochondrial import signals. Residues 24-40 have the potential of forming an amphipathic α-helix, a motif considered essential for translocation of proteins into mitochondria (Lemire et al., 1989; Hartl & Neupert, 1990). The carboxyl-terminal portion of the prepeptide (amino acids 45-64) is homologous with the prepeptide of SGPB.


1 10 20 30 40 50 60
SGPB MRIKRTSNRSN-AARRV-RTTAVLAGLAAVAALAVP----TANA 38
SGPD MCVSRRRNSGRPILRVRAPHLLRARPHRRSKLKHRRISRKRATLAGSAVVALVAAGFTFQTANA 64
MITO MLRLPTVFRQMRPVSRVLAPHLTRAY 26

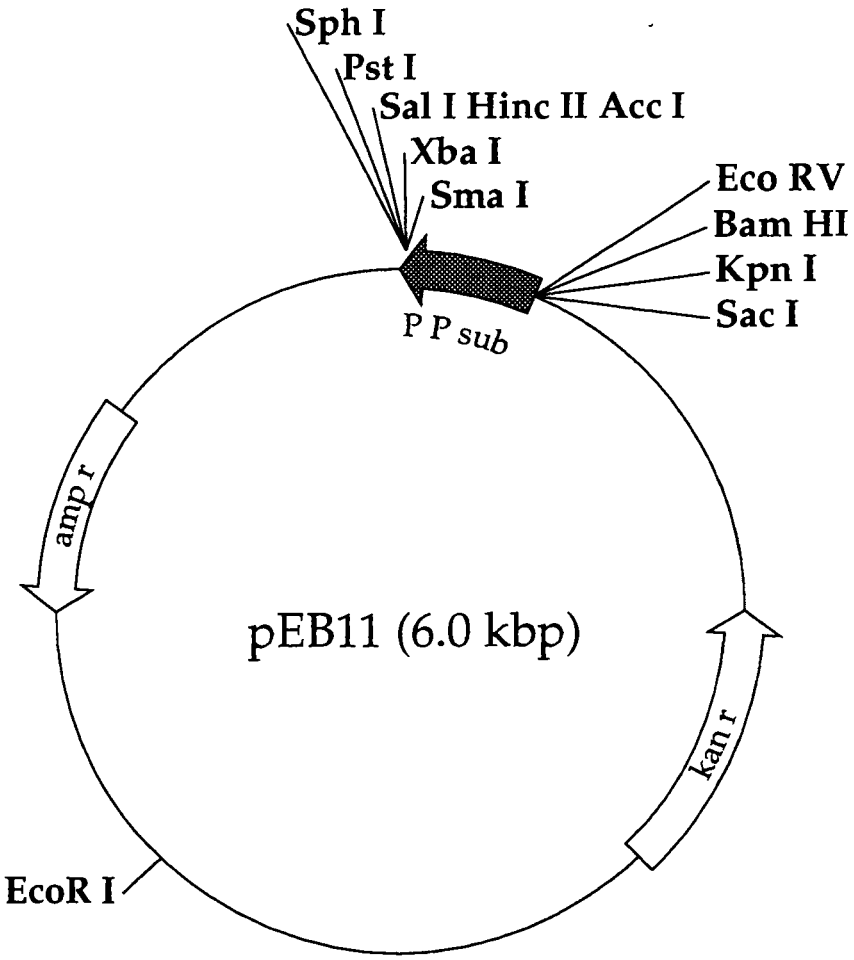
51b

3.3 Recombinant protease secretion in *B. subtilis*

3.3.1 Construction of pEB11 (an *E. coli/B. subtilis* shuttle vector)

B. subtilis is a gram positive organism adapted for the secretion of hydrolytic enzymes (Harwood, 1989); the protease-deficient DB104 strain of *B. subtilis (Kawamura & Doi, 1984)* is an ideal host for recombinant protease secretion. However, low-efficiency labour-intensive transformations make the genetic manipulation of DNA in *B. subtilis* generally more difficult than in *E. coli*. Therefore, we assembled the *E. coli/B. subtilis* shuttle vector pEB11 (Fig. 2-10) from the *E. coli* plasmid pUC18 and the *B. subtilis* plasmid pUB110. pEB11 contains origins of replication recognized by both organisms, an ampicillin resistance gene for selection in *E. coli*, and a kanamycin resistance gene for selection in *B. subtilis*. The transformation efficiencies of pEB11 in *E. coli* and *B. subtilis* were found to be comparable to those of pUC18 and pUB110 in the respective hosts. pEB11 also contains a 330 bp PCR-amplified fragment derived from the *B.amyloliquefaciens* subtilisin gene. This fragment contains the region encoding the subtilisin secretion signal (prepeptide) preceded by 210 bp of 5' untranslated region containing promoter and ribosome-binding sites recognized by *B. subtilis* (Wells et al., 1983). pEB11 was designed in such a way that insertion into the unique *Sma*I site of the vector produces an open reading frame consisting of the subtilisin preregion joined in-frame to the inserted DNA by a proline codon.

Fig. 2-10. Plasmid map of pEB11. Genes are represented by either open or shaded boxes with arrowheads indicating the direction of transcription. The ampicillin resistance gene of pUC18 (*ampr*) or the kanamycin resistance gene of pUB110 (*kanr*) allow for selection in *E. coli* or *B. subtilis*, respectively. A 330-bp fragment from the subtilisin BPN' gene (PPsub) contains the region encoding the subtilisin secretion signal preceded by promoter and ribosome-binding sites recognized by *B. subtilis* (Wells et al., 1983). Plasmid construction is described in section 2.13.1.


Sph I
Pst I
Sal I Hinc II Acc I
Xba I
Sma I
Eco RV
Bam HI
Kpn I
Sac I
P P sub
amp r
pEB11 (6.0 kbp)
kan r
EcoR I

3.3.2 Construction of vectors for protease secretion

Gene fragments encoding promature proteases were cloned into the pEB11 *Sma*I site; the resulting vectors contain open reading frames encoding fusion proteins composed of the prepeptide of subtilisin followed in-frame by promature recombinant protease segments.

The expression vector pEB-E contains an open reading frame encoding the subtilisin preregion connected to promature SGPE (residues 30-355) by a dipeptide sequence (Pro-Gly). Similarly, in pEB-C8 or pEB-D8 the subtilisin prepeptide is connected to promature SGPC (residues 35-457) or SGPD (residues 65-392), respectively, by a dipeptide (Pro-Thr). SGPC is closely related to α-LP (sections 4.4.1 and 4.6); the deletion of as few as six amino acids from the amino terminus of the propeptide of α-LP completely abolishes the secretion of active protease in *E. coli* (Fujishige et al., 1992). Thus, pEB-C8 was designed to secrete a fragment of SGPC initiating from within the putative prepeptide to ensure inclusion of the complete promature enzyme. Finally, in pEB-B8 the subtilisin preregion is connected to promature SGPB (residues 40-299) (Henderson et al., 1987) by a dipeptide (Pro-Ala).

Secretion of active protease was evidenced on YT/milk plates by the appearance of zones of clearing (caused by the degradation of milk proteins) around *B. subtilis* transformants harbouring pEB-E, pEB-B8, pEB-C8, or pEB-D8. Zones of clearing were not observed around pEB11 transformants.

3.4 Purification and characterization of recombinant proteases

3.4.1 Yields of purified recombinant proteases

Optimal protease expression in *B. subtilis* required media containing tryptone, yeast extract, and $CaCl_2$. Because yeast extract contains a pigment which interferes with subsequent purification steps, the media was treated with activated charcoal prior to autoclaving. Pretreatment resulted in the removal of the majority of the pigment without diminishing protease expression. Any remaining traces of the pigment were effectively removed during ion exchange chromatography on Q-sepharose (purification of SGPB, SGPC, or SGPE) or S-sepharose (purification of SGPD), since the pigment bound to the matrix while the proteases were not adsorbed.

The final purified yields of all enzymes were approximately 5 mg/L. Each purified enzyme showed a single band by SDS-PAGE with a mobility corresponding to a molecular mass of 19 kDa for SGPB or SGPE, 26 kDa for SGPC, or 34 kDa for SGPD. Amino-terminal sequence analysis of purified SGPB verified correct processing, since the sequence of recombinant enzyme was identical to the amino-terminal sequence of SGPB purified from *S. griseus* (Henderson et al., 1987). Amino-terminal sequence analysis of SGPC and SGPD confirmed the predicted promature junction locations (Figs. 2-6 and 2-8).

Thus, the development of a *B. subtilis* expression system and simple purification protocols provided substantial yields of highly pure, correctly processed proteases.

3.4.2 Primary specificity of SGPC and SGPD

The primary specificity of SGPB has been well characterized (Bauer, 1978). We compared the primary specificities of SGPC and SGPD to that of SGPB by measuring relative specific activities towards a series of substrates varying solely in the residue at the P1 position (substrates of the form Suc-AAPX-pNA, where X is a variable amino acid). SGPC and SGPD were found to have primary specificities very similar to that of SGPB (Fig. 2-11) and, notably, none of the enzymes were able to hydrolyze a substrate containing glutamic acid in the P1 position.

3.4.3 Quaternary structure of SGPD

Purified SGPE, SGPC, and SGPD were chromatographed separately on a size exclusion column; each sample produced a distinct single peak, and the following retention times were determined: SGPE (54.5 min), SGPC (51 min), SGPD (48 min). A chromatograph of a sample containing approximately equal quantities of SGPE, SGPC, and SGPD is shown in Fig. 2-12. Since retention time is inversely proportional to molecular mass, these results suggest that native SGPE and SGPC exist as monomers with molecular masses of 18 kDa and 26 kDa, respectively, while native SGPD exists as a homodimer with a molecular mass of 36 kDa.

SDS-PAGE of SGPC and SGPD provided further evidence that SGPD exists as a very stable homodimer (Fig. 2-13). Under denaturing conditions, SGPC showed a single band corresponding to a molecular mass of 26 kDa, in good agreement with the monomeric molecular mass predicted from the DNA sequence of *sprC*. Under the same conditions, SGPD (which exhibited a

Fig. 2-11. Substrate specificities of SGPC and SGPD. The specific activity of SGPC or SGPD is plotted relative to that of SGPB which is known to be chymotrypsin-like in activity (Bauer, 1978). Specific activities were measured against a series of substrates having the form *N*-succinyl-Ala-Ala-Pro-X-*p*-nitroanilide; X, the amino acid at the P1 site of the substrate, was varied. The P1 amino acid is indicated to the *right* of each data point. In each case, the data points fall on a line with a slope close to unity, indicating that the substrate specificities of SGPC and SGPD are approximately the same as that of SGPB. Methods are described in section 2.15.2.

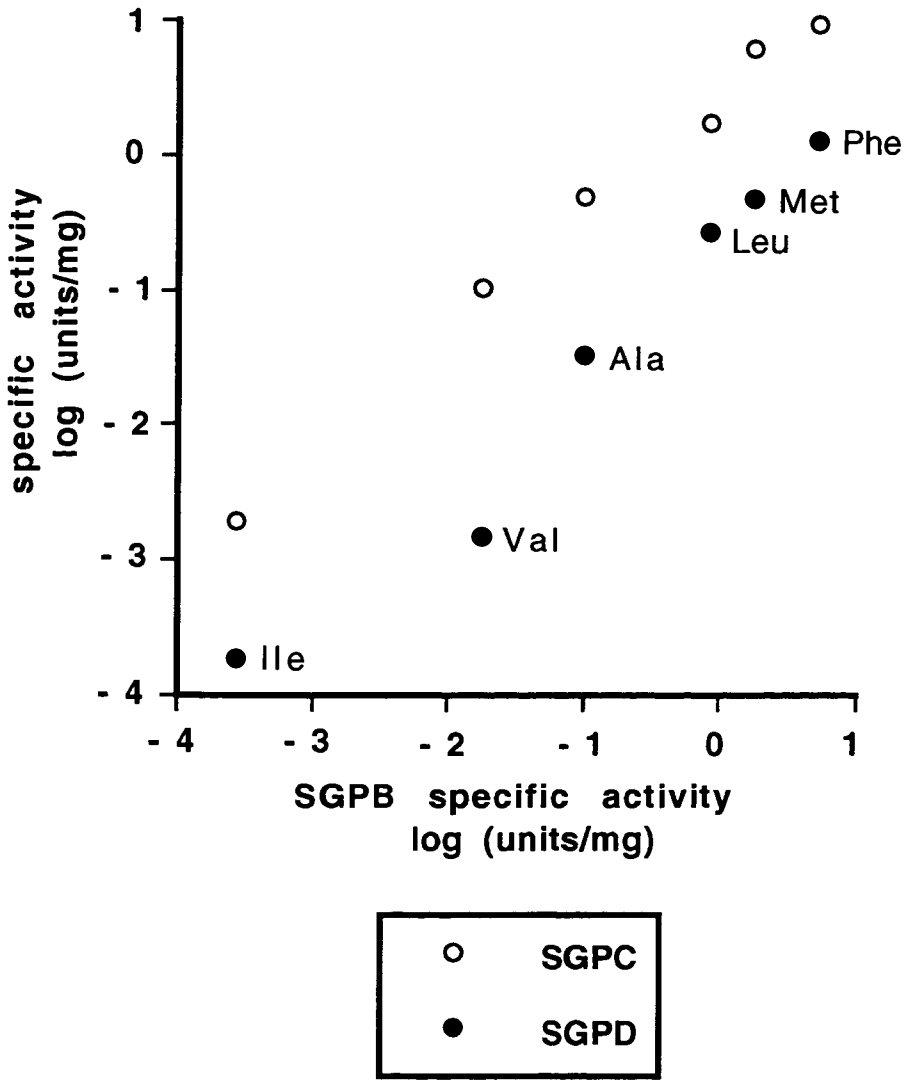

1
0
- 1
- 2
- 3
- 4
specific activity
log (units/mg)
Phe
Met
Leu
Ala
Val
Ile
- 4
- 3
- 2
- 1
0
1
SGPB specific activity
log (units/mg)
SGPC
SGPD

Fig. 2-12. Elution profile from the gel filtration chromatography of a combination of *S. griseus* proteases. SGPD (D) eluted from the column at 48 minutes, followed by SGPC (C) at 51 minutes and SGPE (E) at 54.5 minutes. Retention time is inversely proportional to molecular weight; consistent with SGPE and SGPC existing as monomers with the anticipated molecular weights of 18 kDa and 26 kDa, respectively, and SGPD existing as an α_2 dimer with a molecular weight of 36 kDa. Methods are described in section 2.16.

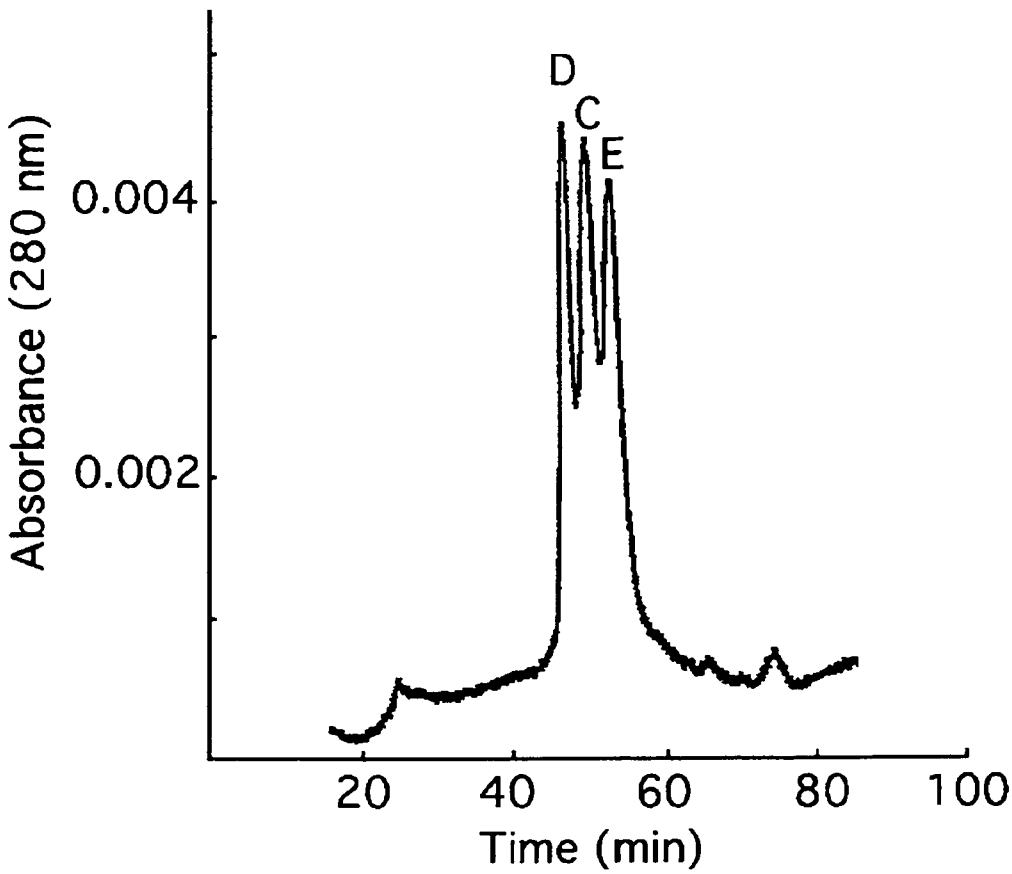
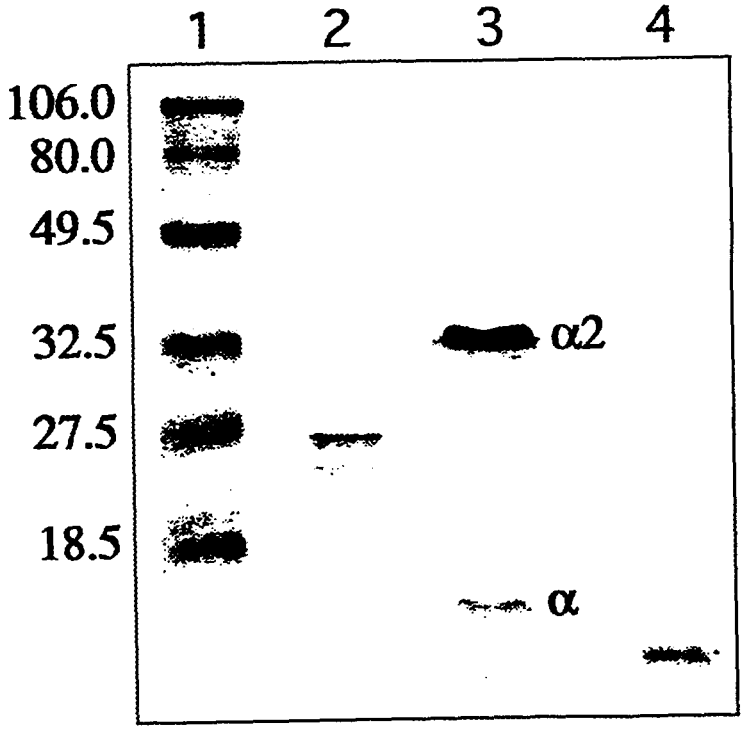

D
C
E
0.004
0.002
Absorbance (280 nm)
20
40
60
80
100
Time (min)

Fig. 2-13. Polyacrylamide gel electrophoresis of the proteases SGPC and SGPD. *Lane 1*, denatured molecular weight standards; *lane 2*, SGPC prepared under "standard denaturing conditions"; *lane 3*, SGPD prepared under standard denaturing conditions; *lane 4*, SGPD prepared under "nondenaturing conditions". Under denaturing conditions, SGPC has an apparent molecular mass of 26 kDa, whereas SGPD resolves into two bands with apparent molecular masses of 17 and 34 kDa, approximating monomeric (α) and dimeric (α_2) masses for the protease. Electrophoresis of SGPD prepared in nondenaturing conditions produced a single band clearly distinguishable from SGPD under denaturing conditions. Methods are described in section 2.17.


1
2
3
4
106.0
80.0
49.5
32.5
27.5
18.5
α2
α

single peak by gel filtration chromatography) resolved into two distinct bands with apparent molecular masses of approximately 17 kDa and 34 kDa, approximating monomeric (α) and dimeric (α_2) molecular masses for the protease. Indeed, even under denaturing SDS-PAGE conditions, SGPD exists predominantly in dimeric form. A sample prepared under nondenaturing conditions was also included in the analysis to establish the position of the native form of the enzyme after electrophoresis. The predominantly negatively charged SGPD (pI = 3.55, Fig. 2-14) had a high electrophoretic mobility in its undenatured form but migrated to a distinct position relative to the monomeric or dimeric forms of the denatured enzyme. Amino-terminal analysis of the 34 kDa band confirmed the position of the promature junction, ruling out the possibility that this band represents an unprocessed promature form of the enzyme.

