Structure, Function and Folding of a Novel Subtilisin Serine Protease

Michael Gamble B.Sc. (Hons)

Being a thesis presented in accordance with the regulations governing the award of degree of Doctor of Philosophy in Cardiff University

May 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date 18/5/2010

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of .....(insert ~~MCh, MD, MPhil~~, PhD etc, as appropriate)

Signed ..... (candidate) Date 18/5/2010

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by explicit references.

Signed ..... (candidate) Date 18/5/2010

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate) Date 18/5/2010

Abstract

ISPs constitute the major cellular proteolytic activity of many *bacilli,* yet their physiological role, mechanism of regulation and 3D structure were unknown. ISP from *B. clausii* is expressed as a dimeric inactive precursor. Dimerisation is not involved in regulating ISP activity. The 3D structure of ISP revealed residues from the C-terminus cross over and interact with the adjoining protomer distant from either active site. The mechanism for ISP activation involves inter-molecular processing of the 18 residue N-terminus. The ISP structure exposed a novel mechanism by which proteolytic activity is regulated: The N-terminal extension binds back over the active site with the proline from the conserved LIPY sequence inducing a kink in the polypeptide backbone positioning the hydrolysable peptide bond beyond reach of the catalytic serine. The N-terminal extension acted as a potent inhibitor when added *in vitro* (K_i of 5×10^{-7} M). The majority of ESPs require a prodomain to fold to a native and active conformation. ISP refolds without a classical prodomain and is therefore thermodynamically stable, in contrast to the kinetically stable ESPs. Removal of calcium from ISP results in loss of activity and protein likely adopting a partially unfolded monomeric state. The structure of ISP indicated that calcium is bound at a high-affinity binding site conserved amongst the subtilases.

ISP preferentially degraded unfolded protein rather than substrates in native conformation. Also, ISP has a preference for large hydrophobic residues at the P1 and P4 substrate binding sites. This supports the hypothesis that ISP is involved in processing of misfolded/unfolded protein. Molecular insights confirm the primary sequence features novel to the ISPs translate to unique structural, functional, folding and regulatory properties amongst the subtilase family. This allows us to postulate the physiological roles for ISPs and how this role differs from their close relatives, the ESPs.

I dedicate this thesis to my family,
for their constant love and support.

Acknowledgements

I would first of all like to thank my supervisor Dr Dafydd Jones for both accepting me as his first PhD student, a unique honour, and for all the advice and support that has been instrumental in completing this project. I have thoroughly enjoyed all aspects of this work, even the 'writing up' which sounds so innocuous a phrase but would have been a mountain to climb without his patient help and guidance. I will fondly remember the lab outings to many parts of the Brecon Beacons, the barbecues during the 'great' British summer, and the many late night trivial pursuit and poker marathons showing us that there is life outside the lab. This was all very much appreciated.

To the current members of the Jones group I am especially grateful to Dr Wayne Edwards for his seemingly unlimited scientific knowledge, and to Dr Amy Baldwin for her good humour and welcome constructive criticism. To James thank you for your enthusiasm and friendship, good luck for the future. And finally Eduardo thank you for showing me physicists are good fun and introducing potato pizza to Wales.

Many thanks to Dr Alan Simm for his help during the first year of this project and insistence that life is nothing without the search for knowledge.

My special thanks to Professor John Kay, for his gift of chromogenic peptides and helpful suggestions. Following some of our conversations I am convinced that a documentary of your scientific career would be fascinating.

To Mrs Joan Hubbard for all your help, many amusing conversations and occasional 'cheer up' sweets.

I would like to thank Professor Rudolf Allemann, Cardiff School of Chemistry for the kind use of his circular dichroism machine. Also thanks to Dr Rob Jenkins, Cardiff School of Chemistry for the use of the MALDI-TOF mass spectrometer and enabling me to fire lasers at my precious proteins.

I am indebted to Professor Keith Wilson and his group at York University, especially Dr Jitka Vévodová, Dr Antonio Ariza and Professor Eleanor Dodson, for persistence solving the structure of ISP.

The financial support of the Cardiff University is gratefully acknowledged.

I would like to thank my brother Chris and sister Gillian for teaching me the healthy art questioning everything told to me, their inquisitive nature and huge support with everything.

I would wholeheartedly like to thank my parents for their support and patience throughout the years and for letting me take over one of their rooms to write this. Who knew we would get here - this is for you.