3.5 Alignment and phylogenetic analyses

As a first step in the generation of a phylogenetic tree for the *S. griseus* and *L. enzymogenes* proteases, mature segments were aligned as shown in Fig. 2-15. This alignment of amino acid sequences was then used to generate the corresponding alignment of the nucleotide sequences. Phylogenetic relationships were examined using a parsimony method which selects a tree that requires the minimum number of mutational changes in order to explain the data set (section 2.20). The DNApars program used in the analysis performed unrooted parsimony (analogous to Wagner trees) on the nucleotide sequences and calculated the number of changes for each base required for a given tree. The result was a single most parsimonious tree

Fig. 2-14. Summary of the known S1 protease family members from *S. griseus* and *L. enzymogenes*. The pre-pro-mature organization of the six homologous proteases is illustrated. The homologous prepeptides are in *solid boxes*, propeptides are in *shaded boxes*, and mature protease domains are *unfilled*. The carboxyl- and amino-terminal extensions of the enzymes SGPC and SGPD are *cross-hatched*. To the *right* of each illustration is the quaternary structure of the mature protease, the isoelectric point (pI) deduced from sequence information using the computer program PC/Gene (IntelliGenetics, Inc., Mountain View, CA), and the primary specificity. References are as follows: SGPA and SGPB (James et al., 1980; Read et al., 1983; Henderson et al., 1987), SGPC (Sidhu et al., 1994), SGPD (Sidhu et al., 1995), SGPE (Nienaber et al., 1993; Sidhu et al., 1993), α-LP (Fujinaga et al., 1985; Silen et al., 1988).

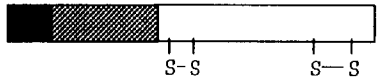
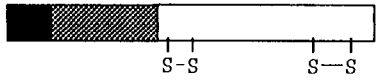
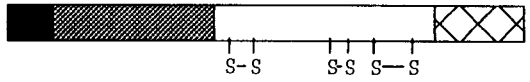
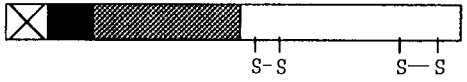
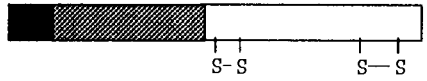
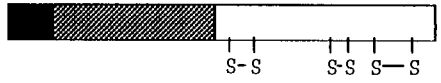
Enzyme	Zymogen Organization	Quaternary Structure	pI	Primary Specificity
SGPA		α	7.10	large aliphatic/ aromatic
SGPB		α	7.01	large aliphatic/ aromatic
SGPC		α	7.59	large aliphatic/ aromatic
SGPD		α_2	3.55	large aliphatic/ aromatic
SGPE		α	7.86	glutamic acid
α-LP		α	9.84	small aliphatic

Fig. 2-15. Alignment of protease amino acid sequences. The best alignment of A, the pro, and B, the mature domains of the proteases SGPA, B, D, E, and C and the α-lytic protease (α-LP) are shown. Regions of significant homology are indicated in *bold* and correspond to identities in at least 2 of 2, 3 of 3, 3 of 4, 4 of 5, or 4 of 6 of the aligned sequences.

A

```
SGPD pro                          SDDVPAFGAKTLSADAAGKLATTLDRDLGADAAGSYYDATAKTLVVNVVDEAGAEQVR·········
SGPE pro                           ADTPPAAPSAIPAPSAYALDAAVERQLGAATAGTYLDAKTGGLVVTVTTDRAEEQAR·········
SGPC pro APPVSADSLSPGMLAALERDLGLDEDAARSRIANEYRAAAVAAGLEKSLGARYAGARVSGAKATL··TVATTDASEAARITEAGARAE
α-LP pro      ADQVDPQLKFAMQRDLGIFPTQLPQYLQTEKLARTQAAAIEREFGAQFAGSWIERNEDGSFKLVAATSGARKSSTL·GGVEVR

SGPA pro    APEAESKATVSQLA·DASSAIL··AADVA·GTAWYTEASTGKIVLTADSTVSKAELAKVSNALAGSKAKLTVKRAEGKFTPL
SGPB pro      ETPRTFSANQLT·AASDAVL··GADIA·GTAWNIDPQSKRLVVTVDSTVSKAEINQIKKSAGANADALRIERTPGKFTKL
SGPD pro QAGGKARIVENSLAELK·SARGTLTE·KATIP·GTSWAVDPVSNKVLVTADSTVDGAAWKKLSAVVEGLGGKAELNRTAGEFTPL
SGPE pro AAGATVRRVARSAAQLD·AAMATLEA·EAKIT·GTSWGVDPRTNRVAVEADSSVSARDMARLEAVAERLGSAVDIKRVPGVFHRE
SGPC pro VVGHSLDRFEGVKKSLDKAALDKAPKN·····VPVWYVDVAANRVVV··NAASPAAGQA·FLKVAGVDRGLVTVARSAEQPRAL
α-LP pro NVRYSLKQLQSAMEQLDAGANARVKGVSKPLDGVQSWYVDPRSNAVVVKVDDGATDAGVD·FVALSGADSAQVRIESSPGKLQTT
```

B

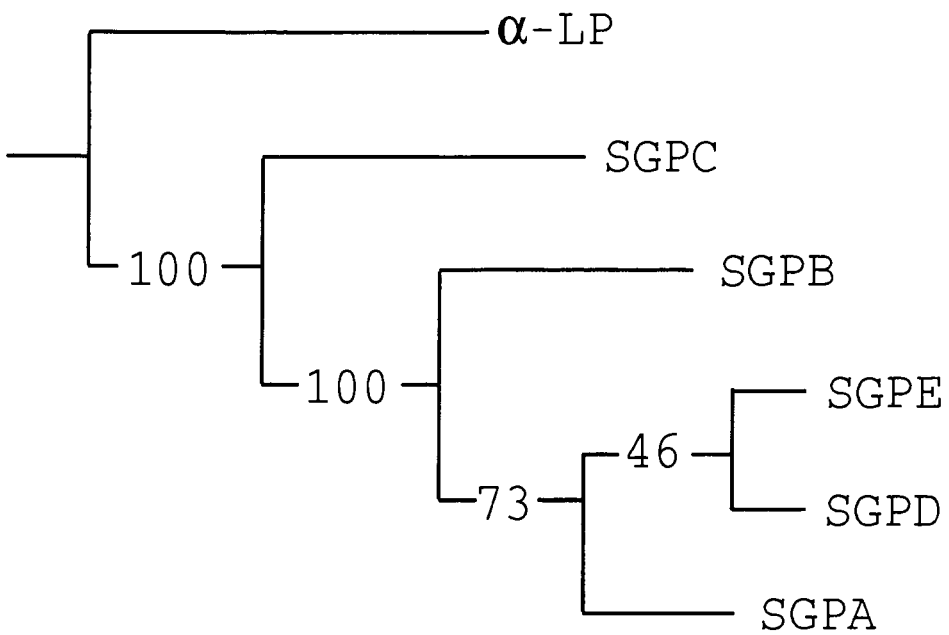
```
SGPA mat   IAGGEAITTG·GSRCSLGFNVSVNGVAHALTAGHCTNISASWS·····IGTRTGTSFPNNDYGIIRHSNPAAA·DGRVYLYNG
SGPB mat   ISGGDAIYSS·TGRCSLGFNVRSGSTYYFLTAGHCTDGATTWWANSARTTVLGTTSGSSFPNNDYGIVRYTNTTIPKDGTVG····
SGPD mat   IAGGDAIWGS·GSRCSLGFNVVKGGEPYFLTAGHCTESVTSWS·DTQGGSEIGANEGSSFPENDYGLVKYTSDTAH·PSEVNLYDG
SGPE mat   VLGGGAIYGG·GSRCSAAFNVTKGGARYFVTAGHCTNISANWS·ASSGGSVVGVREGTSFPTNDYGIVRYTDGSSP·AGTVDLYNG
SGPC mat ADIRGGDAYYMNGSGRCSVGFSVTRGTQNGFATAGHCGRVGTTTN··GVNQQAQGTFQGSTFPGRDIAWVATNANWTP·RPLVNGYGR
α-LP mat ANIVGGIEYSINNASLCSVGFSVTRGATKGFVTAGHCGTVNATAR···IGGAVVGTFAARVFPGNDRAWVSLTSAQTL·LPRVA··NG

SGPA mat S·YQDITTAGNAFVGQAVQRSGSTTGLRSGSVTGLNATVNYGSSGIVYGMIQTNVCAEPGDSGGSLFAGS·TALGLTSGGS·····GN
SGPB mat ··GQDITSAANATVGMAVTRRGSTTGTHSGSVTALNATVNYGGGDVVYGMIRTNVCAEPGDSGGPLYSGT·RAIGLTSGGS·····GN
SGPD mat S·TQAITQAGDATVGQAVTRSGSTTQVHDGEVTALDATVNYGNGDIVNGLIQTTVCAEPGDSGGALFAGD·TALGLTSGGS·····GD
SGPE mat S·TQDISSAANAVVGQAIKKSGSTTKVTSGTVTAVNVTVNYGD·GPVYNMVRTTACSAGGDSGGAHFAGS·VALGIHSGSS·····G·
SGPC mat G·DVTVAGSTASVVGASVCRSGSTTGWHCGTIQQLNTSVTYPE·GTISGVTRTSVCAEPGDSGGSYISGS·QAQGVTSGGS·····GN
α-LP mat SSFVTVRGSTEAAVGAAVCRSGRTTGYQCGTITAKNVTANYAE·GAVRGLTQGNACMGRGDSGGSWITSAGQAQGVMSGGNVQSNGNN

SGPA mat CR···TGGTTFYQPVTEALSAYGATVL
SGPB mat CS···SGGTTFFQPVTEALSAYGVSVY
SGPD mat CS···SGGTTFFQPVPEALAAYGAEIG
SGPE mat CS··GTAGSAIHQPVTEALSAYGVTVY
SGPC mat CS···SGGTTYFQPINPLLQAYGLTLVTS
α-LP mat CGIPASQRSSLFERLQPILSQYGLSLVTG
```

requiring 796 steps. Bootstrap analysis with DNAboot was used to place confidence limits on the phylogeny presented in Fig. 2-16.

Fig. 2-16. Phylogenetic tree of the bacterial proteases SGPA, B, C, D, and E and the α-lytic protease (α-LP). The tree was constructed using nucleotide sequences of the mature regions of the respective proteases. Nucleotide sequence alignments correspond to the amino acid alignment shown in Fig. 2-14B. Prepeptide and propeptide sequences and the sequence of the carboxyl-terminal domain of SGPC were not included in the analysis. The number at each of the forks represents the number of times that the particular grouping (consisting of the species to the right of the fork) was generated during the 100 bootstrap replicates. Methods are described in section 2-20.


α-LP
SGPC
100
SGPB
100
SGPE
46
73
SGPD
SGPA

4. DISCUSSION

This chapter describes the analysis of three *S. griseus* protease genes: *sprE* (the gene encoding SGPE) and two new genes, which we named *sprC* and *sprD*. An *E. coli/B. subtilis* shuttle vector (pEB11) allowing for genetic manipulations in *E. coli* and protease secretion in *B. subtilis* was constructed and used to produce SGPB, SGPE, and also SGPC and SGPD (the gene products of *sprC* and *sprD*, respectively). Subsequently, protocols were developed for recombinant protease purification; preliminary characterization of SGPC and SGPD is reported. Finally, the protein sequences obtained in this study were aligned along with the sequences of previously characterized proteases and the alignment was used to generate a phylogenetic tree describing the evolution of the S2 serine protease family.

4.1 Hybridization studies

The degree of homology between the known *S. griseus* protease genes suggested that members of the family should cross-hybridize with one another. Therefore, we employed a genetic approach designed to detect genes homologous to *sprB* (the gene encoding SGPB). A DNA fragment encoding almost the entire mature SGPB (amino acids 9 to 185) was used to generate radiolabeled probes using a random-priming method. Southern blot analysis of genomic *S. griseus* DNA detected the genes *sprA, sprB,* and *sprE,* and the band intensities were proportional to homology (Fig. 2-4). The *S. griseus* trypsin gene was not detected, but this is not surprising since SGT's homology to SGPB is significantly less than that of the other proteases (Read & James,

1988). Two putative genes which could not be accounted for on the basis of extant literature (*sprC* and *sprD*) were also detected.

4.2 Organization of the gene products

4.2.1 The protease gene products are produced in a pre-pro-mature form

Sequence comparisons of the *sprC*, *sprD*, and *sprE* gene products with those of previously characterized protease genes (Figs. 2-14 and 2-15) suggest a similar pre-pro-mature organization. While the prepeptide sequence of SGPD is unusual (section 3.2.3), the prepeptides of SGPC and SGPE are typical of bacterial secretion signals. They consist of a positively charged amino terminus followed by a hydrophobic region capable of forming a membrane spanning helix, and terminate in a signal peptidase recognition site (von Heijne, 1986). In all three gene products, the preregion is followed by a propeptide; propeptides are found in most bacterial proteases and have been shown to be essential for correct folding in subtilisin (Ohta et al., 1991) and α-LP (Silen et al., 1989).

4.2.2 Evidence for autocatalytic promature junction cleavage

The maturation of the S2 family proteases involves two proteolytic cleavages. The first cleavage, at the junction between the prepeptide and the propeptide, requires the action of a leader peptidase (von Heijne, 1986). The second cleavage, occurring at the promature junction, is thought to be autocatalytic (Henderson et al., 1987). The sequence alignment in Fig. 2-15 supports this supposition since the promature junction P1 residue of each

S. griseus protease is a good substrate for autocatalytic cleavage. The presence of a Leu at the promature junction P1 positions of *S. griseus* proteases A, B, C, and D correlates with the chymotrypsin-like primary specificity of these enzymes (Bauer, 1978) (section 3.4.2). In contrast, the optimal P1 substrate for SGPE is glutamic acid and *sprE* encodes a glutamic acid residue at the promature junction P1 site. Recombinant expression of *sprB, sprC, sprD,* and *sprE* in *B. subtilis* provided further evidence for autocatalytic maturation; a foreign host is unlikely to possess processing enzymes adapted for precise activation of the *S. griseus* proteases, yet each recombinant enzyme was correctly processed.

4.3 *sprE* and SGPE

The amino acid sequence deduced from *sprE* indicates that mature SGPE is 187 amino acids in length, and notably, differs from the previously published amino acid sequence (Svendsen et al., 1991) at three positions. Residues Val132 and Glu177 were previously reported to be Gly and Lys respectively, and the published amino acid sequence contained an additional carboxyl-terminal Leu residue not present in our sequence. We have repeatedly sequenced the regions in which the discrepancies occur, and therefore, we have a high degree of confidence in the derived amino acid sequence presented here.

4.4 *sprC* and SGPC

The sequence of *sprC* predicts that mature SGPC is 255 amino acids in length and consists of two domains connected by a 19-amino acid linker: an

amino-terminal protease domain of 189 amino acids and a 47-amino acid carboxyl-terminal domain with significant homology to putative chitin-binding domains of chitinases A1 and D of *B. circulans* (Fig. 2-7).

4.4.1 Relationship to the α-lytic protease

Sequence comparisons (Fig. 2-15) suggest that, of all the *S. griseus* proteases, SGPC is the one most closely related to the α-lytic protease of *L. enzymogenes*. The propeptide of SGPC (162 amino acids) is almost identical in length to that of α-LP (166 amino acids), while in contrast, the propeptides of SGPA and SGPB contain only 78 and 76 residues respectively. The propeptides of SGPE (139 amino acids) and SGPD (140 amino acids) lie between these two extremes. In addition, the protease domains of SGPC and mature α-LP each contain six cysteine residues which occur in homologous positions; in the crystal structure of α-LP they form three disulfide bridges. SGPA, SGPB, and SGPE each contain only four of the homologous cysteine residues. The crystal structures of SGPA (James et al., 1980) and SGPB (Read et al., 1983) reveal that these residues are involved in disulfide bridges which are topologically equivalent to two of the three bridges in α-LP. Thus it seems likely that SGPC, like α-LP, contains a third disulfide bridge which is absent in *S. griseus proteases* A, B, and E. Finally, SGPC and α-LP possess an additional two amino acids at their mature amino-terminal ends as compared to SGPA, SGPB and SGPE. The sequence of SGPC begins Ala-Asp while α-LP begins with the homologous sequence Ala-Asn.

The similarities between SGPC and α-LP were reflected in phylogenetic analyses of α-LP and the *S. griseus* proteases (section 4.6).

4.4.2 Relationship to chitinases

The Pronase-derived proteases SGPA and SGPB were originally detected by virtue of their activities against synthetic substrates. The substrate specificity of SGPC is similar to that of SGPA and SGPB, and the open reading frame of *sprC* encodes a polypeptide with a well-defined secretion signal. Why then has SGPC never been detected in Pronase? We believe that the explanation may involve the unusual structure of mature SGPC. The presence of a carboxyl-terminal domain absent in all other known proteases suggests that SGPC has a specialized function, and therefore, it may not be secreted constitutively. Unfortunately, the initiation codon of *sprC* is only 34 nucleotides from one end of the *Bam*HI fragment characterized in this study, and thus the promoter and any upstream regulatory sequences of *sprC* are not contained in the fragment. Sequencing of upstream regions and Northern blot analysis of *S. griseus* mRNA with *sprC* DNA as probe should help to answer questions concerning regulated expression.

The carboxyl-terminal domain of SGPC is homologous to putative chitin-binding domains of chitinases A1 and D of *B. circulans* (Watanabe et al., 1992). Recombinant SGPC purified from *B. subtilis* has a molecular mass of 26 kDa as determined by SDS-PAGE. This is in good agreement with the mass predicted from the deduced amino acid sequence, demonstrating clearly that the carboxyl-terminal domain remains stably attached to the active protease. The small linker connecting the domain to the proteolytic domain is rich in threonines and prolines; the presence of prolines is noteworthy as this structure may serve to resist proteolysis.

What is the relationship between SGPC and chitin? Chitin, a eukaryotic structural polysaccharide composed of repeating β(1-4)-linked *N*-

acetyl-D-glucosamine units, is found complexed with proteins in insects, fungi, and nematodes. For example, chitin and protein are present in insect cuticles in equal proportions by mass (Ruiz-Herrera, 1991). Hence, the efficient degradation of chitin in the environment of *S. griseus* is expected to require the participation of proteases. Notably, *S. griseus* secretes a combination of chitinolytic and proteolytic enzymes. SGPC seems to be an enzyme which possesses features of both enzymatic classes, but it is not immediately clear why a protease requires a carbohydrate-binding domain in order to degrade chitin-linked proteins. An answer to this question may be found in the observation that the chitin-binding domains of *B. circulans* chitinases are essential in the hydrolysis of insoluble crystalline chitin but are not required for the hydrolysis of soluble chitin (Watanabe et al., 1990). SGPA, SGPB, and SGPC have very similar substrate specificities, and we speculate that chitin-linked polypeptides may represent poor substrates for these proteases. Hence, the acquisition of a chitin-binding domain by SGPC would be an expedient method of improving specificity and activity. Perhaps, as has been observed with the cellulose-binding domains of certain cellulases (Din et al., 1991; Tapani et al., 1992), the chitin-binding domain also participates in unraveling the polymers, thereby giving the degradative enzymes access to their substrates. The carbohydrate binding capabilities of SGPC and its activity in the degradation of chitin-linked proteins remain to be investigated.

4.5 *sprD* and SGPD

4.5.1 Quaternary structure of SGPD

Recombinant SGPD isolated from *B. subtilis* culture supernatants proved to be much larger than anticipated from the nucleotide sequence of *sprD*. SDS-PAGE gave two bands corresponding to proteins with molecular masses of approximately 34 and 17 kDa (Fig. 2-13). Gel filtration (size exclusion) chromatography subsequently established that the native enzyme exists in the form of a stable homodimer (Fig. 2-12). The molecular mass of a monomeric SGPD should be 18.7 kDa according to sequence data, and consequently, a homodimer should have have a mass of roughly 36 kDa. The high mobility of the protein in SDS-PAGE is most likely due to the high negative charge on the protein (Figs. 2-8 and 2-14). Amino-terminal analysis of the protein ruled out the possibility that the 36-kDa band corresponded to an unprocessed promature form of the protein (section 3.4.3).

It is remarkable that SGPD should have such a high degree of homology to its monomeric cousins and yet form such a stable dimer. Hence, the transition from monomer to dimer (or *vice versa*) involves relatively few residues in the protein. Given the high negative charge on SGPD, one might expect the monomers to repel one another. Therefore, dimerization may involve metal chelation, the formation of intermolecular salt bridges, or both. The physical basis for the extraordinary stability of the SGPD dimer in denaturing conditions remains to be examined.

4.5.2 Implications of the unusual SGPD prepeptide

In prokaryotes, the prepeptide acts to signal translational secretion of extracellular enzymes. Containing 62 amino acids, The predicted prepeptide of SGPD (Fig. 2-8) is significantly longer than that of other bacterial proteases, and it can be divided into two parts: an amino-terminal segment with characteristics of a mitochondrial import signal and a carboxyl-terminal segment characteristic of bacterial secretion signals (section 3.2.3). These characteristics are most evident when the prepeptide is aligned with the prepeptides of SGPB (Henderson et al., 1987) and the mitochondrial heat shock protein hsp60 (Jindal et al., 1989; Venner & Gupta, 1990) (Fig. 2-9). We are aware of no other prokaryotic enzyme with this type of signal sequence. The unusual organization of the SGPD prepeptide suggests that the protease has a function distinct from that of other *S. griseus* proteases.

The unusual prepeptide of SGPD may target the enzyme to a subcellular location; the notion is supported by the fact that SGPD has never been observed in *S. griseus* secretions even though the substrate specificity of the enzyme is similar to that of the well characterized enzyme SGPB (Fig. 2-11). Prokaryotes are not known to contain specific organelles; however subcellular compartments, or mesosomes, have been observed in *streptomyces* (Kurylowicz et al., 1975) and other genera of bacteria (Horne & Tomasz, 1985; Cherepova et al., 1986; Nakasone et al., 1987). Although the significance of mesosomes is controversial, they may have functions similar to the periplasmic spaces of Gram-negative bacteria or even the organelles of eukaryotic cells (Greenwalt & Whiteside, 1975). It is tempting to speculate that SGPD is directed to one of these structures.

The unusual signal sequence of SGPD has implications for the endosymbiont hypothesis which proposes that mitochondria are derived from bacteria. According to this hypothesis, the "mitochondrial" genes of a proto-eukaryotic cell were moved to the nucleus (Ostermann, 1990) where they had targeting sequences attached to them. Therefore, similarities between bacterial and mitochondrial targeting are to be expected. The fact that the prepeptide of SGPD contains features of both mitochondrial and prokaryotic signal sequences lends support to the endosymbiont hypothesis, but it also implies that so-called "mitochondrial" targeting sequences predate the existence of mitochondria (see below).

4.6 Evolutionary relationships

The chymotrypsin clan (SA) has been subdivided into several distinct families (Rawlings & Barrett, 1994). The majority of eukaryotic SA enzymes belong to the chymotrypsin family (S1) while the majority of prokaryotic SA enzymes belong to the α-lytic protease family (S2). Propeptide length and function provide features which are valuable in distinguishing between members of the two families. S1 enzymes possess small propeptides (amounting in some cases to a few amino acids) while S2 enzymes possess large propeptides comparable in length to the mature enzymes. Studies with α-LP have demonstrated the importance of the propeptide in catalyzing the proper folding and maturation of S2 enzymes (Silen & Agard, 1989; Silen et al., 1989). In contrast, the function of the propeptide in S1 enzymes is to block the amino terminus of the protease, holding the enzyme in an inactive state until the propeptide is cleaved from the zymogen. Hence, S1 propeptides appear not to be involved in "catalyzing" the folding process.

There is considerable variation in the lengths of propeptides even within the group of bacterial enzymes compared in Figs. 2-14 and 2-15. According to propeptide length, SGPA and SGPB appear to fall into one group, SGPD and SGPE in another group, and SGPC and α-LP in a third. This arrangement is also reflected in the phylogeny produced using the parsimony (DNApars) and bootstrap (DNAboot) analyses of the mature domains of each protein (Fig. 2-16).

The phylogenetic tree shown in Fig. 2-16 indicates that SGPC and α-LP have diverged the most from the other proteases in the analysis. The six proteases form a monophyletic group beginning with α-LP and followed by SGPC, SGPB, SGPA, and finally SGPD and SGPE. DNA boot-strap analysis strongly supports the relationships, placing excellent confidence limits on the branches between α-LP, SGPC, SGPB, and the tricotomy formed between SGPB and the remaining three proteases (SGPA, SGPD, and SGPE).

We believe that α-LP and SGPC are the two most ancient proteases in our study, primarily because the two enzymes have the most extensive propeptides. Notably, these are also the only two enzymes with three disulfide bonds instead of two. It can be argued that the presence of the two homologous proteases in different genera of bacteria is evidence that they arose from a common ancestor before the organisms diverged.

S. griseus trypsin (SGT) is unusual in that it is produced by a prokaryote but in terms of sequence and structure it is grouped with the predominantly eukaryotic S1 enzymes. The propeptide of SGT is four residues in length (Kim et al., 1991), similar to the propeptides of mammalian trypsins (four residues) (Huber & Bode, 1978) and notably, the gene encoding SGT was not detected by hybridization with *sprB*. The anomolous relationship of SGT to prokaryotic and eukaryotic enzymes has even led some authors to speculate

that SGT was acquired from a mammalian source only recently (Hartley, 1970; Hartley, 1979). Perhaps a more satisfactory explanation is that the propeptides of bacterial proteases are becoming shorter through the course of evolution and SGT is simply furthest of the *S. griseus* enzymes from the ancestral.

A comparison of the prokaryotic and eukaryotic S1 family members produced a dendogram rooted at the bacterial sequences; divergence was estimated at approximately 1300 million years ago (Young et al., 1978). The divergence was far more recent than the estimated divergence between eukaryotes and prokaryotes (3500 million years ago), but instead corresponded to the time at which eukaryotes acquired mitochondria. A more recent study confirmed the date and it was suggested that the eukaryotic S1 enzymes arose through horizontal gene transfer from the proto-mitochondrion (Rawlings & Barrett, 1994). The results of our study reveal that *S. griseus* contains not only a member of the S1 family (*S. griseus* trypsin) but also a protease (SGPD) with a prepeptide homologous to mitochondrial import signals. Thus, horizontal gene tranfer from an *S. griseus*-like proto-mitochondrion could account for both the S1 serine proteases and also the mitochondriol import system of eukaryotic organisms.

CHAPTER 3: DIRECTED EVOLUTION OF SUBSTRATE SPECIFICITY

1 INTRODUCTION

Advances in DNA manipulation techniques have made the engineering of novel substrate specificities into natural enzymes a reality. Site-specific substitutions at key positions selected through the study of sequence and structural data have yielded both impressive (Estell et al., 1986; Fersht, 1987) and unexpected results (Graf et al., 1987). In general, our limited understanding of the complex interactions responsible for an enzyme's conformational stability and catalytic activity does not allow for accurate predictions of mutational effects. Furthermore, an enzyme's catalytic mechanism involves the cooperation of numerous functional groups; thus a number of changes may be necessary to alter specificity while maintaining catalytic efficiency (Graf et al., 1987; Graf et al., 1988). For these reasons, we sought methods for rapidly selecting mutant enzymes with specified, altered activities from large libraries containing variations at multiple sites. Such techniques could be used both in the engineering of enzymes for practical applications, and also as tools for probing general principles of enzyme structure and function.

Serine peptidases, proteolytic enzymes with catalytic activity dependent on a serine residue, are grouped into over 20 families which in turn are grouped into six clans or superfamilies (Rawlings & Barrett, 1994). The chymotrypsin clan (SA) in particular provides an ideal system for the study of enzyme evolution since the structural framework supports primary specificities which span the range of natural amino acids. Indeed, the

extensive resources of structure and sequence data have allowed for the design of several site-specific studies of primary specificity (see chapter 1).

We have developed a novel strategy for probing protease substrate specificity and describe its use in selectively altering the primary specificity of *Streptomyces griseus* protease B (SGPB) towards various substrates. SGPB is a member of the α-lytic protease family (S2) and thus belongs to the SA clan. All members of the S2 family are translated as pre-pro-mature precursors. The prepeptide directs co-translational secretion from the cell and is removed by signal peptidase (von Heijne, 1986). The propeptide is essential for proper folding but is not present in the mature enzyme; self-processive removal of the propeptide is essential for the generation of active protease. Using a visual screening strategy allowing for the detection of active protease secretion, we screened an expression library encoding 29,952 possible SGPB variants with mutations at seven sites implicated in conferring primary specificity. All active variants isolated from the screen had catalytic efficiencies and thermal stabilities comparable with those of wild type SGPB, and significantly, all retained primary specificity for large hydrophobic residues which correlated with the Leu residue at the promature junction P1 site (pmP1). Thus, it was demonstrated that the primary specificity of recombinantly expressed SGPB is constrained by the sequence at pmP1; only those mutants capable of efficient self-processing were capable of producing detectable activity in *E. coli.*

We exploited this relationship through a two-step mutagenesis strategy to select for SGPB mutants with selectively altered primary specificities. In the first step, the wild type Leu at pmP1 was genetically substitued with various amino acids and protease secretion in *E. coli* was monitored. The introduction of poor SGPB substrates at pmP1 reduced or abolished active

protease secretion. In the second step, the original expression library was screened in conjunction with the promature junction mutants for self-processing, catalytically active variants. SGPB mutants were expressed in *B. subtilis* and purified to homogeneity; substrate specificities and thermostabilities were investigated. In general, increases in proteolytic activity correlated with increases in specificity for the residue at pmP1 and thermostabilities were comparable to wild-type. Variant enzymes with increases in specificity towards various hydrophobic and polar residues were isolated. Variant enzymes were not obtained with charged residues at pmP1, suggesting the need for more drastic mutations. This chapter discusses the results of these experiments and implications for the design of enzymes with tailored substrate specificities, improved catalytic efficiency, and increased structural stability.

2 MATERIALS AND METHODS

The materials and methods below are in addition to Chapter 2.

2.1 Enzymes and reagents

Suc-AAPS-pNA was a gift from Dr. C. A. J. Braun, Simon Fraser University, Vancouver, Canada.

2.2 Strains and plasmids

E. coli TB1 and plasmid pMAL-p (Guan et al., 1987; Maina et al., 1988) were obtained from New England Biolabs. *E. coli* TXK-11 (Racher, 1996) was a gift from Ms. K. Racher, Simon Fraser University, Vancouver, Canada.

2.3 Media and growth conditions

For visual screening of protease secretion, *E. coli* were grown on YT plates containing 200 μg/mL of ampicillin, 60 μM IPTG, and 2% w/v skim milk powder (YT/milk plates). For protease expression, *B. subtilis* were grown in YTC broth (Chapter 2, section 2.3) without charcoal treatment; cultures were grown at 30 ^{o}C in a rotary shaker at 200 rpm.

2.4 Electroporation

E. coli were prepared for electroporation as follows. Frozen stocks were streaked on YT plates and grown overnight at 37 °C. A single colony was used to inoculate 2 mL of YT broth, grown overnight at 37 °C with vigorous shaking, and then used to inoculate 1.0 L of YT broth in a 2-L Erlenmeyer flask. The cultures were grown to an optical density of 0.8 at 600 nm and then centrifuged for 5 min at 4000 x g; from this point, all procedures were performed on ice or at 4 °C. The supernatant was decanted and the pellet was resuspended in an equal volume of sterile ddH_2O. The culture was centrifuged as above and the supernatant was discarded. The ddH_2O wash was repeated with a half volume of ddH_2O and then with 20 mL of ddH_2O. Finally, the pellet was resuspended in 1.0 mL of ddH_2O and used for electroporation; the final yield was 1.4 mL of resuspended *E. coli*.

E. coli were electroporated using a Bio-rad Gene Pulser according to the manufacturer's instructions. In a typical electroporation, 200 µL of cells were used in a 1-mL electroporation cell with the following settings: voltage = 2.5 kV; capacitance = 25 µFd; and resistance = 200 Ω.

2.5 Oligonucleotides

The following oligonucleotides (listed from 5' to 3') were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer (nucleotides in brackets indicate equimolar degeneracies; complementary regions of SB1 and SB2 are in ***bold***; identical sequences in SB1 and SB3 or in SB2, SB4, and SB5 are *underlined*):

BF1: GGTACCTGCGCAACCCCCCGGACG

BF2: CTGCAGGCGGCGACGCGATCTACTCC

BR1: AAGCTTTCTAGATCAGTACACGCTGACGCCG

P423: GGATCCAGTACTGATC(CGT)(ACT)CTTGGTGAACTTCC

P508: GGATCCAGTACTGAT(ACGT)TCCTTGGTGAACTTCC

P853: GGATCCAGTACTGAT(CG)T(GT)CTTGGTGAACTTCC

P857: GGATCCAGTACTGAT(CG)A(ACT)CTTGGTGAACTTCC

P890: GGATCCAGTACTGAT(CG)(CT)ACTTGGTGAACTTCC

SB1: AGAGGATCCGACGTCGTCTACGGCATGATCCGCACCAACGTGTG

C(AGT)(CT)GG(AGC)G(GC)(GC)GGGCGACTCCGGCGGTCCGC(AT)CTACTC

CGGCACCCGGGCGATCGG

SB2: TTTTCTAGAGCCTGCAGTCAGTACACGCTGACGCCGTACGCGCTC

AGCGCCTCGGTGACCGGCTGG(AT)(AG)GAAGGTCGTGCCGCCGGAGGAG

CAGTTGCCGCTG(GC)(CT)GCCGGA(GC)(AGT)(GCT)CAGA**CCGATCGCCCG**

GGTGCC

SB3: AGAGGATCCGACGTCGTCTAC

SB4: TTTTCTAGAGCCTGCAGTCAG

SB5: GGAATTCAAGCTTTTCTAGAGCCTGCAG

2.6 Expression of SGPB in *E. coli* (construction of pMP-B)

Plasmid pDS-B8 (Chapter 2, section 2.13.2.1) was digested with *Fsp*I and *Pst*I, and the fragment encoding promature SGPB was gel purified and ligated into pMAL-p (digested with *Stu*I and and *Pst*I and treated with CIP). The ligation mixture was used to transform *E. coli* TB1, and a vector containing the correct insert in the correct orientation (isolated on the basis of restriction enzyme analysis) was designated pMP-B. Expression and secretion of active SGPB from *E. coli* TB1 harbouring pMP-B was verified on YT/milk plates.

2.7 Construction of *sprB* genes with substitutions at pmP1

Oligonucleotides BF2 and BR1 were used as PCR primers with the plasmid pNC-B (a pUC19 derivative containing *sprB*) (Carson, 1994) as template to amplify a fragment of *sprB* encoding from the third residue of mature SGPB to the termination codon; the amplified 568 bp product was ligated into pUC18 that had been digested with *Pst*I, treated with T4 DNA polymerase to produce blunt ends, and treated with CIP. The ligation mixture was used to transform *E. coli* TB1 and a vector with the correct insert oriented with the BF2 priming site closest to the *Eco*RI site of pUC18 was identified on the basis of restriction digests. Correct amplification was verified by DNA sequencing and the plasmid was designated pMB-P.

2.7.1 Phe, Ile, Met, or Val at pmP1

Oligonucleotides BF1 and P857 were used as PCR primers with plasmid pNC-B as template to amplify fragments of *sprB* encoding from the second residue of the SGPB propeptide to the second amino acid of mature SGPB (with substitutions at pmP1). The amplified 252 bp product was digested with *Kpn*I and *Bam*HI and ligated into similarly digested pUC18 that had been treated with CIP. Vectors with inserts of the correct size were identified by restriction analysis; plasmids containing *sprB* fragments encoding SGPB propeptides with Phe (pPM-75), Ile (pPM-72), Met (pPM-76), or Val (pPM-77) at the promature junction P1 site were identified by DNA sequencing.

Plasmid pPM-n (where n = 72, 75, 76, or 77) was digested with *Eco*RI and *Sca*I and the 252 bp band was ligated into pMB-P which had been digested with *Pst*I, treated with T4 DNA polymerase to produce blunt ends,

digested with *Eco*RI, and treated with CIP. The ligation mixture was transformed into *E. coli* TB1 and vectors with the correct inserts were identified on the basis of restriction enzyme digests; vectors derived from pPM-75, pPM-72, pPM-76, or pPM-77 were designated pDS-BF, pDS-BI, pDS-BM, or pDS-BV respectively. Plasmid pDS-BX (where X = F, I, M, or V) was digested with *Fsp*I and *Hind*III, and the 792 bp band, containing *sprB* with a mutant promature P1 site, was ligated into pMAL-p which had been digested with *Stu*I and *Hind*III and treated with CIP; the ligation mixture was used to transform *E. coli* TB1. Plasmids containing inserts were identified on the basis of restriction analyses and verified by sequencing the promature region of the *SprB* open reading frame. Plasmids derived from pDS-BF, pDS-BI, pDS-BM, or pDS-BV were designated pMP-BF, pMP-BI, pMP-BM, or pMP-BV respectively.

2.7.2 Ala, Gly, Lys, Ser, or Thr at pmP1

The procedures in section 2.7.1 were repeated using P423 in place of P857. Plasmids containing *sprB* fragments encoding SGPB propeptides with Ala (pPM-3), Gly (pPM-28), Lys (pPM-2), Ser (pPM-11), or Thr (pPM-6) at pmP1 were identified by DNA sequencing.

pUC18 or pMAL-p derivatives containing *sprB* genes with Ala, Gly, Lys, Ser, or Thr at pmP1 were named pDS-BX or pMP-BX (where X = A, G, K, S, or T).

2.7.3 Asp or Glu at pmP1

The procedures in section 2.7.1 were repeated using P508 in place of P857. Plasmids containing *sprB* fragments encoding SGPB propeptides with Asp (pPM-41) or Glu (pPM-37) at pmP1 were identified by DNA sequencing.

pUC18 or pMAL-p derivatives containing *sprB* genes with Asp or Glu at pmP1 were named pDS-BX or pMP-BX (where X = D or E).

2.7.4 His, Asn, or Gln at pmP1

The procedures in section 2.7.1 were repeated using P853 in place of P857. Plasmids containing *sprB* fragments encoding SGPB propeptides with His (pPM-62), Asn (pPM-67), or Gln (pPM-65) at pmP1 were identified by DNA sequencing.

pUC18 or pMAL-p derivatives containing *sprB* genes with His, Asn, or Gln at pmP1 were named pDS-BX or pMP-BX (where X = H, N, or Q).

2.7.5 Cys, Trp, or Tyr at pmP1

The procedures in section 2.7.1 were repeated using P890 in place of P857. plasmids containing *sprB* fragments encoding SGPB propeptides with Cys (pPM-94), Trp (pPM-95), or Tyr (pPM-89) at pmP1 were identified by DNA sequencing.

pUC18 or pMAL-p derivatives containing *sprB* genes with Cys, Trp, or Tyr at pmP1 were named pDS-BX or pMP-BX (where X = C, W, or Y).

2.8 Mutant *sprB* libraries

2.8.1 Library construction

2.8.1.1 Wild-type promature junction

A 50 µL reaction (1X vent DNA polymerase buffer containing SB1 and SB2 at 20 nM each; SB3 and SB4 at 0.80 µM each; dATP, dCTP, dGTP, and dTTP at 0.50 mM each; and 1.0 units of vent DNA polymerase) was subjected to the following thermocycle in a Biometra TRIO-thermoblock: step 1) 97 °C for 120 s, step 2) 97 °C for 30 s, step 3) 50 °C for 30 s, step 4) 72 °C for 30 s; Cycle to step 2) 35 times. The contents of 60 such reactions were combined and a 215 bp DNA fragment (encoding a fragment of mature SGPB extending from residue 175 to the carboxyl terminus with variations at seven sites, section 3.1) was gel purified.

The 215 bp fragment was digested with *Aat*II and *Pst*I and ligated into pMP-B (section 2.5) which had been digested with *Aat*II and *Pst*I and treated with CIP (200 ng insert and 2.0 µg vector in a total ligation volume of 30 µL, incubated with 1.0 units of T4 DNA ligase at 16 °C for 24 h). The ligation mixture was used to electroporate *E. coli* TXK-11 and the electroporated bacteria were flash frozen in 15% glycerol.