Contents

Units and Abbreviations

Units

A	Ampere
bp	Base pair
Da	Dalton
F	Farad
g	Gravitational acceleration (9.8 m/s/s)
g	Grams
h	Hour
k	Kilo
l	Litre
m	Milli
M	Molar
min	Minute
n	Nano
OD/D	Optical density/Attenuance
p	Pico
psi	Pounds per square inch
s	Second
μ	Micro
V	Volts
Ω	Ohms
°C	Degree Celsius

Abbreviations

AP	Alkaline phosphatase
APS	Ammonium persulphate
BSA	Bovine serum albumin
CV	Column volume
CD	Circular dichroism
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediamine-tetraacetic acid
ESP	Extracellular subtilisin protease
FPLC	Fast protein liquid chromatography
IMAC	Immobilised metal affinity chromatography
ISP	Intracellular subtilisin protease
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LDH	Lactate dehydrogenase
Kav	SEC partition coefficient
LB	Luria Bertani
MDH	Malate dehydrogenase
Mw	Molecular weight
Nle	Norleucine
Orn	Ornithine
pNA	*para*-nitroaniline
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulphonyl fluoride
pNA	*Para*-nitroanilide
PVDF	Polyvinylidene difluoride
rpm	Revolutions per minute
RNA	Ribonucleic acid
SEC	Size exclusion chromatography
SDS	Sodium dodecyl sulphate
Suc	Succinyl
TAE	Tris-acetate EDTA buffer
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tris	2-amino-2-hydroxymethylpropane
uv	Ultraviolet
vis	Visible
w/v	weight/volume
w/w	weight/weight

Chapter 1

Introduction

1. Introduction

1.1 Proteolytic enzymes

Proteases are an important class of enzymes that are ubiquitous in nature and play numerous important biological and non-biological roles. They catalyse hydrolysis of peptide bonds using a number of different reaction mechanisms. Found in all organisms their functions range from the initiation and propagation of the blood coagulation cascade (Davie *et al.*, 1991) to the control of cellular sodium balance (Huntington 2008) to blood pressure maintenance (Vallet *et al.*, 1997) and gastric digestion (Richter *et al.*, 1998). Proteases also have extensive uses within the biotechnology, industry and medicine. For example trypsin, a protease normally found in the digestive tract, is used in cell culture to separate bound cells (Ricardo *et al.*, 2008) and tissue plasminogen activator, which catalyses the activation of plasmin to breakdown blood clots, is used in the treatment of an ischemic stroke (Yepes *et al.*, 2009).

The term protease can be used synonymously with the term peptidase, which until clarified by Barrett and McDonald (1986) was used exclusively for enzymes that hydrolysed small polypeptides. At present the terms peptidase and protease can be used generally for all peptide bond hydrolase enzymes whereas proteinase is used exclusively for endopeptidases.

1.1.1 The peptide bond

The peptide bond (amide bond; Figure 1) is formed via a condensation reaction joining the amine group of one amino acid and the carboxyl group of another. Polymerisation of amino acid residues creates a polypeptide chain with a repeating backbone consisting of an amide nitrogen, alpha carbon, to which a side chain group is attached, and carbonyl carbon. The lone pair of electrons on the nitrogen delocalises with the carbonyl oxygen, which results in the amide bond (1.32 Å) being shorter than a normal C-N bond (1.47 Å) but longer than a normal C=N bond (1.25 Å). This confers a partial double bond character that restricts the free rotation around the amide bond creating a very stable, rigid, planar bond. There is therefore two degrees of freedom per residue along the polypeptide chain: the nitrogen and alpha carbon bond, called the Phi

bond (ϕ) and the carbonyl carbon and alpha carbon bond, called the Psi bond (ψ). These form a number of dihedral angles that define the proteins tertiary structure.

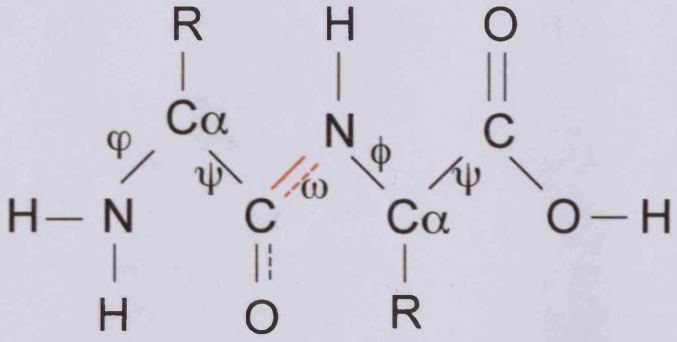


Figure 1.1 Structure of the basic unit of a polypeptide chain. The amide nitrogen, α-carbon and carbonyl carbon repeating unit with amide bond (red, ω) and rotational bonds Phi (ϕ) and Psi (ψ) labelled.

1.1.2 Breaking the peptide bond

Hydrolysis of the amide bond in water is very slow with a half-life in solution of centuries (Radzicka and Wolfenden, 1996). This process can be accelerated under acidic or basic conditions (Figure 1.2). A strong acid causes the protonation of the amide nitrogen, which breaks the delocalisation of the nitrogen lone pair allowing attack and cleavage of the peptide bond by the conjugate base. The mechanism for peptide bond cleavage within an alkali solution involves direct nucleophilic attack by a hydroxyl group, formed from the dissociation of a base (Figure 1.2B). While the peptide bond is broken in both mechanisms, the reaction requires extreme conditions such as high acid or base concentrations together with high temperatures (>100°C) for more than 24 hours, which are far beyond biological systems.