2.8.1.2 Mutant promature junctions

The 215 bp DNA fragment described above (section 2.8.1.1) was used as template in a PCR with SB3 and SB5 as primers (in order to add *Hin*dIII and *Eco*RI sites at the 3' end of the gene fragment). Twenty such independent

reactions (50 µL each) were combined and the 226 bp product was gel purified. The 226 bp product was digested with *Aat*II and *Hind*III and ligated into pMP-BX (where X = A, D, E, F, G, H, K, M, N, Q, S, T, or V) which had been similarly digested and treated with CIP (100 ng isert and 250 ng vector in a total ligation volume of 60 µL, incubated with 1.0 units of T4 DNA ligase at 16 ^{o}C for 24 h). The ligation mixture was used to electroporate *E. coli* TXK-11 and the electroporated bacteria were flash frozen in 15% glycerol.

2.8.2 Library screening

Representatives from the above libraries were plated on YT/milk plates (100 x 15 mm) at a density of 1.0×10^3 cfu/plate. The plates were incubated 16 h at 30 ^{o}C and then 48 h at 20 ^{o}C. After every 24 hours at 20 ^{o}C, colonies with distinct zones of clearing were restreaked on fresh plates. The restreaked plates were incubated as per the original plates. Plasmid DNA isolated from individual colonies with reproducible zones of clearing was transformed into *E. coli* TB1. Plasmid DNA was isolated from *E. coli* TB1 transformants exhibiting zones of clearing. Mutant plasmids were identified on the basis of restriction analysis and the insert regions were sequenced.

Mutant plasmids were named "pMP-BXrn" where "pMP-B" is the parent plasmid, "X" denotes the promature junction P1 site (if mutated), and "n" is a numeral distinguishing different variants with the same promature junction P1 site.

2.9 Expression and purification of proteases from *B. subtilis*

Each selected pMAL-p derivative carrying a mutant *sprB* gene was first digested with *Eag*I and treated with T4 DNA polymerase to produce blunt ends and then digested with *Pst*I. The fragment containing the mutant *sprB* gene was gel purified and ligated into pEB11 (Chapter2, section 2.13.1) which had been digested with *Sma*I and *Pst*I and treated with CIP. The ligation mixture was used to transform *E. coli* TB1, and vectors containing the correct insert in the correct orientation were identified on the basis of restriction analysis. The pEB-11 derivatives were transformed into *B. subtilis* DB104 (Kawamura & Doi, 1984). pEB11 derivatives were named "pEB-BXrn" and mutant proteases were named "SGPBXrn" in correspondence with the "pMP-BXrn" parent plasmid (section 2.8.2).

B. subtilis DB104 transformants from frozen stocks were streaked on YT plates containing 50 µg/mL kanamycin and grown overnight at 30 °C. A single colony was used to inoculate 2 mL of YT broth containing 50 µg/mL kanamycin, grown approximately 12 h at 30 °C with vigorous shaking, and then used to inoculate 200 mL of YTC broth containing 50 µg/mL kanamycin in a 500-mL Erlenmeyer flask. After 60 h growth at 30 °C in a rotary shaker at 200 rpm, the culture was centrifuged for 30 min at 10,000 x g and the pellet was discarded. Ammonium sulphate was added to the supernatant to a concentration of 2.0 M. After stirring for 10 min, the mixture was centrifuged at 14,000 x g for 15 min, and the pellet was discarded. Ammonium sulphate was added to the supernatant to a final concentration of 3.0 M and the mixture was stirred and centrifuged as above. The pellet, which contained active protease, was resuspended in 5.0 mL of 50 mM sodium acetate (pH 4.8). The sample was next subjected to acetone fractionation and cation exchange

chromatography as described for SGPB (Chapter 2, section 2.13.3.1). All enzymes were purified to homogeneity as evidenced by SDS-PAGE (Chapter 2, section 2.16).

2.10 Determination of protease thermostability

Enzyme samples (50.0 μL of approximately 5 nM enzyme solution in 50 mM Tris, pH 8.0) were incubated at 55 °C for various periods of time and then placed on ice. Enzyme activity at 20 °C was measured spectrophotometrically at 412 nm using the following assay mixture: 1.00 mL of 160 μM Suc-AAPF-pNA in 50 mM Tris buffer, pH 8.0 + 10 μL of enzyme solution.

2.11 Determination of kinetic parameters

Enzyme assays were performed at 20 °C in 50 mM tris, 5% methanol (pH 8.0) and were monitored spectrophotometrically at 412 nm.

Enzyme and substrate concentrations used were as follows: Suc-AAPF-pNA (15-600 μM) with enzyme at 5-250 nM, Suc-AAPL-pNA (65-700 μM) with enzyme at 15-250 nM, Suc-AAPM-pNA (25-300 μM) with enzyme at 10-35 nM, Suc-AAPV-pNA (85-650 μM) with enzyme at 50-150 nM, Suc-AAPA-pNA (150-600 μM) with enzyme at 50-150 nM, and Suc-AAPS-pNA (50-150 μM) with enzyme at 50 nM.

Kinetic constants and their associated standard errors were determined using the computer program "Enzfitter" (R. J. Leatherbarrow). In cases where the substrate concentrations were insufficient to allow an accurate fit to the Michealis-Menten equation, k_{cat}/K_m values were determined from a linear

regression analysis at substrate concentrations far less than K_m. All other values were determined from a non-linear regression analysis of the Micheals-Menten equation. Protein concentrations were determined by the method of Lowry (Lowry et al., 1951) ; We assumed that all protein was active enzyme. Molecular weights were calculated using the computer program PC/Gene (IntelliGenetics, Inc., Mountain View, Calif.) from the primary sequence of mature SGPB (Henderson et al., 1987) with mutations taken into account.

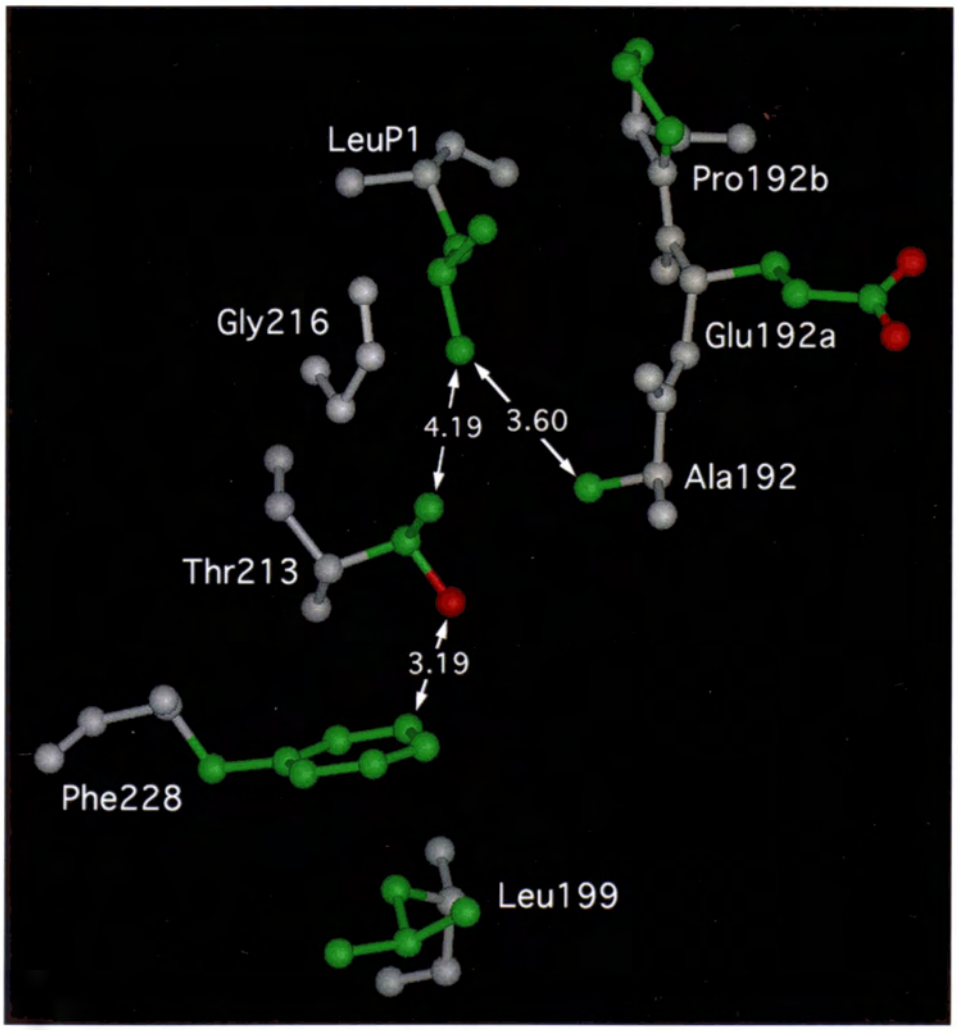
3 RESULTS AND DISCUSSION

3.1 A library of SGPB primary specificity variants

Primary specificity (a preference for substrate peptide bonds immediately following particular amino acids) depends on a protease's S1 pocket, the region which binds the substrate P1 side-chain. Crystallographic studies of SGPB have revealed two segments of polypeptide (residues 192-192b and residues 213-216) which are directly involved in forming the S1 pocket (Read et al., 1983) (Fig. 3-1). Residues 192 and 213 are of particular importance since the side-chains of these residues project into the pocket and, as a result, can interact directly with the P1 side chain. In *S. griseus* proteases A, B, C, and D, small side chains (Ala192 and Thr213) at these positions produce a large S1 pocket; these enzymes prefer large hydrophobic P1 side-chains (Bauer, 1978). In the α-lytic protease of *Lysobacter enzymogenes* (α-LP), Met192 and Met213 greatly reduce the size of the S1 pocket (Fujinaga et al., 1985); thus, α-LP prefers small P1 substrates. Finally, SGPE exhibits primary specificity for glutamate partly due to the presence of the polar residues Ser192 and His213. The crystal structure of SGPE (Nienaber et al., 1993) suggests two additional primary specificity determinants; His199 and His228 (replacing Leu199 and F228 in SGPB) interact with His213 to form a hydrogen bonded "His triad" which extends through the hydrophobic core of the protein and is putatively involved in stabilizing the negatively charged P1 glutamate.

Based on an alignment of the S2 serine protease family (Fig. 2-15) we selected seven residues as mutagenic targets (Fig. 3-1). The selected residues fulfilled two criteria: they were putative primary specificity determinants and they were variable within the family. Thus, we theorized that mutations at

Fig. 3-1. The S1 pocket of SGPB with the P1 Leu residue of bound turkey ovomucoid inhibitor third domain (Read et al., 1983). Atomic coordinates were obtained from the Brookhaven Protein Data Bank (entry set 3SGB) and represented using the program package Insight II version 2.3.0 (Biosym Technologies, San Diego, CA). SGPB residues selected for mutagenesis and the P1 Leu residue (LeuP1) are labeled. Backbone atoms are depicted in light grey while side-chain atoms are in colour (carbon, green; oxygen, red). Interatomic distances are in angstrom units.


LeuP1
Pro192b
Gly216
Glu192a
4.19
3.60
Ala192
Thr213
3.19
Phe228
Leu199

these sites may alter primary specificity without perturbing the overall structural and catalytic integrity of the enzyme. Specifically, degeneracies were introduced at positions 192, 192a, 192b, 199, 213, 216, and 228 by means of a DNA cassette encoding from residue 175 to the carboxyl terminus. The cassette could be inserted into the *sprB* gene by means of an unique *Aat*II restriction site and a restriction site following the termination codon. In addition, a silent mutation in the codon for Gly197 introduced a *Rsr*II restriction site which could be used to distinguish between mutant and wild-type *sprB* genes. The library was designed such that, as a minimum, each variable residue would be represented by either the wild-type SGPB residue or the corresponding residue in the sequence of *S. griseus* protease A, C, D, or E or α-LP; potentially, the library contained 41,472 nucleotide combinations encoding 29,952 unique protein sequences (Table 3-1).

The mutagenic DNA cassette was synthesized using a method which did not involve any mismatched base pairing, thus eliminating a major source of bias in the library. Two large oligonucleotides were synthesized; the sequence of SB1 corresponded to a segment of *sprB* encoding from residue 175 to residue 211 of mature SGPB while that of SB2 was complementary to a segment of *sprB* encoding from residue 202 to the carboxyl terminus. The 3' termini of SB1 and SB2 were perfectly complementary so that annealing and extension with polymerase allowed for the synthesis of the mutagenic cassette. The cassette was further amplified through a PCR using oligonucleotides SB3 and SB4 as primers. A high template concentration (approximately 10^{11} copies per PCR) was used and the products of 60 separate PCRs were combined to minimize the introduction of a bias during amplification. Twenty individual clones were sequenced; all the designed degeneracies were present and no other mutations were observed at the

Table 3-1. Sequence alignment of variable residues in the *sprB* library.

Enzyme[a]	pmP1	192	192a	192b	199	213	216	228
SGPB	L	A	E	P	L	T	G	F
SGPA	L	A	E	P	L	T	G	Y
SGPC	L	A	E	P	L	T	G	F
SGPD	L	A	E	P	L	T	G	F
SGPE	E	S	A	G	H	H	S	H
α-LP	T	M	G	R	L	M	G	F
SGPB LIBRARY		A **L** **M** **S** **T** **V**	**A** E **G**	**A** **G** P **R**	**H** L	**A** **D** **E** **H** **I** **K** **L** **M** **N** **P** **Q** T **V**	**A** G **S** **T**	F **H** **L** **Y**

[a]Sequences for the known *S. griseus* members of the S2 serine protease family and the α-lytic protease (α-LP) are shown along with the possible variations at each position within the *sprB* expression library. The pmP1 residue is also shown for each enzyme. Mutations are in *bold* text.

degenerate sites. Furthermore, no obvious bias was observed at any of the mutagenized positions.

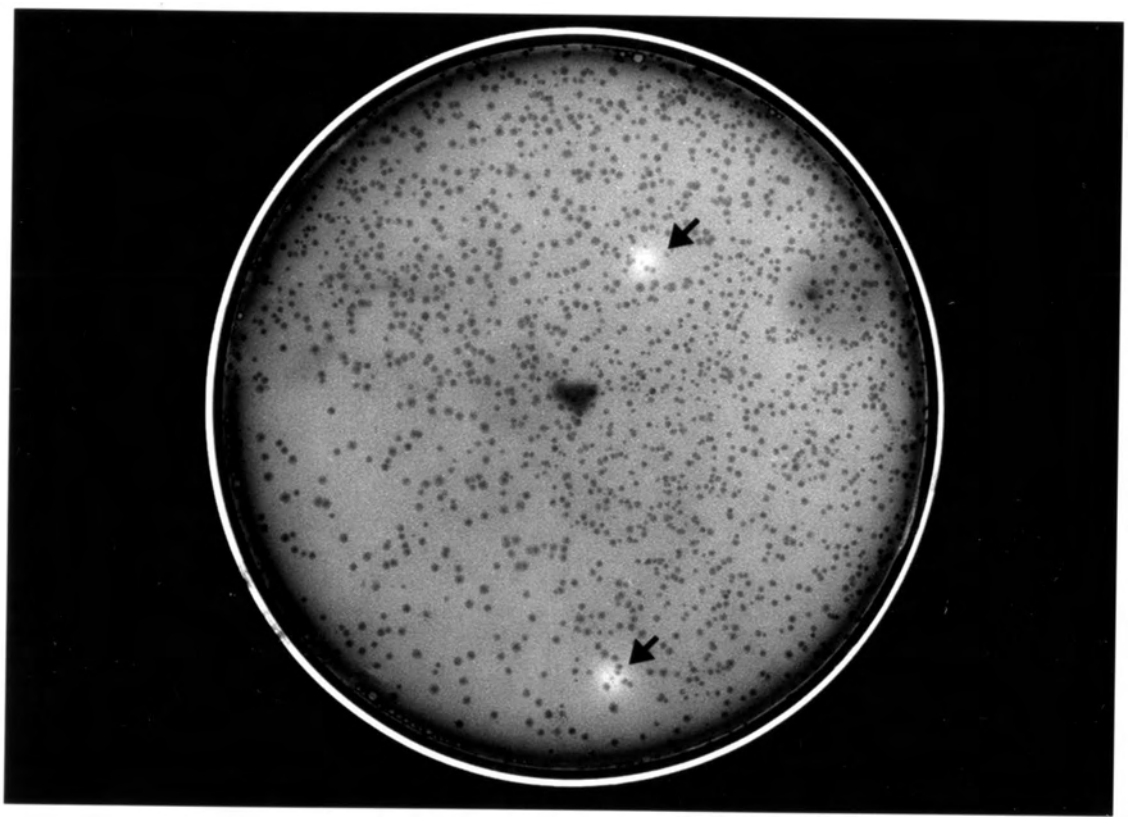
3.2 Visualization of proteolytic activity

SGPB is capable of degrading skim milk proteins, and active SGPB secreted from *E. coli* accumulates in the extracellular media. Thus, *E. coli* transformants secreting active SGPB produce "zones of clearing" when grown on plates containing skim milk. An *E. coli* colony secreting active protease can be visually detected against a background of non-secreting colonies, thereby providing a rapid screen for catalytically active mutant proteases (Fig. 3-2). Since zones of clearing are produced by proteases with primary specificity towards large hydrophobic residues (*S. griseus* proteases A, B, C, and D), small residues (α-LP), acidic residues (SGPE), and basic residues (*S. griseus* trypsin), the detection of proteolytic activity is not dependent on a particular specificity.

3.3 Substitutions at the promature junction P1 site (pmP1)

S. griseus proteases A, B, C, D, and E, like all members of the α-lytic protease family, are characterized by the presence of a large amino-terminal propeptide which is necessary for correct folding but is not present in the mature enzyme. Activation requires a cleavage at the promature junction, and two lines of evidence suggest that the enzymes are self-processing. Firstly, the P1 residue at each enzyme's promature junction correlates with the primary specificity of the enzyme; interestingly, all enzymes with primary specificity for large hydrophobic substrates (*S. griseus proteases* A, B, C, and D)

Fig. 3-2. *E. coli* TB1 colonies grown on a YT/milk plate. A typical YT/milk plate (100 mm x 15 mm) containing 1.9 x 10^3 individual *E. coli* transformants from the expression library with wild-type Leu at pmP1. Colonies secreting active protease (arrows) produce zones of clearing which are easily visible against a background of inactive variants. The libraries were screened at a density of 1.0 x 10^3 cfu/plate, allowing for the discernation and isolation of individual clones of interest.



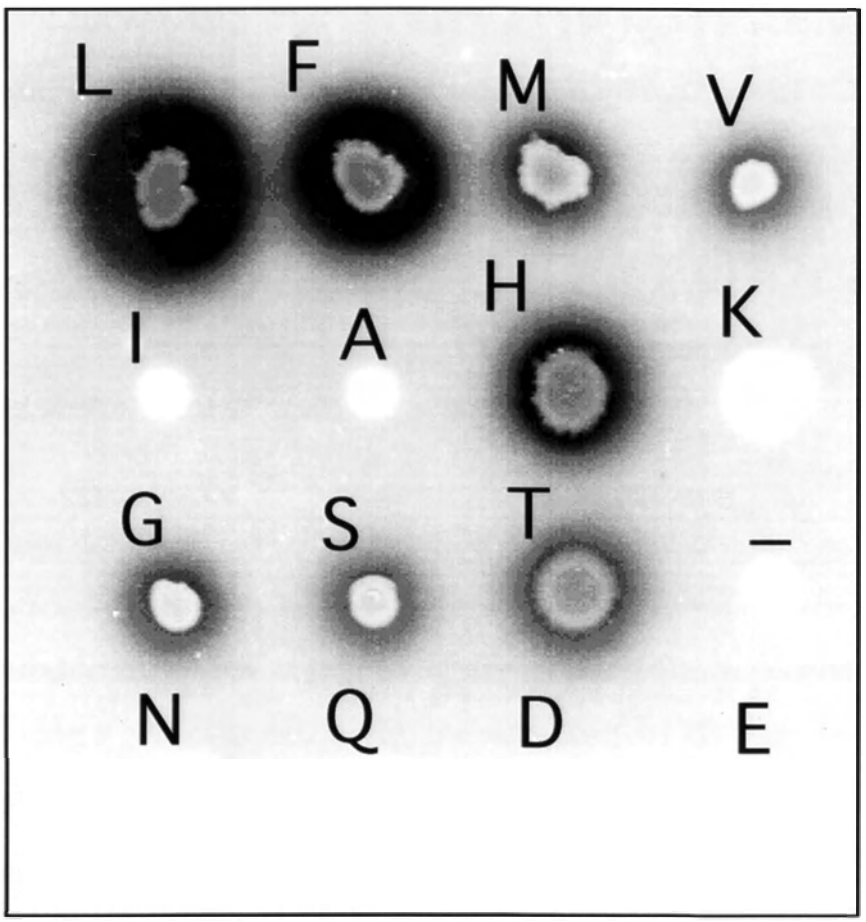
95b

have a Leu at this position. Secondly, recombinant expression of catalytically active proteases in *E. coli* (Silen et al., 1989) and *B. subtilis* (Chapter 2, section 3.3) results in the production of correctly processed enzymes, while catalytically inactive mutants of α-LP (Silen et al., 1989) expressed in *E. coli* accumulate as unprocessed proenzymes. In a recombinant host, therefore, the catalytic activity of recombinant protease is necessary and sufficient for proenzyme activation. We constructed the vector pMP-B (a derivative of the IPTG inducible secretion vector pMAL-p (Guan et al., 1987; Maina et al., 1988)) to secrete promature SGPB in *E. coli* and the plasmids pMP-BX (where X = A, D, E, F, G, H, I, K, M, N, Q, S, T, or V) to secrete mutant promature SGPB polypeptides with substitutions at the promature junction P1 site (pmP1); the effects of these substitutions on *sprB* expression levels are shown in Fig. 3-3. The results are discussed below and implications for the folding mechanism of promature SGPB are discussed in section 3.7.