To allow breakdown and turnover of protein in a biologically relevant timescale (milliseconds to seconds) biological systems use enzymes known as proteases. All proteases utilize a reactive nucleophile for peptide bond hydrolysis. The majority of nucleophiles contain a pair of electrons that act as a Lewis base donating electrons to an electrophile. The hydrolysis of an amide or ester bond is highly dependent on the efficacy of the nucleophile and, therefore, the availability of the electrons.

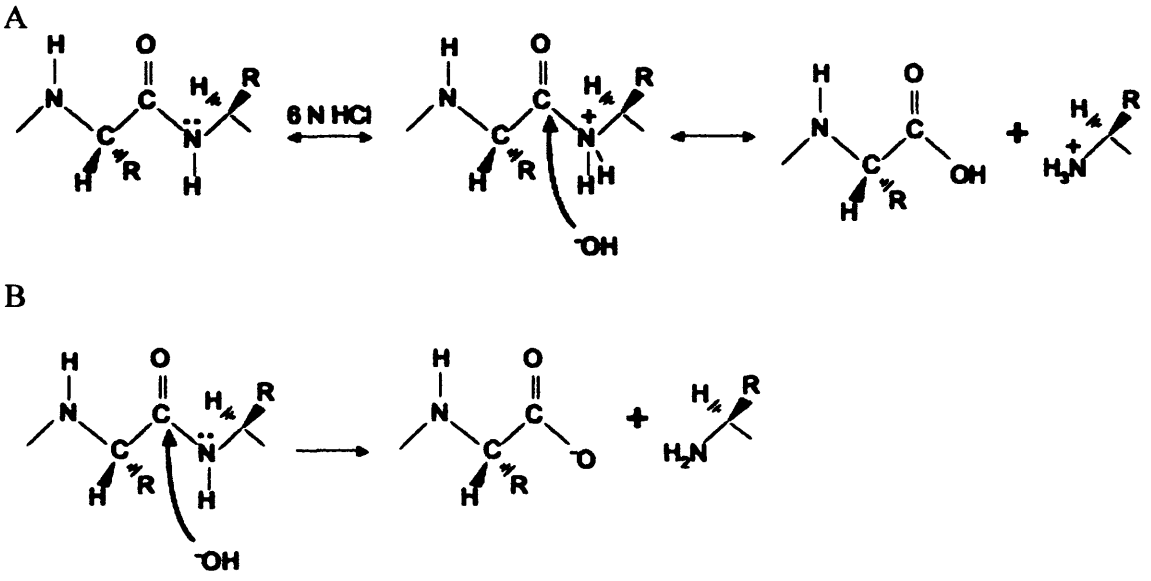


Figure 1.2 Breakdown of the peptide bond by (A) acid or (B) alkali hydrolysis. Acid hydrolysis involves protonation of the amide nitrogen and attack by the hydroxide anion, whereas in a highly basic solution the hydroxide anion is available for direct attack of the carbonyl group.

The relative strength of a nucleophile relies on its electronegativity and polarisation. An increased electronegativity translates to a higher affinity for electrons and therefore decreased nucleophilicity since the atom is less likely to donate electrons. The polarity of the nucleophile also changes the availability of the electron pair. As the distance between the valance electrons and the positively charged nucleus increases, the atom becomes polar, increasing the ability to donate electrons. Another important factor in nucleophilicity is the solvent; hydrogen-bonding solvents act to decrease nucleophilicity by interacting with the free electrons of the nucleophile, thus reducing their reactivity.

1.1.3 The protease families

Proteases are classified according to their catalytic mechanism into six main classes. Aspartic, cysteine, glutamic, metallo, serine and threonine type proteases. The serine, cysteine and threonine proteases activate their respective active site residues (hydroxyl, thiol and hydroxyl groups) for initial nucleophilic attack via deprotonation by a general base, normally a histidine (Storer and Menard 1994, Hedstrom 2002). The reaction mechanisms are similar with acylation and deacylation steps, occurring during peptide

bond cleavage, and the use of a water molecule as a second nucleophile (described in detail later). The aspartic proteases contain two aspartate residues in their active site together with a water molecule hydrogen bonded to the carboxyl groups of both aspartate residues. Due to differing pK_a values one aspartate acts as a general base deprotonating a water molecule to use as a nucleophile to attack the carbonyl carbon. The other aspartate acts as a general acid donating a proton to the α-carbon forming a tetrahedral intermediate (Davies 1990). This breaks down releasing the carboxyl termini. The protonated aspartate is deprotonated by the amide nitrogen on the leaving group and the enzyme is regenerated by the entrance of a water molecule (Coates *et al.*, 2008). The metallo proteases coordinate a metal ion, mainly zinc, at their active site using two histidine residues, a water molecule and either third histidine or a glutamic acid. The water molecule is deprotonated by a separate glutamic acid acting as a general base, which promotes nucleophilic attack by the hydroxyl oxygen on the carbonyl carbon of the peptide bond. A tetrahedral intermediate is formed similarly to that seen with the aspartic proteases, followed by its breakdown and release of the amide leaving group after protonation by the glutamic acid this time acting as a general acid. The metal ion is used to stabilise the negative charge generated on the peptide carbonyl oxygen (Matthews 1988, Hernick and Fierke 2005). The glutamic acid proteases also employ a water molecule as a nucleophile. The water is deprotonated by the active site glutamic acid that in turn attacks the peptide carbonyl carbon in a mechanism similar to the metallo protease except stabilisation of the oxyanion is by a glutamine residue (Fujinaga *et al.*, 2004).

These classes are further arranged into clans that are subdivided into families and species of homologous proteins. To date over 3000 species of proteases have been identified totalling over 140,000 sequences found within bacteria, archaea, protozoa, fungi, plants, animals and viruses (Rawlings *et al.*, 2008, http://merops.sanger.ac.uk).

1.2 Serine proteases

The serine proteases are the largest of all the classes, representing over 1,200 proteins. Of these, the 3D structure of over 200 has been solved, making this class of enzymes one of the most studied in biology. This class is further divided into approximately 46 families with five main families: chymotrypsin, subtilisin, carboxypeptidase Y, Clp and sedolisin proteases. The Chymotrypsin (S1) family is the

largest with over 500 proteins involved in food digestion (chymotrypsin A and trypsin) (Blow 1976) and blood coagulation cascade (factors VII and IX-XII) (Davidson *et al.*, 2003). The subtilisin family (S8) is the second largest with nearly 200 proteins are predominately secreted from the cell for nutrition, however the intracellular serine proteases form the basis of this thesis. The carboxypeptidase Y family (S10) differ from most serine proteases as they function under acidic rather than alkaline conditions. A conserved glutamic acid residue forms a glutamic acid bridge that increases stability at low pH (Mortensen and Breddam 1994, Bullock *et al.*, 1996). The proteases within this family are important in the innate immune response and structural assembly (Pejler *et al.*, 2007 and Pshezhetsky *et al.*, 2009). The Clp family (S14) is mainly involved in the degradation of incomplete polypeptides when translation has been halted or interrupted (Gottesman *et al.*, 1998). The sedolisin family (S53) use a serine residue as their nucleophile however the traditional histidine is replaced by a glutamic acid that acts as a general acid and base during hydrolysis. In prokaryotes these proteases are involved in proteolysis for nutrition whereas in eukaryotes they have been shown to be the cause of an autosomal recessive neurodegenerative disease. The deficiency of tripeptidyl-peptidase I, a lysosomal protease, causes the accumulation of material in the brain that causes axon degeneration (Sleat *et al.*, 2004).

1.3 Catalytic mechanism of serine proteases

An aspartate and a histidine residue in addition to the active site serine form a catalytic triad (Figure 1.3) required for hydrolysis of peptide bonds. The catalytic triad forms a hydrogen-bonded network between the $N^{\delta 1}$ of histidine and the carboxylate oxygen on the aspartate and between $N^{\varepsilon 2}$ of histidine and hydroxyl of serine.

The first serine proteases identified (chymotrypsin, trypsin and elastase) had high sequence and structural similarities suggesting that these proteins evolved via divergent evolution from a common ancestor. The discovery of additional serine proteases from different sources, such as the subtilisin class, that had entirely different primary and tertiary structures but nevertheless contained a virtually super imposable catalytic triad demonstrated convergent evolution. The order of the catalytic residues within the primary sequence (Table 1.1) of these proteins is quite different suggesting it is unlikely that these proteins evolved from a common ancestor. It therefore appears

that nature has evolved the same catalytic mechanism on at least four separate occasions from different starting points.

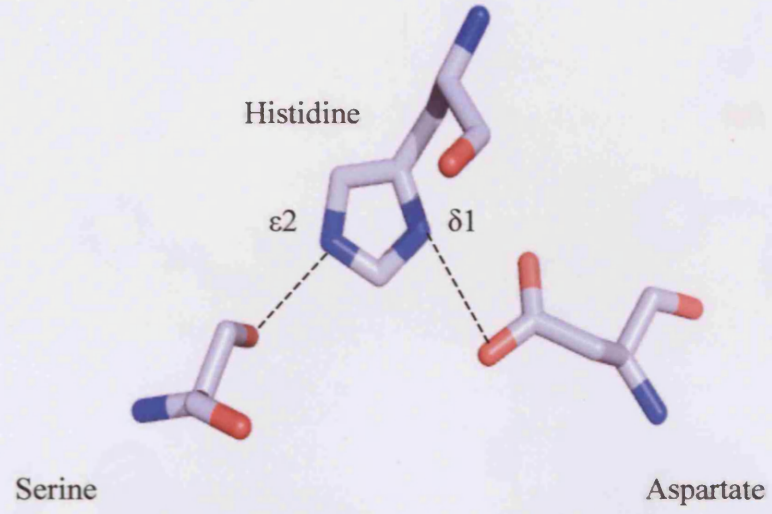


Figure 1.3 Structural arrangement of the catalytic triad of serine proteases. The serine, histidine and aspartate residues from a representative of the serine proteases (BPN') are shown. Hydrogen bonds are shown as dotted lines.