3.3.1 Large hydrophobic residues at pmP1

As expected, the substitution of Phe for Leu at pmP1 had no appreciable effect on expression levels, since Phe is a better P1 substrate than Leu (Bauer, 1978). On the other hand, the substitution of Leu by the poor substrate Val greatly reduced activity, while substitution with Ile (an even poorer substrate) abolished activity. Interestingly, while Met seems to be as good a P1 substrate as Leu in a synthetic substrate (Table 3-4), transformants harbouring pMP-BM produced greatly reduced levels of activity relative to pMP-B. There are at least two possible explanations for this result: either the introduction of Met at the promature junction P1 site sterically hinders the folding of SGPB, or constraints imposed by the promature junction render Met a much poorer

Fig. 3-3. ***E. coli*** **transformants secreting promature SGPB polypeptides with mutations at pmP1.** The residue at pmP1 is labeled (single-letter amino acid code) in each case; the colony labeled "-" is a negative control harbouring the plasmid pMal-p. The transformant secreting promature SGPB with wild-type Leu at pmP1 harboured the plasmid pMP-B while all others harboured plasmids of the form pMP-BX (where "X" is the label above each colony). Transformants were incubated on a YT/milk plate for two days at 20 °C and then for five days at 4 °C. The diameter of the "zone of clearing" around each colony is proportional to the amount of active enzyme secreted.


L
F
M
V
I
A
H
K
G
S
T
–
N
Q
D
E

intramolecular substrate than anticipated on the basis of *in vitro* studies. Our kinetic analyses support the latter explanation (section 3.4.1.3).

3.3.2 Small residues at pmP1

The introduction of Thr, Ser, or Gly at pmP1 reduced activity to levels comparable to those obtained with Met and Val. The result is somewhat surprising since Gly, Ser and Thr are expected to be relatively poor substrates for the chymotrypsin-like specificity of SGPB (see section 3.7 for further discussion). The introduction of Ala at pmP1 abolished activity; this result is not readily explicable since it would be expected that Ala (a better P1 substrate than Gly, Ser, or Thr) should allow expression levels at least comparable to those obtained with the other small residues at pmP1.

3.3.3 Large polar or charged residues at pmP1

The introduction of His at pmP1 slightly reduced active protease secretion. Having a pKa near neutrality, the side-chain of His may be either neutral or positively charged in a protein environment; in its uncharged state, the bulky His side-chain should be a good P1 substrate for SGPB. the ntroduction of Asn, Gln, Asp, Glu, or Lys at pmP1 completely abolished ctivity; the results are reasonable, since these large, charged or polar side-hains should be poor substrates for SGPB.

3.4 Selection of active SGPB variants with pmP1 mutations

Since the presence of a poor substrate at pmP1 reduces proteolytic activity, perhaps activity could be regained through complementary changes in the S1 pocket of the enzyme. Such a relationship could allow for the selection of mutant proteases with substrate specificities shifted in favour of the pmP1 residue (Fig. 3-4).

The mutant specificity library described above (section 3.1) was screened with mutations at pmP1 for variants capable of increased levels of proteolytic activity relative to wild type SGPB (as evidenced by clearing on YT/milk plates, Fig. 3-2). The number of individual clones screened in each selection and the results of each screen are summarized in Table 3-2. Plasmid DNA isolated from potential positives was used to transform *E. coli* TB1 to ensure that the ability to produce zones of clearing was a plasmid borne trait. Plasmid DNA from *E. coli* TB1 transformants with reproducible zones of clearing was purified and the nucleotide sequence of the mutations was determined. The entire insert region was sequenced for each mutant, and no mutations were observed outside the degenerate positions. For kinetic analysis, proteases were produced in *B. subtilis* using the secretion-expression vector pEB11 (chapter 2, section 3.3.1).

3.4.1 Variants with large hydrophobic residues at pmP1

Screens with Leu, Phe, Met, or Val at pmP1 all resulted in the isolation of variants exhibiting increased proteolytic activity in *E. coli*. Depending on the nature of the promature junction, proteolytic activity was increased through three different effects: increased thermostability, increased specificity

Fig. 3-4. The selection strategy. In each panel, an *E. coli* colony secreting a particular promature protease is shown on a YT/milk plate (top) along with a schematic representation of the promature enzyme (bottom). The propeptide is filled black while the mature enzyme is unfilled. The promature junction P1 residue (pmP1) and the S1 binding pocket (S1) are labeled. A, Secretion of wild-type promature SGPB from plasmid pMP-B. The Leu residue at pmP1 is a good substrate for the chymotrypsin-like primary specificity of SGPB. Efficient folding and junction processing results in high proteolytic acitivity which is detected as a large "zone of clearing" around the *E. coli* colony (due to hydrolysis of milk proteins). B, Secretion of promature SGPB with Ser at pmP1 from plasmid pMP-BS. Ser is a poor substrate for SGPB; reduction in proteolytic activity results in a greatly reduced zone of clearing. C, Secretion of promature SGPB with Ser at pmP1 and a mutated S1 pocket (plasmid pMP-BSr2 secreting SGPBSr2, see Table 3-5). Mutations in the S1 pocket complementary to the mutation at pmP1 restore efficient promature junction processing. The result is an increase in proteolytic acitivity which is again visualized as a large zone of clearing.

A
S1
pmP1
B
S1
pmP1
C
S1
pmP1

Table 3-2. Summary of expression library screens with variations at pmP1.

		cfu with zones of clearing[b]		
pmp1 residue	cfu screened[a]	wild-type	mutant	total
A	1.58×10^5	0	0	0
D	1.16×10^5	0	0	0
E	1.07×10^5	0	0	0
F	7.6×10^4	3	4	7
G	5.1×10^4	5	4	9
H	1.22×10^5	0	2	2
K	9.5×10^4	0	0	0
L	2.25×10^5	48	7	55
M	5.9×10^4	0	2	2
N	2.0×10^5	0	0	0
Q	6.8×10^4	0	0	0
S	4.7×10^4	13	3	16
T	1.2×10^5	9	5	14
V	9.0×10^4	0	1	1

[a]Total number of colony forming units (cfu) screened for each promature junction P1 site (pmP1) variant. [b]Number of unique cfu harbouring a plasmid capable of conferring "zones of clearing" to *E. coli* transformants. Mutant plasmids were differentiated from wild-type contaminants on the basis of restriction analyses. The insert region of each mutant plasmid was sequenced. Procedures are described in section 2.8.2.

towards the promature junction P1 residue, or increased non-specific catalytic efficiency.

3.4.1.1 Variants with Leu at pmP1

A screen of the specificity library in conjunction with wild-type Leu at pmP1 resulted in the isolation of seven unique SGPB mutants (Table 3-3) producing proteolytic activity comparable to that of SGPB in *E. coli* ; mutants producing increased proteolytic activity were not obtained. The sequence at positions 192a and 213 divides the mutants into two groups, those with wild-type amino acids at these positions (E192a/T213) and those with mutations at these positions (E192aAorG/T213L). In no case is a mutation at one of these sites accompanied by a wild-type residue at the other. Furthermore, T213L is the only mutation observed at position 213, despite the possibility of 13 different residues within the random library (Table 3-1).

Four mutant proteases (SGPBr1, SGPBr3, SGPBr6, and SGPBr8) were kinetically characterized (Table 3-4, Figs. 3-5 and 3-6). Primary specificity was essentially conserved among the mutants, but single mutant SGPBr1 (F228H) did exhibit a three-fold increase in k_{cat}/K_m against LeuP1. While k_{cat}/K_m against LeuP1 was reduced for the two variants with the mutation T213L (SGPBr3 and SGPBr6), it was increased for the proteases containing wild-type Thr213 (SGPBr1 and SGPBr8).

3.4.1.2 Variants with Phe at pmP1

Phe is an excellent P1 substrate for SGPB (Table 3-4) and expression levels from pMP-B and pMP-BF were essentially identical (Fig. 3-3); therefore,

Table 3-3. SGPB mutants isolated with Leu, Phe, Met, or Val at pmP1.

Enzyme[a]	pmP1	192	192a	192b	199	213	216	228
SGPB LIBRARY		A L M S T V	A E G	A G P R	H L	A D E H I K L M N P Q T V	A G S T	F H L Y
SGPB	L	A	E	P	L	T	G	F
SGPBr1	L	A	E	P	L	T	G	**H**
SGPBr3	L	A	**G**	P	L	**L**	G	**Y**
SGPBr5	L	A	**A**	**R**	L	**L**	G	**Y**
SGPBr6	L	A	**A**	**A**	L	**L**	G	**F**
SGPBr8	L	A	E	**A**	L	**T**	G	**F**
SGPBr11	L	A	**G**	**R**	L	**L**	G	**Y**
SGPBr14	L	A	E	**G**	L	T	G	F
SGPBFr1	**F**	A	E	**G**	L	**L**	G	F
SGPBFr12	**F**	A	E	**G**	L	**L**	G	**Y**
SGPBFr13	**F**	A	E	P	L	**L**	G	**H**
SGPBFr14	**F**	A	E	P	L	**A**	G	**Y**
SGPBMr2	**M**	A	E	P	L	**L**	G	**F**
SGPBMr3	**M**	A	E	**G**	L	**L**	G	**Y**
SGPBVr2	**V**	A	E	**R**	L	**L**	G	**Y**

[a]The possible variations at each position within the expression library are shown along with the sequences of SGPB and the SGPB mutants. The pmP1 residue is also shown for each enzyme. Mutations are in *bold* text.

Table 3-4. Kinetic constants and associated errors for SGPB and variants selected with Leu, Phe, Met, or Val at pmP1.

Enzyme	Substrate[a]	k_{cat} (s^{-1})[b]	K_m (μM)[b]	k_{cat}/K_m ($s^{-1}M^{-1}$)[b]
SGPB	Phe	31.4 ± 1.7	62:6 ± 1.5	501,000 ± 39,000
	Leu	6.36 ± 0.21	124.5 ± 4.8	51,100 ± 3,600
	Met	16.15 ± 0.74	306 ± 12	52,600 ± 4,500
	Val	0.292 ± 0.015	555 ± 25	526 ± 50
SGPBr1	Phe	31.9 ± 1.8	45.32 ± 0.73	703,000 ± 52,000
	Leu	6.51 ± 0.22	42.9 ± 1.6	152,000 ± 11,000
	Met	13.65 ± 0.49	80.8 ± 2.2	169,000 ± 11,000
	Val	0.318 ± 0.014	200.1 ± 7.2	1,590 ± 130
SGPBr3	Phe	15.63 ± 0.91	37.4 ± 1.3	418,000 ± 39,000
	Leu	2.01 ±0.12	116.6 ± 4.4	17,200 ± 1,700
	Met	11.44 ± 0.66	71.4 ± 1.7	160,000 ± 13,000
SGPBr6	Phe	22.64 ± 0.87	39.6 ± 1.2	572,000 ± 39,000
	Leu	1.76 ± 0.06	64.1 ± 1.7	27,400 ± 1,700
	Met	10.00 ± 0.36	46.2 ± 1.0	216,000 ± 13,000
SGPBr8	Phe	26.9 ± 1.3	64.6 ± 1.7	416,000 ± 31,000
	Leu	6.18 ± 0.28	109.6 ± 4.0	56,400 ± 4,600
	Met	8.70 ± 0.39	200.9 ± 4.1	43,300 ± 2,800
	Val	0.124 ± 0.0057	517.2 ± 9.7	240 ± 16
SGPBFr1	Phe	14.93 ± 0.45	54.5 ± 1.3	274,000 ± 14,000
	Leu	2.09 ± 0.07	86.1 ± 3.3	24,300 ± 1,700
	Met	9.43 ± 0.37	133.4 ± 4.6	70,700 ± 5,200
SGPBFr12	Phe	24.3 ± 1.0	45.2 ± 1.4	538,000 ± 39,000
	Leu	1.88 ± 0.07	34.1 ± 1.0	55,100 ± 3,600
	Met	16.82 ± 0.70	36.9 ± 1.5	456,000 ± 37,000
	Val	0.069 ± 0.022	7,700 ± 2,400	**7.90 ± 0.24**
SGPBFr13	Phe	13.26 ± 0.49	58.9 ± 1.9	225,000 ± 15,000
	Leu	2.38 ± 0.09	107.4 ± 3.9	22,200 ± 1,600
	Met	11.25 ± 0.54	146.8 ± 6.4	76,700 ± 7,000

continued

Table 3-4. *continued*

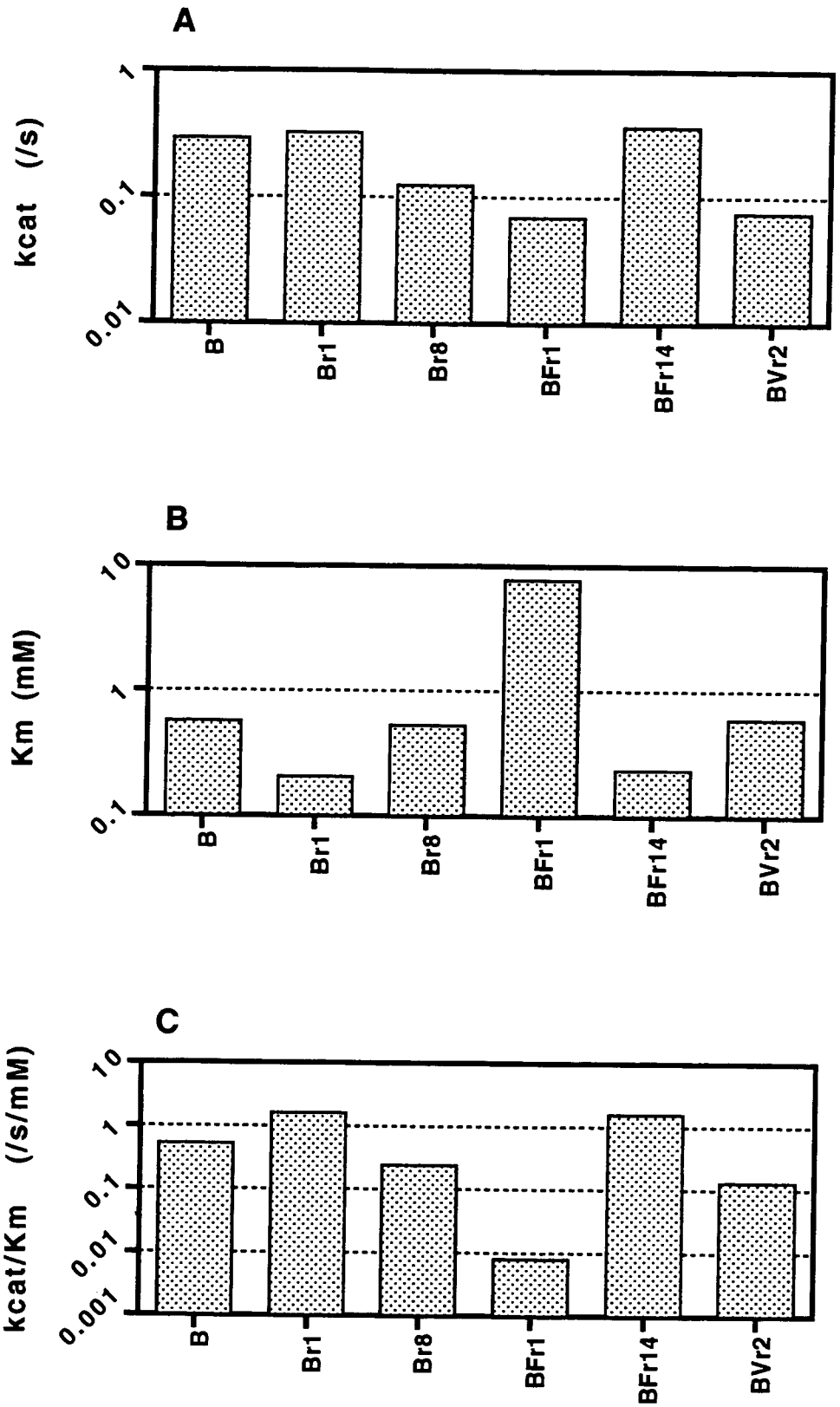
Enzyme	Substrate[a]	kcat (s^{-1})[b]	Km (μM)[b]	kcat/Km ($M^{-1}s^{-1}$)[b]
SGPBFr14	Phe	18.85 ± 0.85	139.3 ± 3.5	135,000 ± 9,500
	Leu	3.30 ± 0.13	38.6 ± 1.2	85,500 ± 6,000
	Met	7.30 ± 0.28	168.4 ± 1.9	43,400 ± 2,200
	Val	0.362 ± 0.013	230.0 ± 1.2	1,570 ± 63
SGPBMr2	Phe	20.64 ± 0.66	57.1 ± 1.3	362,000 ± 20,000
	Leu	2.47 ± 0.15	37.6 ± 1.4	65,700 ± 6,500
	Met	10.67 ± 0.55	24.03 ± 0.44	444,000 ± 31,000
SGPBVr2	Phe	12.91 ± 0.84	15.57 ± 0.49	829,000 ± 80,000
	Leu	2.80 ± 0.14	8.74 ± 0.19	320,000 ± 23,000
	Met	19.4 ± 1.3	24.25 ± 0.79	799,000 ± 78,000
	Val	.075 ± .0051	594 ± 19	126 ± 12

[a]Substrates were of the form Suc-AAPX-pNA (where X = Phe, Leu, Met, or Val as indicated). [b]Standard errors are shown to two significant figures and constants are shown to the corresponding accuracy. Constants in *bold* were determined from a linear regression analysis, all others were determined from a non-linear regression analysis of the Michealis-Menten equation. Procedures are described in section 2.11.

Fig. 3-5. Kinetic constants for SGPB and variants selected with Leu, Phe, Met, or Val at pmP1. Substrates were of the form Suc-AAPX-pNA (where x = Phe, Leu, or Met as indicated). Data from Table 3-4; the enzyme name is indicated along the horizontal axis (*e.g.* "B" designates SGPB). A, k_{cat} values (in units of s^{-1}). B, K_m values (in units of μM). C, k_{cat}/K_m values (in units of $s^{-1}\mu M^{-1}$).

A
kcat (/s)
0
10
20
30
40
B
Br1
Br3
Br6
Br8
BFr1
BFr12
BFr13
BFr14
BMr2
BVr2
B
Km (uM)
0
100
200
300
400
C
kcat/Km (/s/uM)
0
0.25
0.5
0.75
1
Phe
Leu
Met

Fig. 3-6. Kinetic constants against Suc-AAPV-pNA for SGPB and variants selected with Leu, Phe, or Val at pmP1. Data from Table 3-4. The constants are plotted on logarithmic scales; the enzyme name is indicated along the horizontal axis (e.g. "B" designates SGPB). A, k_{cat} values (in units of s^{-1}). B, K_m values (in units of mM). C, k_{cat}/K_m values (in units of $s^{-1}mM^{-1}$).