Table 1.1 Order of catalytic residues in the primary structure within serine protease families.

Serine protease family	Catalytic residues
Chymotrypsin	His-Asp-Ser
Subtilisin	Asp-His-Ser
Caboxypeptidase Y	Ser-Asp-His
Clp protease	Ser-His-Asp

This display of convergent evolution is highlighted by the comparison of chymotrypsin and subtilisin, whose structural folds are very different but the arrangement of active site residues is almost identical (Figure 1.4) (Dodson and Wlodawer 1998). Despite these two protease families being structurally distinct, they

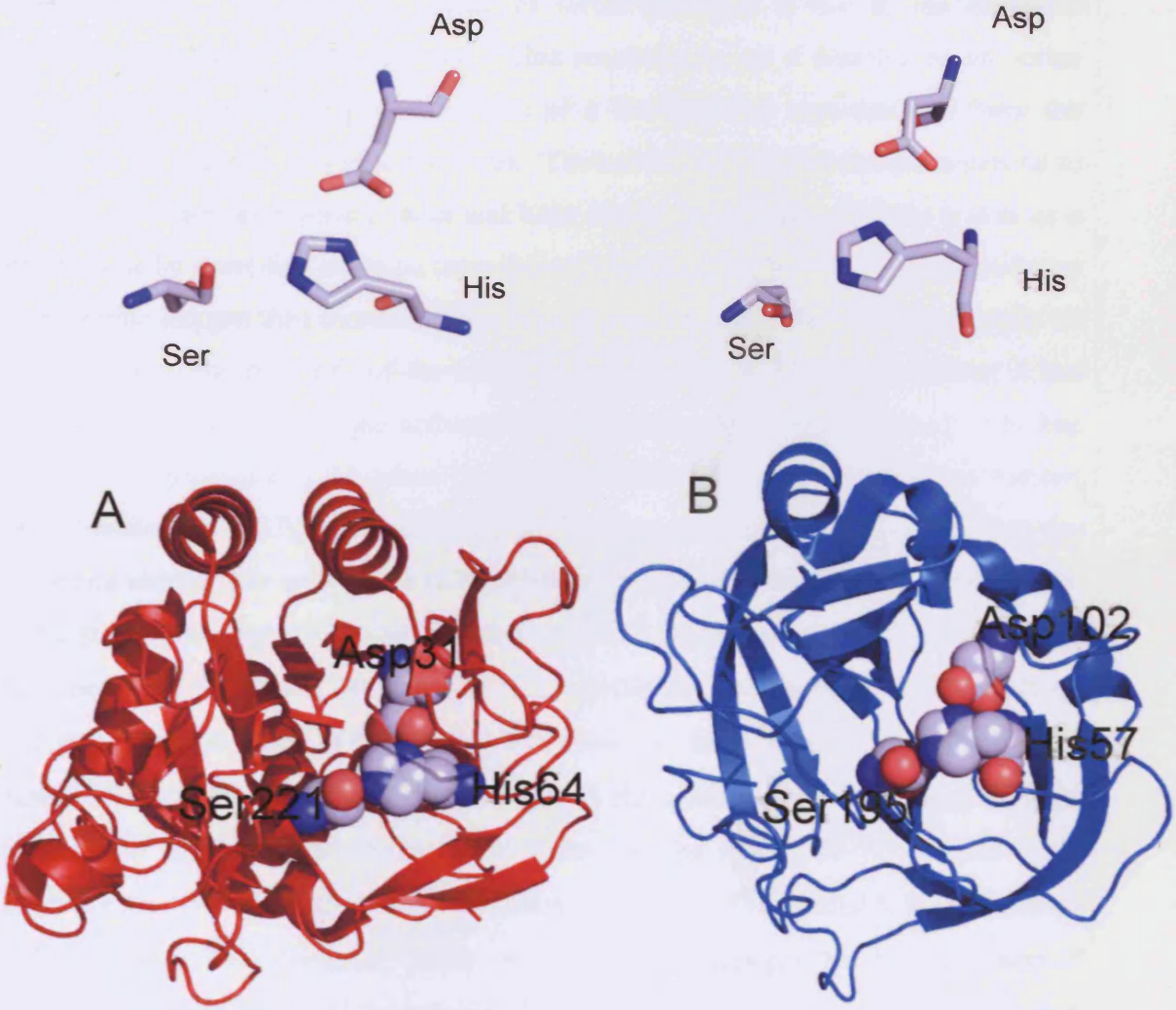


Figure 1.4 Structure and catalytic triad of serine proteases. (A) subtilisin BPN' from *Bacillus amyloliquefaciens* (PDB code 1TO2) (red) with catalytic residues shown above and (B) Trypsin from *Streptomyces griseus* (PDB code 1SGT) with catalytic residues shown above (cyan).