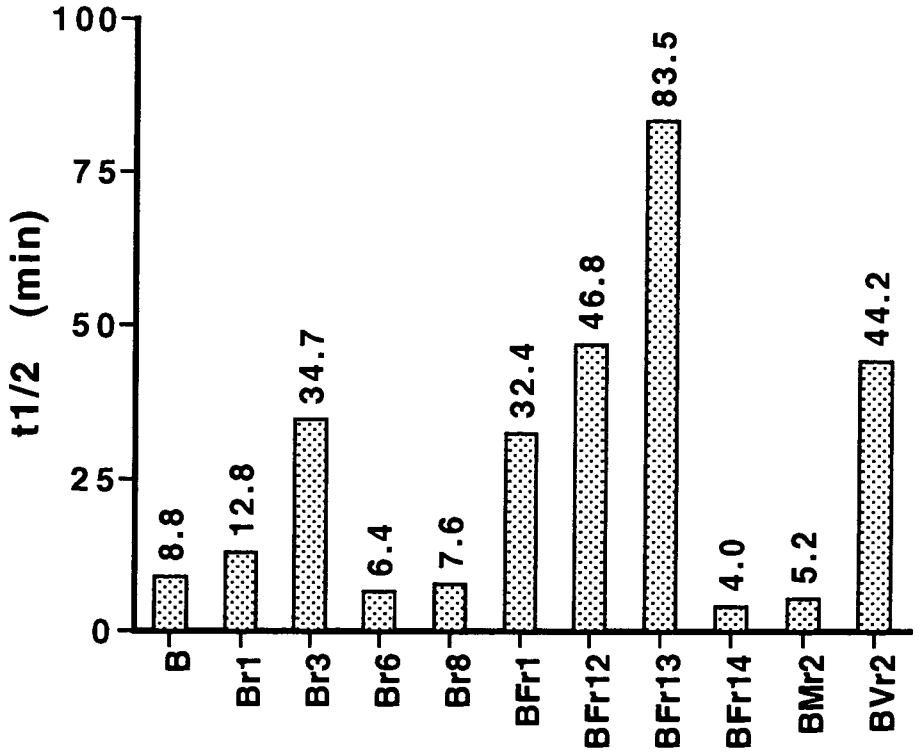

A
kcat (/s)
1
0.1
0.01
B
Br1
Br8
BFr1
BFr14
BVr2
B
Km (mM)
10
1
0.1
B
Br1
Br8
BFr1
BFr14
BVr2
C
kcat/Km (/s/mM)
10
1
0.1
0.01
0.001
B
Br1
Br8
BFr1
BFr14
BVr2

the isolation of variants with increased proteolytic activity from the pMP-BF screen was not anticipated. Nevertheless, four unique variants producing significantly increased zones of clearing (in comparison with both pMP-B and pMP-BF) were isolated (Table 3-3). None of the variants exhibited increased k_{cat}/K_m values towards PheP1 (Table 3-4, Fig. 3-5), suggesting that the primary specificity of wild-type SGPB towards Phe may already be near optimal and thus difficult to improve. However, three of the variants (SGPBFr1, SGPBFr12, and SGPBFr13) had greatly increased half-lives at 55 ^{o}C in comparison with SGPB (Fig. 3-7); in addition to other mutations, these variants share a common T213L mutation. SGPBFr14 had a reduced half-life which may correlate with this variant being the only one with the substitution T213A. It appears that while SGPB is optimized for primary specificity towards Phe, proteolytic activity from constructs with Phe at the promature junction P1 site can be increased by increasing the thermostability of the mature enzyme.

3.4.1.3 Variants with Met at pmP1

Two unique variants producing increased zones of clearing (SGPBMr2 and SGPBMr3) were isolated from the screen with Met at pmP1 (table 3-3); the sequence of SGPBMr3 was identical to that of SGPBFr12. Relative to SGPB, Both SGPBMr2 and SGPBFr12 exhibited ten-fold increases in k_{cat}/K_m against MetP1 with minimal changes in k_{cat}/K_m values against PheP1 and LeuP1 (Table 3-4, Fig. 3-5). In fact, the k_{cat}/K_m values for SGPBMr2 and SGPBFr12 against MetP1 were almost identical to that of SGPB against PheP1; therefore, the two mutants have primary specificities towards Met which are comparable with that of SGPB towards Phe, the optimal P1 substrate.

Fig. 3-7. Thermostabilities of SGPB and SGPB mutants at 55 °C. All enzymes exhibited linear exponential decays with time, allowing for the determination of half-lives ($t_{1/2}$, in units of min) from the slopes of the lines. Numerical values are noted above each bar.


t1/2 (min)
100
75
50
25
0
8.8
12.8
34.7
6.4
7.6
32.4
46.8
83.5
4.0
5.2
44.2
B
Br1
Br3
Br6
Br8
BFr1
BFr12
BFr13
BFr14
BMr2
BVr2

Interestingly, in both cases the increase in k_{cat}/K_m is due entirely to a decrease in K_m; SGPBMr2 and SGPBFr12 bind Met more efficiently but k_{cat} is not increased. SGPBMr2 differs from SGPB only by the substitution T213L and, since the side-chain of residue 213 protrudes into the S1 pocket (Fig. 3-1), it is possible that the replacement of Thr213 by a bulkier Leu side-chain narrows the pocket and provides a better fit for a long, unbranched MetP1 side-chain. However, the explanation is complicated by the fact that not all variants with the mutation T213L exhibit increased primary specificity for Met; mutations at different sites have cooperative effects. For example, in going from the single mutant SGPMr2 (T213L) to the double mutant SGPBFr1 (T213L/P192bG) and finally to the triple mutant SGPBFr12 (T213L/P192bG/F228Y), k_{cat}/K_m towards MetP1 goes from 0.44 $s^{-1}\mu M^{-1}$ down to 0.070 $s^{-1}\mu M^{-1}$ and then back up to 0.46 $s^{-1}\mu M^{-1}$ (Table 3-4). The results suggest that the decrease in proteolytic activity due to the substitution of Met for Leu at pmP1 is caused by a decrease in the enzyme's specificity for the processing site; mutations which increase the primary specificity of SGPB towards Met compensate for this effect and allow for increased protease production.

3.4.1.4 Variants with Val at pmP1

Due to a hundred-fold reduction in k_{cat} and a ten-fold increase in K_m, wild type SGPB exhibits a k_{cat}/K_m towards ValP1 which is three orders of magnitude lower than that towards PheP1 (Table 3-4); thus, it is not surprising that the introduction of Val at pmP1 results in the reduction of proteolytic activity (Fig. 3-3). The screen with Val at pmP1 produced one isolate with restoration of proteolytic activity (SGPBVr2, Table 3-3). Against

substrates with Phe, Leu, and Met at the P1 position, SGPBVr2 exhibits enhanced activity which is due entirely to a reduction in K_m (Table 3-4, Fig. 3-5); the reduction in K_m is probably caused by interactions with the substrate backbone since the effect is not restricted to a particular side-chain. Interestingly, SGPBVr2 did not exhibit increased primary specificity towards ValP1 and, in fact, the enzymes with the highest k_{cat}/K_m values against ValP1 all contain wild type Thr213 (SGPB, SGPBr1, and SGPBr8) or the mutation T213A (SGPBFr14) (Table 3-4, Fig. 3-6). With the exception of SGPBVr2, all enzymes with the substitution T213L exhibited greatly reduced k_{cat}/K_m values against ValP1 (less than 3% wild-type specific activity against 200 μM substrate in 50 mM Tris, 5% methanol; kinetic constants were determined only for SGPBFr1; Table 3-4, Fig. 3-6). It is possible that SGPBVr2 exhibits close to wild-type activity against ValP1 due to two compensating effects: a decrease in primary specificity towards Val caused by the mutation T213L and a general increase in primary specificity towards all substrates. Primary specificity towards Val is favoured by small side-chains (Thr or Ala) and disfavoured by Leu at position 213; the same effect was observed for primary specificity towards Leu and may be generally true for branched side-chain substrates. It is possible that the library used in this study does not contain a member with greatly enhanced primary specificity towards Val.

3.4.2 Variants with small residues at pmP1

Screens with Gly, Ser, or Thr at pmP1 all resulted in the isolation of variants producing increased proteolytic activity. In addition to variants with mutations similar to those isolated from the screens with large hydrophobic pmP1 residues, several new variants were also obtained. The screen with Ala

at pmP1 did not produce variant enzymes exhibiting zones of clearing in *E. coli*. This is not surprising in light of the fact that Ala at pmP1 completely abolishes wild-type SGPB production while Gly, Ser, or Thr at pmP1 only reduce SGPB production (Fig. 3-3); however, the difference in effect due to Ala as opposed to the other small side-chains is not clear.

3.4.2.1 Variants with Ser or Thr at pmP1

Three variants were isolated from the screen with Ser at pmP1; five variants were isolated from the screen with Thr at pmP1 (Table 3-5). All variants contained the substitution T213L, but they could be divided into two groups on the basis of the residue at position 192. Group 1 contained wild-type Ala192; the combination A192/T213L had previously been observed in screens with large hydrophobic residues at pmP1 (section 3.4.1). Group 2 contained the substitution A192S; the combination A192S/T213L should reduce the volume and hydrophobicity of the S1 pocket and thus complement the change at pmP1.

SGPB, a representative from Group 1 (SGPBFr1, identical to SGPBTr3) and a representative from Group 2 (SGPBTr4) were kinetically characterized against a panel of six synthetic substrates varying solely in the P1 position (Table 3-6, Fig. 3-8). The primary specificity profile of SGPBFr1 was similar to that of SGPB; both enzymes prefer large hydrophobic substrates. The primary specificity of SGPBTr4 relative to SGPB (as measured by the ratio of k_{cat}/K_m, Fig. 3-9) was reduced against all substrates except SerP1. The primary specificity profile of SGPB is as follows: Phe>Met, Leu>Ala>Val, Ser. SGPBTr4 exhibits drastic decreases in k_{cat}/K_m against PheP1, LeuP1, and

Table 3-5. SGPB mutants isolated with Ser, Thr, Gly, or His at pmP1.

Enzyme[a]	pmP1	192	192a	192b	199	213	216	228
SGPB LIBRARY		A L M S T V	A E G	A G P R	H L	A D E H I K L M N P Q T V	A G S T	F H L Y
SGPB	L	A	E	P	L	T	G	F
SGPBSr2	S	S	E	G	L	L	G	Y
SGPBSr3	S	A	G	P	L	L	G	F
SGPBSr8	S	A	E	R	L	L	G	Y
SGPBTr1	T	A	G	A	L	L	G	Y
SGPBTr3	T	A	E	G	L	L	G	F
SGPBTr4	T	S	E	A	L	L	G	Y
SGPBTr6	T	S	E	G	L	L	G	Y
SGPBTr7	T	A	E	R	L	L	G	F
SGPBGr2	G	S	E	P	L	L	G	Y
SGPBGr4	G	A	E	R	L	L	G	Y
SGPBGr8	G	S	E	G	L	L	G	Y
SGPBGr12	G	A	E	R	L	H	G	Y
SGPBHr1	H	A	E	R	L	M	G	Y
SGPBHr2	H	A·	E	G	L	L	G	F

[a]The possible variations at each position within the expression library are shown along with the sequences of SGPB and the SGPB mutants. The pmP1 residue is also shown for each enzyme. Mutations are in *bold* text.

Table 3-6. Kinetic constants and associated errors for SGPB and variants selected with Ser, Thr, Gly, or His at pmP1.

Enzyme	Substrate[a]	k_{cat} (s^{-1})[b]	K_m (μM)[b]	k_{cat}/K_m ($M^{-1}s^{-1}$)[b]
SGPB	Phe	31.4 ± 1.7	62.6 ± 1.5	501,000 ± 39,000
	Leu	6.36 ± 0.21	124.5 ± 4.8	51,100 ± 3,600
	Met	16.15 ± 0.74	306 ± 12	52,600 ± 4,500
	Val	0.292 ± 0.015	555 ± 25	526 ± 50
	Ala	10.36 ± 0.37	3,020 ± 56	3,400 ± 180
	Ser	ND	ND	**510 ± 19**
SGPBFr1	Phe	14.93 ± 0.45	54.5 ± 1.3	274,000 ± 14,000
	Leu	2.09 ± 0.07	86.1 ± 3.3	24,300 ± 1,700
	Met	9.43 ± 0.37	133.4 ± 4.6	70,700 ± 5,200
	Val	0.069 ± 0.022	7,700 ± 2,400	7.90 ± 0.24
	Ala	ND	ND	**810 ± 28**
	Ser	ND	ND	**134.2 ± 7.4**
SGPBTr4	Phe	4.64 ± 0.37	5.12 ± 0.39	910 ± 140
	Leu	0.399 ± 0.043	3,740 ± 430	107 ± 24
	Met	11.98 ± 0.45	713 ± 28	16,800 ± 1,300
	Val	ND	ND	**1.290 ± 0.027**
	Ala	3.93 ± 0.65	14,000 ± 2,300	**264.0 ± 4.0**
	Ser	1.09 ± 0.18	940 ± 170	**950 ± 10**
SGPBGr12	Phe	1.88 ± 0.16	1,680 ± 150	1,120 ± 190
	Leu	0.227 ± 0.017	4,260 ± 290	**47.5 ± 1.2**
	Met	2.259 ± 0.083	614 ± 18	3,680 ± 240
SGPBHr1	Phe	7.88 ± 0.20	375 ± 12	21,000 ± 1,200
	Leu	0.744 ± 0.023	315 ± 15	2,360 ± 190
	Met	4.66 ± 0.21	182 ± 13	25,550 ± 300

[a]Substrates were of the form Suc-AAPX-pNA (where X = Phe, Leu, Met, Val, Ala, or Ser as indicated). [b]Standard errors are shown to two significant figures and constants are shown to the corresponding accuracy. "ND" denotes constants which could not be determined because the substrate concentrations were not high enough to allow a fit to the Michealis-Menten equation. Constants in *bold* were determined from a linear regression analysis, all others were determined from a non-linear regression analysis of the Michealis-Menten equation. Procedures are described in section 2.11.

Fig. 3-8. Kinetic constants for SGPB and variants selected with Ser or Thr at pmP1. Substrates were of the form Suc-AAPX-pNA (where X = Phe, Leu, Met, Val, Ala, or Ser as indicated). Data from Table 3-6. The constants are plotted on logarithmic scales; the enzyme name is indicated along the horizontal axis (*e.g.* "B" designates SGPB). A, k_{cat} values (in units of s^{-1}). B, K_m values (in units of mM). C, k_{cat}/K_m values (in units of $s^{-1}mM^{-1}$).

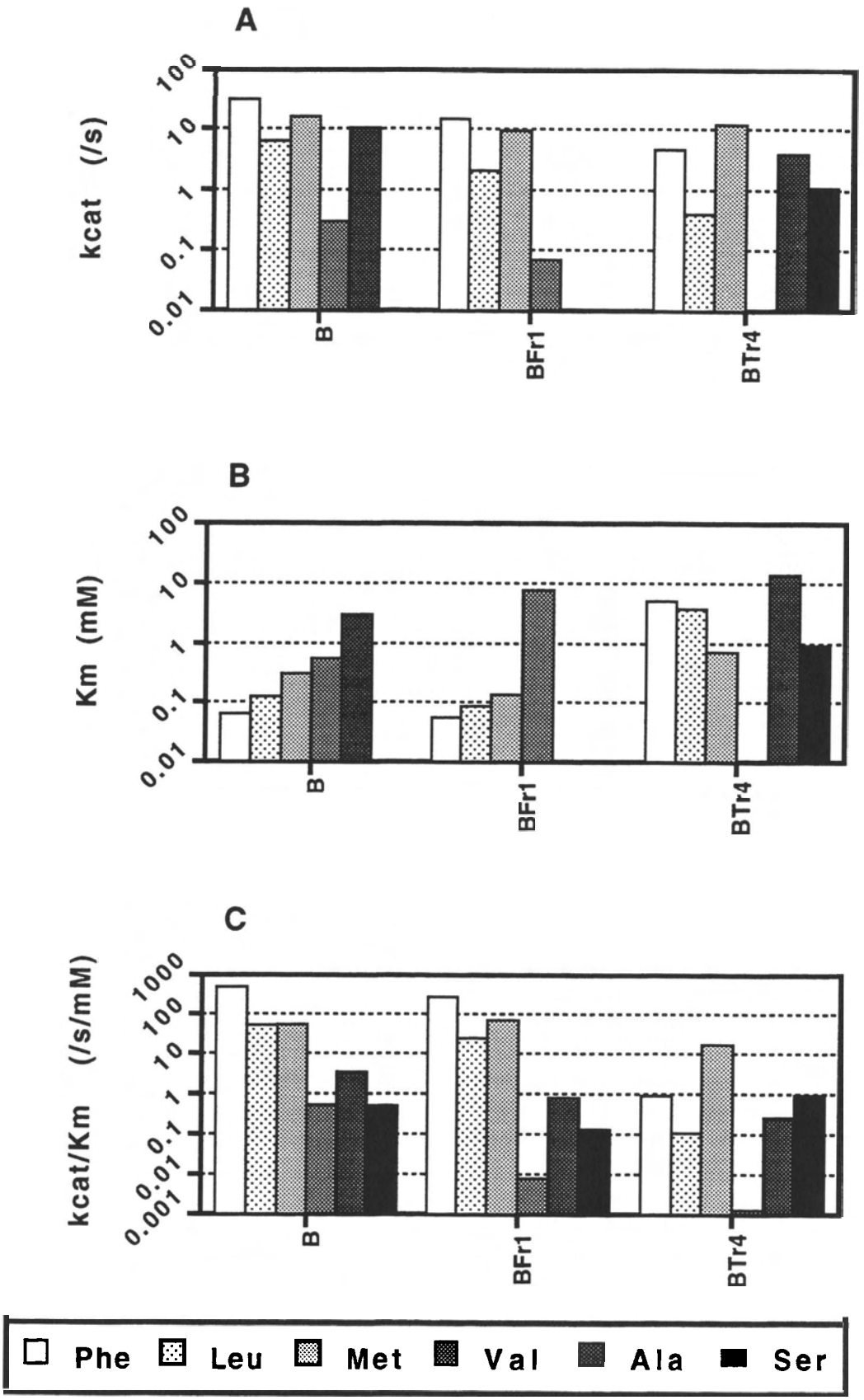
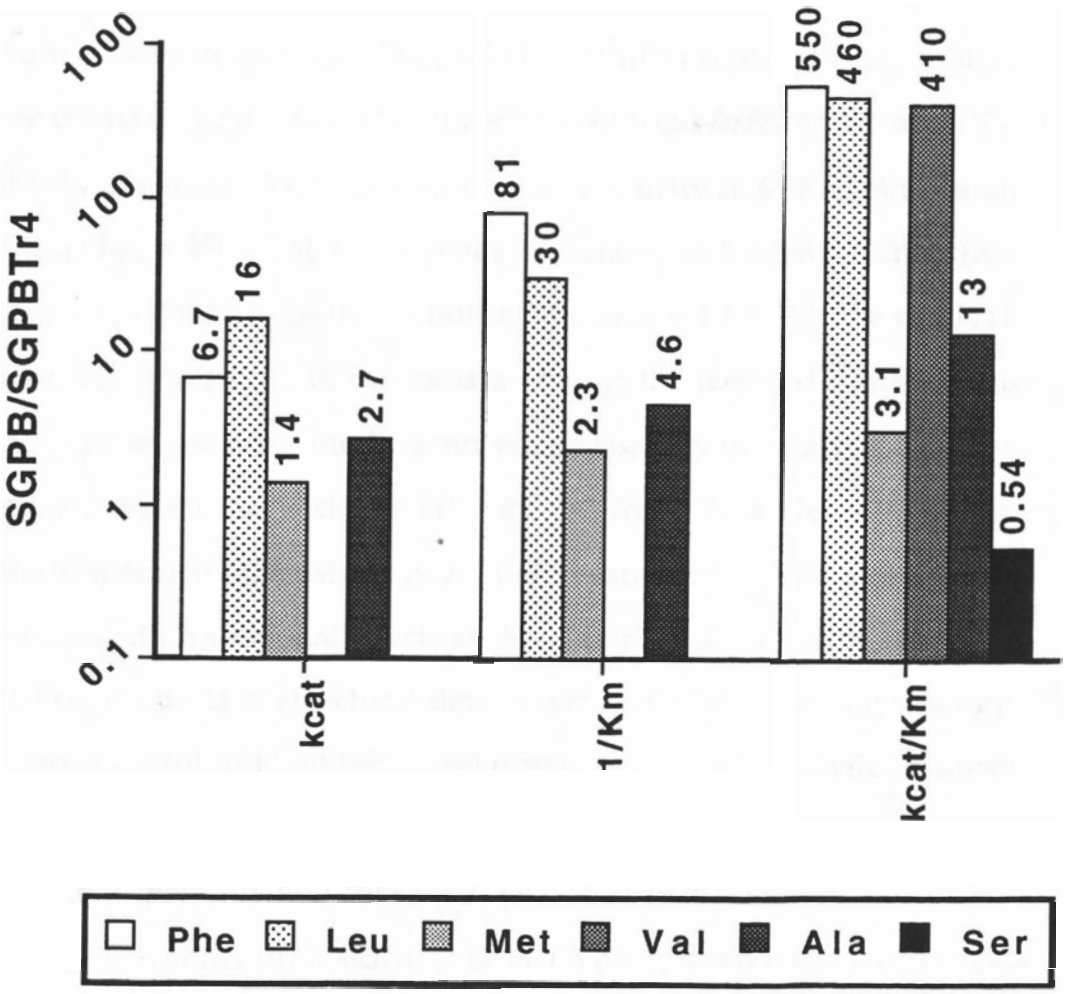

A
kcat (/s)
100
10
1
0.1
0.01
B
BFr1
BTr4
B
Km (mM)
100
10
1
0.1
0.01
B
BFr1
BTr4
C
kcat/Km (/s/mM)
1000
100
10
1
0.1
0.01
0.001
B
BFr1
BTr4
Phe
Leu
Met
Val
Ala
Ser

Fig. 3-9. Ratios of kinetic constants (SGPB/SGPBTr4). Substrates were of the form Suc-AAPX-pNA (where X = Phe, Leu, Met, Val, Ala, or Ser as indicated). Data from Table 3-6. The ratios are plotted on a logarithmic scale; numerical values are noted above each bar. Some k_{cat} or K_m ratios could not be determined because the relevant constants for SGPB or SGPBTr4 could not be determined.


SGPB/SGPBTr4
1000
100
10
1
0.1
6.7
16
1.4
2.7
81
30
2.3
4.6
550
460
3.1
410
13
0.54
kcat
1/Km
kcat/Km
Phe
Leu
Met
Val
Ala
Ser

ValP1; a substantial decrease against AlaP1; and a moderate decrease against MetP1. Combined with a two-fold increase in k_{cat}/K_m against SerP1, the changes shift the primary specificity profile of SGPBTr4 (Met>Ser, Phe>Ala, Leu>Val) substantially in favour of Ser.

The closely related α-LP prefers small P1 side-chains and thus provides a useful comparison with SGPBTr4. The primary specificity of α-LP towards its favoured substrate Ala (k_{cat}/K_m = 26,800 $s^{-1}M^{-1}$) is provided by a high turnover constant (kcat = 62.8 s^{-1}) but relatively weak binding affinity (K_m = 2.34 mM) (Mace et al., 1995). Therefore, the specificity of SGPBTr4 towards SerP1 (k_{cat}/K_m = 950 $s^{-1}M^{-1}$) compares favourably in binding affinity (K_m = 0.94 mM) but is deficient in the turnover rate (k_{cat} = 1.1 s^{-1}). The result is consistent with the nature of the variant library; the targeted mutation sites are implicated in substrate binding but not necessarily in catalysis. Recent studies have shown that residues far removed from the active site contribute to substrate specificity (Hedstrom et al., 1992; Mace et al., 1995). Catalytic contributions and mutational effects at distal positions cannot be reliably predicted on the basis of structural data; a semi-random screening strategy should prove useful in identifying and optimizing distal catalytic residues (section 3.5).

A quadruple mutation consensus (A192S/P192bAorG/T213L/F228Y) (Table 3-5) which shifts the specificity of SGPB away from large hydrophobic residues towards small polar residues (Table 3-6, Fig. 3-9) was identified from 29,952 possibilities. The S1 pocket of SGPBTr4 discriminates against large hydrophobic P1 side-chains while exhibiting a binding affinity for SerP1 which is near optimal for small P1 side-chains. The turnover rate most likely depends on residues not targeted in our library; SGPBTr4 can be used as a

starting point for improving k_{cat} against SerP1 through the variation of other putative specificity determinants.

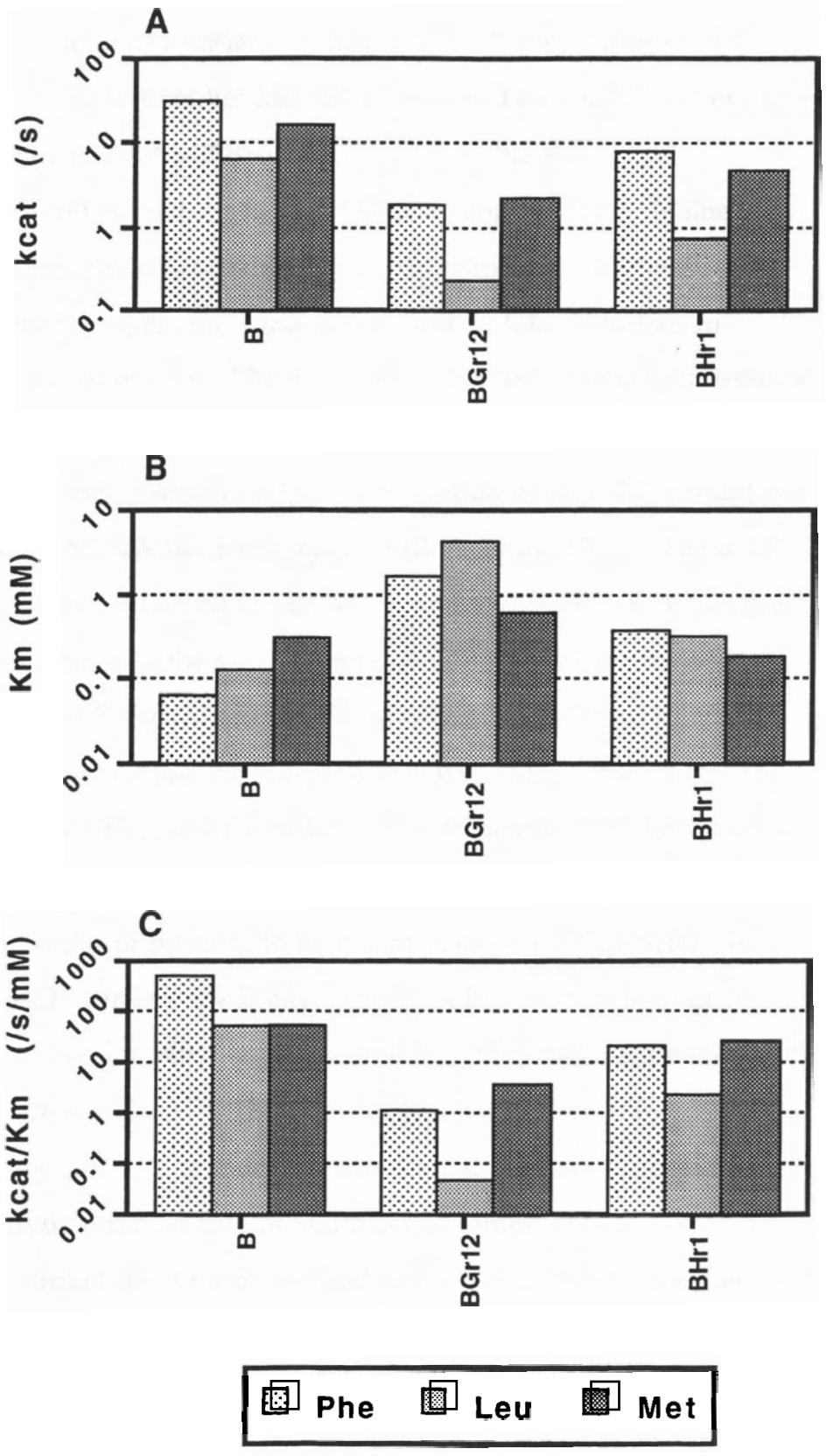
3.4.2.2 Variants with Gly at pmP1

Four variants were isolated from the screen with Gly at pmP1 (Table 3-5). One variant contained the mutation T213L together with wild-type Ala192; two variants contained the mutations A192S/T213L. However, the fourth variant (SGPBGr12) contained a mutation (T213H) not observed in previous screens. The presence of a bulky His side-chain in place of Thr213 at the base of the S1 pocket (Fig. 3-1) should cause unfavourable steric interactions with large P1 side-chains; SGPBGr12 exhibits substantial reductions in activity against substrates with Phe, Leu, or Met at the P1 position (Table 3-6, Fig. 3-10).

3.4.3 Variants with large polar or charged residues at pmP1

Screens with Asn, Gln, Asp, Glu, or Lys at pmP1 failed to produce any clones with zones of clearing. The result was reasonable since the presence of these residues at pmP1 had completely abolished proteolytic activity (Fig. 3-3). It is likely that the library did not contain any members with substantial activity against these P1 substrates (see section 3.5.3 for further discussion). The presence of His at pmP1 had only slightly reduced proteolytic activity, so it was not surprising that the His screen resulted in two variants with increased zones of clearing.

Fig. 3-10. Kinetic constants for SGPB and variants selected with Gly or His at pmP1. Substrates were of the form Suc-AAPX-pNA (where x = Phe, Leu, or Met as indicated). Data from table 3-6. The constants are plotted on logarithmic scales; the enzyme name is indicated along the horizontal axis (*e.g.* "B" designates SGPB). A, k_{cat} values (in units of s^{-1}). B, K_m values (in units of mM). C, k_{cat}/K_m values (in units of $s^{-1}mM^{-1}$).


A
kcat (/s)
100
10
1
0.1
B
BGr12
BHr1
B
Km (mM)
10
1
0.1
0.01
B
BGr12
BHr1
C
kcat/Km (/s/mM)
1000
100
10
1
0.1
0.01
B
BGr12
BHr1
Phe
Leu
Met

3.4.3.1 Variants with His at pmP1

While both His pmP1 variants retained Ala192, they differed in the replacement of Thr213 with either Met (SGPBHr1) or Leu (SGPBHr2) (Table 3-5). The T213M mutation of SGPBHr1 was unique to the His screen. Unfortunately, a synthetic substrate with His at P1 could not be obtained, but SGPBHr1 was kinetically characterized against substrates with Phe, Leu, or Met at P1. Relative to SGPB, the variant exhibited 20-fold decreases in k_{cat}/K_m against substrates with Phe or Leu at P1 but only a two-fold decrease against MetP1 (Table 3-6 and Fig 3-10).

Interestingly, construction and characterization of an α-LP variant with the single mutation M192A has been reported (Bone et al., 1989). The α-LP variant and SGPBHr1 contain A192 and M213, and therefore, the S1 pockets and primary specificities of the two enzymes should be similar. Indeed, the K_m values for the α-LP variant against Suc-AAPF-pNA (1.50 mM), Suc-AAPL-pNA (600 μM), and Suc-AAPM-pNA (220 μM) (Mace et al., 1995) are similar to those of SGPHr1, and the order of k_{cat}/K_m values is the same for both enzymes (Met>Phe>Leu). However, the k_{cat} values for the α-LP variant are more than an order of magnitude higher than those for SGPBHr1. In a recent study of α-LP primary specificity, mutations in a surface loop of 19 amino acids were found to effect both k_{cat} and K_m; the effects were attributed to changes in active site plasticity (Mace et al., 1995). Notably, the corresponding loop of SGPB is much smaller, containing only 11 residues. Perhaps the catalytic turnover rate of SGPBHr1 is compromised relative to that of the α-LP variant due to a more rigid active site conformation imposed by the shortened surface loop. An active site with increased conformational

freedom may more readily accomodate the dynamic changes associated with catalytic turnover.

3.5 General trends in the selection results

3.5.1 Specificities are affected in K_m rather than k_{cat}

We have characterized thirteen unique SGPB variants selected from screens with Leu, Phe, Met, Val, Thr, Gly, or His at pmP1; some general trends are evident. Relative to SGPB, increased primary specificity as measured by k_{cat}/K_m was invariably caused by changes in K_m, k_{cat} values were rarely improved (Tables 3-4 and 3-6). The variant library was designed on the basis of static crystallographic data; for the most part, residues chosen for mutagenesis were those capable of interacting directly with the substrate P1 residue. Such interactions are likely to effect K_m (binding affinity) but may not effect k_{cat}. Residues influencing k_{cat} may not directly contact substrate but may instead alter the positions of critical catalytic residues (*e.g.* the catalytic triad) or modulate the flexibility and plasticity of the enzyme. To significantly increase primary specificity against substrates where k_{cat} limits the maximum practically obtainable k_{cat}/K_m, additional mutations designed to increase k_{cat} may be necessary. The identification and optimization of residues contributing to k_{cat} should lend itself well to a screening strategy analysis.

3.5.2 Mutational tolerance of particular positions

Information has been gained as to which positions are capable of tolerating substitutions. At positions 192a and 192b, while certain amino acids are favoured, all possible substitutions within the library are tolerated. The side-chains of these residues point away from the S1 pocket and do not interact with the substrate P1 side-chain (Fig. 3-1), but certain variations help to stabilize other mutations and thus contribute to primary specificity. Position 228 is tolerant to substitutions; while not directly involved in forming the S1 pocket, residue 228 interacts with residue 213 and cooperative changes at the two positions have substantial effects on primary specificity. In contrast, positions 199 and 216 may not be suitable targets for mutagenesis; at these positions, the wild-type sequence is absolutely conserved among the variants. Perhaps changes at position 216 perturb the interactions of Ser214, a structurally important residue conserved within the S2 serine protease family (Delbaere et al., 1975). Leu199 is located within the hydrophobic core of the protein and changes at this site may not be possible without compensatory changes in surrounding residues.

The side-chains of residues 213 and 192 are expected to contribute significantly to primary specificity; they are located at the base of the S1 pocket and, thus, can interact directly with the P1 side-chain (Fig. 3-1). The substitution T213L seems highly favourable, since it is present in many of the mutants. The crystal structure of SGPB complexed with OMTKY3 shows a distance of only 3.60 Å between Ala192-C^{β} and LeuP1-$C^{\gamma 1}$ (Fig. 3-1); it is not surprising, therefore, that the side-chain at position 192 has significant effects on primary specificity. All variants selected with large hydrophobic pmP1 residues retained wild-type Ala192, providing a hydrophobic S1 pocket and

maintaining chymotrypsin-like specificity. The presence of Ser or Thr at pmP1 allowed for the selection of variants containing the mutation A192S. The reduction in volume and hydrophobicity due to the replacement of Ala192 with Ser increased primary specificity toward Ser while dramatically decreasing specificity towards Phe, Leu, and Val.

3.5.3 Selection success or failure depends on the library

Library screens with large hydrophobic or small polar pmP1 residues were successful, but those with large, polar or charged substrates were unsuccessful. These results may be attributable to the nature of the variant library. The large, hydrophobic S1 pocket of wild-type SGPB is partly formed by the small residues Ala192 and Thr213 at the base. Moderate alterations in specificity towards various hydrophobic substrates can be achieved through minor changes at these positions. The substitution of Ser for Ala192 increased primary specificity towards Ser. The result is reasonable; the mutation should decrease the pocket's volume and hydrophobicity, and thus favour small, polar substrates while discriminating against large, hydrophobic side-chains. An increase in specificity towards large, polar or charged side-chains (*i.e.* Asn, Gln, Asp, Glu, Lys, or Arg) would require the introduction of favourable enzyme/substrate interactions while maintaining a large S1 pocket volume. However, the introduction of a large side-chain at position 192 or 213 (*e.g.* to provide a counter-ion for a charged P1 side-chain) necessarily reduces the pocket volume. Thus, our library likely cannot provide a steric and electrostatic environment complementary to such bulky, hydrophilic substrates; a library varying residues beneath the present base of

the pocket combined with mutations designed to alter the pocket's shape (*e.g.* shifting of the disulfide between Cys191 and Cys220) may be necessary.

3.5.4 Improvement of enzyme thermostability

The folding pathway of the serine proteases requires the propeptide (Silen et al., 1989); thus unfolding of the mature enzymes alone is an irreversible process. We monitored decays in activity at 55 °C and found that the plot of the logarithm of activity against time produced a straight line for all enzymes. Thus, unfolding is a unimolecular process independent of protease concentration and half-lives could be calculated (Fig. 3-7). Our results suggest that there is great potential for increasing the thermostability of SGPB through site-directed mutagenesis; several of the variants exhibit dramatic increases in half-life at 55 °C. Like substrate specificity, thermostability depends on cooperative mutational effects; for example, the single mutation T213L destabilizes SGPBMr2 but is highly stabilizing in conjunction with the substitution F228H (as in SGPBFr13). A general trend appears to be an increase in stability due to the substitution T213L in conjunction with the replacement of Phe228 by an amino acid with a polar, cyclic side-chain (as in SGPBr3, SGPBFr12, SGPBFr13, and SGPBVr2).

3.6 Advantages and disadvantages of the methodology

Our strategy for the isolation of proteases with altered specificities uses the autocatalytic proenzyme activation event for the selection of variants with a desired specificity and the hydrolysis of milk proteins for the detection

of active enzyme against an inactive background. These two factors largely determine the success of the strategy and thus merit further discussion.

3.6.1 Skim milk as a substrate for protease detection

The use of skim milk as a general detection system for proteolytic activity is well established (Wells et al., 1983) and has proven to be a reliable indicator for catalytic efficiency and structural stability (Chen & Arnold, 1993; Graham et al., 1993). The production of a visible zone of clearing requires at least 24 hours and thus is dependent on protein stability. We exploited this fact to select enzymes with thermostabilities comparable to wild-type, but it may be disadvantageous for other purposes. While our studies (section 3.2) and others (Graham et al., 1993) have shown that a wide range of specificities are detectable by milk protein hydrolysis, it cannot be guaranteed that all active enzymes produce comparable zones of clearing. Nonetheless, the low cost and high stability of skim milk relative to most defined synthetic substrates make it an attractive choice. The low cost in particular allowed for the use of several hundred plates (*e.g.* 230 plates in the case of the wild-type promature junction screen) and, even at plating densities of 1.0×10^3 cfu/plate, individual clones could be easily discerned and isolated (Fig. 3-2).

The overall strategy is rapid and facile; the entire process from library design to kinetic analysis required less than two months. However, substrates with lower detection limits and defined cleavage recognition sequences may provide useful improvements to the methodology.

3.6.2 A link between proenzyme activation and specificity

The use of the promature junction sequence as a constraint for substrate specificity has not been previously reported. The strategy is simple and, in theory, allows for the screening of all 20 natural amino acids. The autocatalytic activation mechanism is poorly understood and crystal structures for promature SGPB or related enzymes are unavailabe; thus our description of the process is necessarily simplistic. The nature of the promature junction P1 residue may affect folding events apart from the final processing event; for example, certain side-chains may destabilize intermediates on the folding pathway. Enzymes with the mutation T213L in combination with wild-type Glu192a were prevalent in the screens with Phe, Met, or Val at the promature junction P1 site. These variants exhibited thermostabilities and Leu primary specificity comparable to wild-type, yet they were never observed in the screen with Leu at pmP1. In these cases, the folded state may not be accessible with Leu at the promature junction P1 site. Again, this may be advantageous or disadvantageous depending on the application. A thermodynamically stable conformation which is not kinetically accessible cannot be predicted, and thus, it may be best to allow the selection system to eliminate such combinations.

The substrate specificity of the α-lytic protease has been investigated using a similar approach in which four residues involved in forming the S1 pocket were randomly mutated (Graham et al., 1993). The library was initially screened for proteolytic activity on skim milk plates and the specificity of individual active variants was subsequently tested with a synthetic substrate overlay method. The active variants formed two groups: those with parental primary specificity and those with increased specificity for His and Met. The promature junction P1 site was not varied in this study and, in light of our

results, it is possible that some variants with altered specificities may have been undetectable due to inefficient folding or processing. An experiment varying the S1 pocket and the promature junction simultaneously may produce a greater variety of specificity variants.

In any *in vivo* selection strategy, complexities of a biological system (*i.e.* the bacterial cell) may complicate the process. For example, it is possible that some variant enzymes may be secreted less efficiently than others and, as a result, produce smaller zones of clearing. While such complications cannot be predicted, they should be kept in mind.

3.6.3 Screening strategies and computational methods

While visual detection methods such as the one described in this chapter can be used to rapidly screen large libraries of variant enzymes, the numbers are still relatively small in comparison with the possible variations for even a small protein. For investigations of macromolecular structure, such screens are most powerful in conjunction with methods which reduce the amount of sequence space to be searched. Computational methods and structural data can be used to predict the effects of defined mutations (Bone & Agard, 1991) and, while these predictions are seldom completely reliable, they can provide valuable guidelines for library design. An ideal experiment would be one in which predictive methods are used to identify a key site-directed mutation necessary to produce a desired function, and a screening strategy is used to vary surrounding residues to compensate for perturbations caused by the original mutation. For example, crystallographic studies of trypsin have identified Asp189 at the base of the S1 pocket as the negatively charged side-chain compensating for the substrate positive charge (Read &

James, 1988), but substitutions at this position have produced enzymes with greatly reduced catalytic efficiency (see chapter 1). Perhaps SGPB could be converted to a trypsin-like enzyme by first positioning a negatively charged residue at the base of the S1 pocket, as in trypsin, and then screening a library with variations at surrounding positions to select for stable, catalytically active variants.