both use a common mechanism to hydrolyse peptide bonds in a two-step process: (1) The initial nucleophilic attack on the peptide bond and (2) the creation of a covalently linked intermediate that is more susceptible to nucleophilic attack by a deprotonated water molecule. The catalytic ability of serine proteases is due to the enhanced reactivity of a specific serine residue. This reactivity is not a function of all serine residues but is a result of the orientation of a histidine and aspartate that form the catalytic triad found in all serine proteases. The active site histidine residue is pivotal to catalysis as it acts as a general acid and base during hydrolysis. Initially it acts as a general base by accepting a proton from the active site serine, leaving a negative charge on the serine oxygen thus increasing its nucleophilic potential. This would normally be unlikely due to the high pK_a of the serine hydroxyl group (~13-15). However, it has been shown that the pK_a of the active site histidine rises from approximately 7 in free enzyme to approximately 11 when the enzyme binds a substrate (Harris and Turner, 2002; Malthouse, 2007), thus increasing the likelihood of accepting a proton from the active site serine. The orientation of the histidine is central to its role as a general base. Histidine residues can exist in two tautomeric forms: one with hydrogen on the $N^{\delta 1}$ and the other with hydrogen on the $N^{\varepsilon 2}$. The dominant tautomeric form consists of hydrogen on the $N^{\varepsilon 2}$ due to its higher pK_a. However, in serine proteases, it has been shown that the hydrogen bond formed between the aspartate carboxyl group and the histidine $N^{\delta 1}$ hydrogen orientates the histidine into its more rare $N^{\delta 1}$-H tautomeric conformation. Therefore, the lone pair of electrons on the $N^{\varepsilon 2}$ is available to deprotonate the active site serine (Day *et al.*, 2003). It was originally thought that the orientation of the catalytic residues created the potential for a charge relay system where a proton was transferred from the active site serine to the aspartate via the histidine thus transferring the charge on the aspartate to the serine (Blow *et al.*, 1969; Hunkapiller *et al.*, 1976). However, NMR analysis revealed that the proton remains bound to $N^{\delta 1}$ of the histidine during catalysis suggesting that the histidine accepts a proton from the serine but does not donate one to the aspartate. Alternatively the pK_a of histidine increases as the dielectric constant of the solvent decreases. Therefore, on substrate binding the exclusion of water decreases the dielectric constant and allows the full electrostatic interaction with the aspartate to increase the pK_a of the histidine (Malthouse, 2007; Shokhen *et al.*, 2007; Shokhen *et al.*, 2008).

The mechanism by which serine proteases catalyse the cleavage of amide bonds (Figure 1.5) begins with substrate binding and deprotonation of the serine hydroxyl by histidine (as described earlier). Nucleophilic attack on the carbonyl carbon of the target peptide bond by the deprotonated serine hydroxyl oxygen, results in the formation of a covalent tetrahedral transition state. The charge developed on the histidine is stabilised by hydrogen bonding to the negatively charged aspartate. The inherently unstable tetrahedral intermediate formed, bears a negative charge on the O atom derived from the carbonyl group. This charge is stabilised by a hydrogen-bonding pocket consisting of NH groups from surrounding main chain residues (trypsin-like) or the side chain (subtilisin) of an asparagine in the oxyanion hole. The importance of the oxyanion binding site was shown with site directed mutagenesis of the asparagine in subtilisin BPN', which resulted in a 10^2 decrease in k_{cat} (Bryan *et al.*, 1986). The reaction continues with a proton being transferred from the positively charged histidine, acting now as a general acid, to the peptide bond amide nitrogen creating a protonated amine. The negative charge on the peptide oxygen is not stable and breaks down forming a more stable acyl-enzyme intermediate. The amine product from the leaving peptide segment is released via steric repulsion between the amine product and carbonyl group of the acyl intermediate. A water molecule is recruited as the second nucleophile and following removal of a proton, by histidine acting as a general base, carries out nucleophilic attack on the carbonyl carbon of the acyl intermediate. This creates a further transient tetrahedral intermediate. The positive charge developed on the histidine is stabilised by hydrogen bonding with the aspartate residue. The collapse of the tetrahedral intermediate is associated with donation of a proton from histidine to the serine oxygen and release of the remaining peptide fragment, which releases the peptide and restores the enzyme to its original state.