3.7 Implications for the folding pathway of SGPB

All members of the S2 serine protease family are secreted in a promature form; the propeptide is extensive and plays a crucial role in the folding pathway (Silen & Agard, 1989; Silen et al., 1989). Previous studies have established that the propeptide acts to lower the energy barrier from a folding intermediate to the native, folded state (Baker et al., 1992). Cleavage of the promature junction to release active, mature protease is an essential event believed to be autocatalytic. Our results support a model in which the promature junction occupies the active site and is cleaved in an intramolecular process.

Leu and Met are equally good P1 substrates in synthetic substrates (Table 3-4), but the introduction of Met at pmP1 greatly reduces proteolytic acitvity (Fig. 3-3). The result is inconsistent with an intermolecular promature junction cleavage, because in such a mechanism the junction must be exposed on the protein surface where it would likely resemble a free peptide substrate. In an intramolecular mechanism, the S1 binding pocket would fold around the pmP1 side-chain which could, therefore, interact with folding intermediates and transition states. In such a process, pmP1 side-chains which are good substrates for mature SGPB may nonetheless hinder

the dynamic folding pathway through unfavourable interactions which do not occur in the native enzyme-substrate complex.

The hypothesis is further supported by the effects of Gly, Ser, or Thr at pmP1. These side-chains are poor substrates for SGPB yet their effect on active enzyme production is similar to that of the good substrate Met. Small P1 side-chains are unlikely to provide favourable interactions with the large S1 pocket of SGPB, but they are also unlikely to provide unfavourable steric interactions along the folding pathway.

The side-chain of Gly (a hydrogen atom) at pmP1 likely does not contribute to the folding pathway; the effects of other substitutions relative to Gly at pmP1 provide indications of the balance between favourable and unfavourable interactions. Large, polar or charged pmP1 residues completely abolish proteolytic acitivity; unfavourable interactions predominate in such cases. The presence of Ser, Thr, Val, or Met at pmP1 greatly diminishes but does not abolish active enzyme secretion. In these cases, proteolytic activity is similar to that provided by Gly at pmP1, suggesting that favourable and unfavourable interactions are closely balanced. Phe, His, or wild-type Leu at pmP1 allow high proteolytic activity and thus contribute favourably to the folding mechanism. Phe is a better P1 substrate than Leu, but the Phe side-chain is bulkier than that of Leu. The presence of Leu at the promature junction P1 site of SGPB (as well as SGPA, SGPC, and SGPD) may be a compromise maximizing favourable interactions with the S1 pocket while minimizing unfavourable steric interactions during folding.

4 CONCLUSION

The autocatalytic activation of promature SGPB links substrate specificity to enzyme activity; only those enzymes capable of efficient self-processing produce detectable proteolytic activity in *E. coli.* The reduction in activity caused by the introduction of a poor substrate at the promature junction P1 site can be restored by compensatory changes in the S1 pocket. Screening of a mutant SGPB expression library in conjunction with large hydrophobic or small polar residues at pmP1 resulted in the selection of enzymes with altered substrate specificities and improved thermostabilities. Even unsuccessful selections (*e.g.* those with charged P1 residues) provided information useful for the design of new libraries.

Both substrate specificity and thermostability were found to be dependent on cooperative mutational effects, supporting the view that multiple changes are necessary to alter substrate specificity while maintaining structural stability and catalytic efficiency. Current limits on the ability to predict the effects of a given mutation underscore the need for a larger mutant protein database relating catalytic activity to tertiary structure. Our methodology can be used to screen large libraries for mutant proteases with desired alterations in substrate specificity but with thermostabilities and catalytic efficiencies comparable to natural enzymes. Furthermore, levels of protein expression from our *B. subtilis* expression system are sufficient for crystallographic studies which should shed light on the observed mutational effects.

REFERENCES

Awad, W. M., Soto, A. R., Siegel, S., Skiba, W. E., Bernstrom, G. G., & Ochoa, M. S. (1972) The proteolytic enzymes of the K-1 strain of *Streptomyces griseus* obtained from a commercial preparation (Pronase)
1. Purification of four serine endopeptidases. *J. Biol. Chem.* **247**, 4144-4154.

Baker, D., Sohl, J. L., & Agard, D. A. (1992) A protein-folding reaction under kinetic control. *Nature* **356**, 263-265.

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

Bauer, C.-A. (1978) Active centers of *Streptomyces griseus* protease 1, *Streptomyces griseus* protease 3, and α-chymotrypsin: Enzyme-substrate interactions. *Biochemistry* **17**, 375-380.

Bauer, C. A., Brayer, G. D., Sielecki, A. R., & James, M. N. G. (1981) Active site of α-lytic protease. Enzyme-substrate interactions. *Eur. J. Biochem.* **120**, 289-294.

Bode, W., Papamokos, E., Musil, D., Seemuller, U., & Fritz, H. (1986) Refined 1.2 Å crystal structure of the complex between subtilisin Carlsberg and the inhibitor eglin C. Molecular structure of eglin and its detailed interaction with subtilisin. *EMBO J.* **5**, 813-818.

Bone, R., & Agard, D. A. (1991) Mutational remodeling of enzyme specificity. *Methods Enzymol.* **202**, 643-671.

Bone, R., Fujishige, A., Kettner, C. A., & Agard, D. A. (1991) Structural basis for broad specificity in α-lytic protease mutants. *Biochemistry* **30**, 10388-10398.

Bone, R., Silen, J. L., & Agard, D. A. (1989) Structural plasticity broadens the specificity of an engineered protease. *Nature* **339**, 191-195.

Carson, D. (1994). Recombinant expression of *Streptomyces griseus* genes and their variants in *Bacillus subtilis* and *Escherichia coli*. Simon Fraser University, Vancouver, Canada.

Carter, P., & Wells, J. A. (1987) Engineering enzyme specificity by 'substrate-assisted catalysis'. *Science* **237**, 394-399.

Carter, P., & Wells, J. A. (1988) Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564-568.

Chang, S., & Cohen, S. N. (1979) High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. *Mol. Gen. Genet.* **168**, 111-115.

Chen, K., & Arnold, F. H. (1993) Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catlysis in dimethylformamide. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5618-5622.

Cherepova, N. V., Baykousheva, S. P., & Ilieva, K. Z. (1986) Ultracytochemical localization of ATP-hydrolysing activity in vegetative cells, spores and isolated cytoplasmic membranes of *Bacillus subtilis* 168. *J. Gen. Microbiol.* **132**, 669-675.

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**, 1784-1790.

Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G., & Thiessen, W. E. (1975) Tertiary structural differences between microbial serine proteases and pancreatic serine proteases. *Nature* **257**,

Din, N., Gilkes, N. R., Tekant, B., Miller, R. C. J., Warren, A. J., & Kilburn, D. (1991) Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Biotechnol.* **9**, 1096-1099.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., & Wells, J. A. (1986) Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering. *Science* **233**, 659-663.

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**, 6659-6663.

Felsenstein, J. (1989) *Cladistics* **5**, 164-166.

Fersht, A. R. (1985) in *Enzyme Structure and Mechanism*(New York, W. H. Freeman and Company.

Fersht, A. R. (1987) Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. *Biochemistry* **26**, 8031-8037.

Fujinaga, M., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1985) Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.* **183**, 479-502.

Fujishige, A., Smith, K. R., Silen, J. L., & Agard, D. A. (1992) Correct folding of α-lytic protease is required for its extracellular secretion from *Escherichia coli*. *J. Cell Biol.* **118**, 33-42.

Gavel, Y., & von Heijne, G. (1990) Cleavage-site motifs in mitochondrial targeting peptides. *Protein Eng.* **4**, 33-37.

Graf, L., Craik, C. S., Patthy, A., Roczniak, 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**, 2616-2623.

Graf, L., Jancso, A., Szilagyi, L., Hegyi, G., Pinter, K., Naray-Szabo, G., Hepp, J., Medzihradszky, K., & Rutter, W. J. (1988) Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4961-4965.

Graham, L. D., Haggett, K. D., Jennings, P. A., Brocque, D. S. L., Whittaker, R. G., & Schober, P. A. (1993) Random mutagenesis of the substrate-binding site of a serine protease can generate enzymes with increased activities and altered primary specificities. *Biochemistry* **32**, 6250-6258.

Greenwalt, J. W., & Whiteside, T. L. (1975) Mesosomes: membranous bacterial organelles. *Bacteriol. Rev.* **39**, 405-463.

Greer, J. (1990) Comparative modeling methods: Application to the family of mammalian serine proteases. *Proteins* **7**, 317-334.

Guan, C., Li, P., Riggs, P. D., & Inouye, H. (1987) Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67**, 21-30.

Hartl, F. U., & Neupert, W. (1990) Protein sorting to mitochondria: evolutionary conservation of folding and assembly. *Science* **247**, 930-938.

Hartley, B. S. (1970) Homologies in serine proteases. *Philos. Trans. R. Soc. Ser. B.* **257**, 77-87.

Hartley, B. S. (1979) Evolution of enzyme structure. *Proc. R. Soc. Lond. Ser B.* **205**, 443-452.

Harwood, C. R. (1989). Introduction to the biotechnology of *Bacillus*. Bacillus. New York and London, Plenum Press. 1-4.

Hedstrom, L., Perona, J. J., & Rutter, W. J. (1994) Converting trypsin to chymotrypsin: Residue 172 is a substrate specificity determinant. *Biochemistry* **33**, 8757-8763.

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

Henderson, G., Krygsman, P., Liu, C. J., Davey, C. C., & Malek, L. T. (1987) Characterization and structure of genes for proteases A and B from *Streptomyces griseus. J. Bacteriol.* **169**, 3778-3784.

Hopwood, D. A., Bibb, M. J., Chater, K. F., Kieser, T., Bruton, C. J., Kieser, H. M., Lydiate, D. J., Smith, C. P., Ward, J. M., & Schrempf, H. (1985) in *Genetic manipulation of Streptomyces: a laboratory manual.*(Norwich, U.K., The John Innes Foundation.

Horne, D., & Tomasz, A. (1985) Pneumonococcal Forssman antigen: Enrichment in mesosomal membranes and specific binding to the autolytic enzyme of *Streptococcus pneumoniae. J. Bacteriol.* **161**, 18-24.

Huber, R., & Bode, W. (1978) *Acc. Chem. Res.* **11**, 114-122.

Hutter, R., & Eckhardt, T. (1988). Genetic Manipulation. Actinomycetes in Biotechnology. London, Academic Press. 89-184.

James, M. N. G., Delbaere, L. T. J., & Brayer, G. D. (1978) Amino acid sequence alignment of bacterial and mammalian pancreatic serine proteases based on topological equivalences. *Can. J. Biochem.* **56**, 396-402.

James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J., & Bauer, C. A. (1980) Structures of product and inhibitor complexes of *Streptomyces griseus* protease A at 1.8 Å resolution. *J. Mol. Biol.* **144**, 43-88.

Jindal, S., Dudani, A. K., Singh, B., Harley, C. B., & Gupta, R. S. (1989) Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. *Mol. Cell Biol.* **9**, 2279-2283.

Kawamura, F., & Doi, R. H. (1984) Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. *J. Bacteriol.* **160**, 442-444.

Kim, J. C., Cha, S. H., Jeong, S. T., Oh, S. K., & Byun, S. M. (1991) Molecular cloning and nucleotide sequence of *Streptomyces griseus* trypsin gene. *Biochem. Biophys. Res. Commun.* **181**, 707-713.

Kraut, J. (1977) Serine proteases: Structure and mechanism of catalysis. *Ann. Rev. Biochem.* **46**, 331-358.

Kurylowicz, W., Kurzatkowski, W., Williams, S. T., Woznicka, W., & Paszkiewicz, A. (1975) in *Atlas of Ultrastructure of Streptomyces in Course of Biosynthesis of Antibiotics*(Warszawa, Poland, PZWL-Polish Medical Publishers.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lemire, B. D., Fankhauser, C., Baker, A., & Schatz, G. (1989) The mitochondrial targeting function of randomly generated peptide sequences correlates with predicted helical amphipathicity. *J. Biol. Chem.* **264**, 20206-20215.

Lovett, P., & Amulos, N. P. J. (1989). Plasmid DNA isolation. Bacillus. New York and London, Plenum. 355.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265-275.

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

Maina, C. V., Riggs, P. D., Grandea, A. G. I., Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A., & Guan, C. (1988) A vector to express and purify foreign proteins in *Escherichia coli* by fusion to, and separation from, maltose binding protein. *Gene* **74**, 365-373.

Matthews, B. W., Sigler, P. B., Henderson, R., & Blow, D. M. (1967) Three-dimensional structure of tosyl-α-chymotrypsin. *Nature* **214**, 652-656.

McPhalen, C. A., & James, M. N. G. (1988) Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin C-subtilisin Carlsberg and CI-2-subtlisin Novo. *Biochemistry* **27**, 6582-6598.

Miller, J. H. (1972) in *Experiments in molecular genetics*(Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Nakasone, N., Masuda, K., & Kawata, T. (1987) Characterization of mesosomes of Micrococcus luteus: Isolation and properties of mesosomal ribosomes and location of penicillin-binding proteins in mesosomal membranes. *Microbiol. Immunol.* **31**, 403-415.

Neurath, H. (1984) Evolution of proteolytic enzymes. *Science* **224**, 350-357.

Nienaber, V. L., Breddam, K., & Birktoft, J. J. (1993) A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry* **32**, 11469-11475.

Norrander, J., Kempe, T., & Messing, J. (1983) Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* **26**, 101-106.

Ohta, Y., Hojo, H., Aimoto, S., Kobayashi, T., Zhu, X., Jordan, F., & Inouye, M. (1991) Pro-peptide as an intermolecular chaperone: Renaturation of denatured subtilisin E with a synthetic pro-peptide. *Mol. Microbiol.* **5**, 1507-1510.

Ostermann, J. (1990). Bio/Technology & Genetic Engineering Reviews. Andover, Intercept. 219-249.

Perona, J. J., & Craik, C. S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Science* **4**, 337-360.

Perona, J. J., Evnin, L. B., & Craik, C. S. (1993) A genetic selection elucidates structural determinants of arginine versus lysin specificity in trypsin. *Gene* **137**, 121-126.

Perona, J. J., Tsu, C. A., Craik, C. S., & Fletterick, R. J. (1993) Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J. Mol. Biol.* **230**, 919-933.

Polgar, L. (1989). Structure and function of serine proteases. Mechanisms of protease action. Boca Raton, Florida, CRC Press. Chapter 3.

Poulos, T. L., Alden, R. A., Freer, S. T., Birktoft, J. J., & Kraut, J. (1976) Polypeptide halomethyl ketones bind to serine proteases as analogs of the tetrahedral intermediate. *J. Biol. Chem.* **251**,

Racher, K. I. (1996). Introduction of novel amino acids into proteins: A study of specificity and editing in the yeast isoleucyl-tRNA synthetase. Simon Fraser University, Vancouver, Canada.

Randall, L. L., & Hardy, S. J. S. (1989) Unity in function in the absence of consensus in sequence: role of leader peptides in export. *Science* **243**, 1156-1159.

Rawlings, N. D., & Barrett, A. J. (1994) Families of serine peptidases. *Methods Enzymol.* **244**, 19-61.

Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry* **22**, 4420-4433.

Read, R. J., & James, M. N. G. (1988) Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. *J. Mol. Biol.* **200**, 523-551.

Ruhlmann, A., Kukla, D., Schwager, P., Bartels, K., & Huber, R. (1973) Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **77**, 417-436.

Ruiz-Herrera, J. (1991) in *Fungal Cell Wall: Structure, synthesis and assembly* London, CRC Press.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*(N. Irwin) Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Schwartz, L. B. (1994) Tryptase: A mast cell serine protease. *Methods Enzymol.* **244**, 88-100.

Seidah, N. G., & Chretien, M. (1994) Pro-protein convertases of subtilisin/kexin family. *Methods Enzymol.* **244**, 175-188.

Sidhu, S., Kalmar, G., & Borgford, T. (1993) Characterization of the gene encoding the glutamic acid-specific protease of *Streptomyces griseus*. *Biochem. Cell Biol.* **71**, 454-461.

Sidhu, S. S., Kalmar, G. B., Willis, L., & Borgford, T. J. (1994) *Streptomyces griseus* protease C: A novel enzyme of the chymotrypsin superfamily. *J. Biol. Chem.* **269**, 20167-20171.

Sidhu, S. S., Kalmar, G. B., Willis, L., & Borgford, T. J. (1995) Protease Evolution in *Streptomyces griseus*: Discovery of a novel dimeric enzyme. *J. Biol. Chem.* **270**, 7594-7600.

Silen, J. L., & Agard, D. A. (1989) The α-lytic protease pro-region does not require a physical linkage to activate the protease domain *in vivo*. *Nature* **341**, 462-464.

Silen, J. L., Frank, D., Fujishige, A., Bone, R., & Agard, D. A. (1989) Analysis of prepro-α-lytic protease expression in *Escherichia coli* reveals that the pro region is required for activity. *J. Bacteriol.* **171**, 1320-1325.

Silen, J. L., McGrath, C. N., Smith, K. R., & Agard, D. A. (1988) Molecular analysis of the gene encoding α-lytic protease: evidence for a preproenzyme. *Gene* **69**, 237-244.

Simon, M. M., & Kramer, M. D. (1994) Granzyme A. *Methods Enzymol.* **244**, 68-79.

Stroud, R. M. (1974) A family of protein-cutting proteins. *Sci. Am.* **231**, 74-88.

Svendsen, I., Jensen, M. R., & Breddam, K. (1991) The primary structure of the glutamic acid-specific protease of *Streptomyces griseus*. *FEBS Lett.* **292**,

Tapani, R., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T. A., Knowles, J. K. C., & Teeri, T. T. (1992) Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolyase. *Proteins* **14**, 475-482.

Venner, T. J., & Gupta, R. S. (1990) Nucleotide sequence of mouse HSP60 (chaperonin, GroEL homolog) cDNA. *Biochim. Biophys. Acta* **1087**, 336-338.

Vogelli, G., & Kaytes, P. S. (1987) Amplification, storage and replication of libraries. *Methods Enzymol.* **152**, 407-415.

von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683-4690.

Watanabe, T., Oyanagi, W., Suzuki, K., Ohnishi, K., & Tanaka, H. (1992) Structure of the gene encoding chitinase D of *Bacillus circulans* WL-12 and possible homology of the enzyme to other prokaryotic chitinases and class III plant chitinases. *J. Bacteriol.* **174**, 408-414.

Watanabe, T., Oyanagi, W., Suzuki, K., & Tanaka, H. (1990) Chitinase system of *Bacillus circulans* WL-12 and importance of chitinase A1 in chitin degradation. *J. Bacteriol.* **172**, 4017-4022.

Watson, H. C., Shotton, D. M., Cox, J. C., & Muirhead, H. (1970) Three-dimensional Fourier synthesis of tosyl elastase at 3.5 Å resolution. *Nature* **225**, 806-811.

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. USA* **84**, 5167-5171.

Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., & Chen, E. Y. (1983) Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. *Nucleic Acids Res.* **11**, 7911-7925.

Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P., & Estell, D. A. (1987) Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. USA* **84**, 1219-1223.

Wilkins, R. J., & Snell, R. G. (1987) Centrifugal transfer and sandwich hybridisation permit 12-hour Southern blot analyses. *Nucleic Acids Res.* **15**, 7200.

Wilson, C., Gregoret, L. M., & Agard, D. A. (1993) Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. *J. Mol. Biol.* **229**, 996-1006.

Wilson, C., Mace, J. E., & Agard, D. A. (1991) Computational method for the design of enzymes with altered substrate specificity. *J. Mol. Biol.* **220**, 495-506.

Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K., & Makisumi, S. (1988) Purification and characterization of an acidic amino acid specific endopeptidase of *Streptomyces griseus* obtained from a commercial preparation (Pronase). *J. Biochem.* **104**, 451-456.

Young, C. L., Barker, W. C., Tomaselli, C. M., & Dayhoff, M. O. (1978). Atlas of Protein Sequence and Structure. Washington D. C., National Biomedical Research Foundation. 73.