The importance of each catalytic residue was investigated by Carter and Wells (1988) in one of the first and seminal examples in the use of protein engineering to dissect enzyme catalysis. Site directed mutagenesis was used to mutate the serine, histidine or aspartate (catalytic residues), of subtilisin BPN' from *Bacillus amyloliquefaciens*, to an alanine and measure changes in enzymatic hydrolysis of a peptide substrate compared to that of the wild-type enzyme and uncatalysed hydrolysis. The total rate enhancement when comparing catalysed versus uncatalysed was 10^9-10^{10} illustrating the power of enzymes to catalyse the hydrolysis of peptide bonds on a biologically relevant timescale.

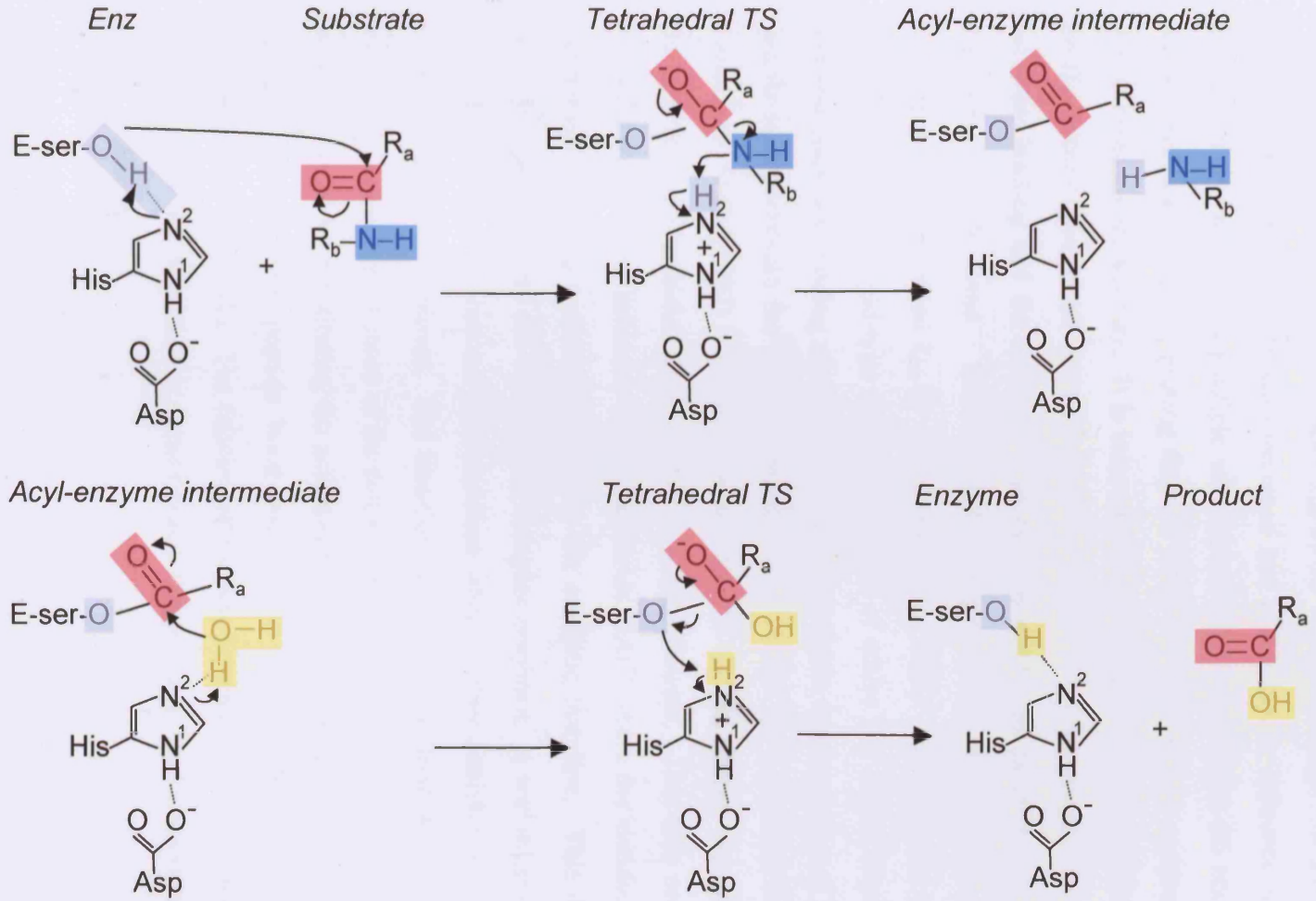


Figure 1.5 The reaction mechanism for serine proteases. The reaction begins with nucleophilic attack on the target peptide bond by the serine hydroxyl oxygen. The resulting tetrahedral intermediate collapses leaving an acyl-intermediate with the expulsion of the amine leaving group. An activated water molecule attacks the acyl-intermediate creating further transient tetrahedral intermediate that collapses thus releasing the serine and carboxylic product.

The substitution of the serine residue resulted in a decrease in the turnover number (k_{cat}) by 10^6. A similar decrease (10^6) was also observed with histidine substitution, revealing the fundamental importance of these residues to the catalytic mechanism. Substitution of the aspartate decreased k_{cat} by 10^4 which, although less than the serine or histidine, represents a significant decrease and highlights its influence on hydrolysis. These data also reveal that multiple substitutions of these residues did not result in any further decrease in k_{cat}, suggesting that the three residues act in a cooperative manner that is essential for activity. It is interesting to note that the residual activity observed on the substitution of all three catalytic residues (10^3) over uncatalysed suggest that substrate binding and the associated interactions contribute more than just orientating the target peptide bond. Site directed mutagenesis of the aspartate to a neutral asparagine also decreased k_{cat} by 10^4 (Craik *et al.*, 1987). The structure of this mutant was identical to the wild-type structure in terms of active site serine, oxyanion binding site and substrate binding site; however, the tautomeric conformation of histidine was unable to deprotonate the serine (Sprang *et al.*, 1987). In a previous experiment by Carter and Wells (1987) activity of the histidine to alanine mutant was recovered with the addition of a histidine at position P2 of the substrate. This was one of the first studies demonstrating substrate-assisted catalysis (SAC) where the histidine supplied by the substrate acts in a similar manner to the catalytic histidine. This discovery had wider implications into the evolution of complex enzymes. It was suggested that SAC might represent an evolutionary intermediate where further catalytic residues are added increasing substrate turnover. The function of the catalytic triad has also been altered by substitution of one or more of the active site residues. For example, subtilisin BPN' was engineered by substituting the active site serine to a cysteine. The function of this enzyme changed from peptide bond cleavage to catalyse the ligation of peptides. (Abrehmsen *et al.*, 1991). The ligation efficiency, however, is limited and only ligates ester peptides but nevertheless demonstrates the plasticity and evolutionally potential within proteases.

1.4 Subtilisin family

The subtilisin family of peptidases, initially named after the subtilisin protease from *Bacillus subtilis*, is comprised of six subfamilies (subtilisin, thermitase, proteinase K, lantibiotic, kexin and pyrolysin) comprised of over 200 proteases being found in bacteria, archaea, eukarya and viruses (Rawlings, 1994; Siezen and Leunissen, 1997). The subfamilies are divided due to their functional role and origin. The term subtilase encompasses proteases from all subfamilies whereas subtilisin refers only to the bacterial subtilisins. The subtilase proteases are predominately secreted out of the cell and have a number of biological roles. For example the proprotein convertase Furin, which belongs to the kexin subfamily and found in humans, is involved in the processing and bioactivation of many precursor proteins (Van de Ven *et al.*, 1993). Also part of the kexin subfamily, the proprotein convertases PC1 and PC2 are fundamental in the specific post-translational processing of hormones within the endocrine and secretory pathways in eukaryotes (Rockwell and Fuller 2002, Rockwell and Thorner 2004). Subtilases also act as virulence factors; the AB5 subtilase toxin expressed in the Shiga toxigenic strains of *Escherichia coli* has a subtilase subunit that has been shown to cleave BiP protein triggering cell death (Paton *et al.*, 2006). Pf-SUB1 found in Plasmodium species has roles in merozoite formation and invasion of the red blood cell in the malaria parasite life cycle (Withers-Martinez *et al.*, 2004). Pf-SUB1 is highly expressed during the sexual reproductive phase of the parasite and thought to be involved in protein processing analogous to the proprotein convertases.

1.4.1 Extracellular subtilisin proteases

The bacterial subtilisins form the largest group within the subtilase family and are normally associated with Gram-positive Bacillus or Clostridium species. The majority have low substrate specificity and are secreted out of the cell to degrade proteins for defence and nutrition and are therefore called extracellular subtilisin proteases (ESPs). The ESPs are amongst the most extensively studied of all the peptidases that include BPN' from *Bacillus amyloliquefaciens*, savinase from *Bacillus lentus* and subtilisin E from *Bacillus subtilis*. The ESPs have provided an important model system for understanding the mechanism of enzyme action, protein structure-function relationships and protein folding (Bryan *et al.*, 1987; Bryan *et al.*, 2002; Eder

and Fersht, 1995; Wells and Estell, 1988). Their broad substrate specificity suits their role as scavenging proteases and their general robustness to harsh environments coupled with extensive protein engineering has resulted in their exploitation by industry for such applications as active ingredients in laundry detergents.

1.4.2 Intracellular subtilisin proteases

The bacterial intracellular subtilisins (ISPs) are distinctive class of the subtilisin family as they are the only known subtilisin to work exclusively within cell cytosol. They were first discovered in *Bacillus megaterium* (Millet 1971), and have since been identified in many other bacilli and related bacteria (Strongin *et al.*, 1978. Strongin *et al.*, 1979, Sheehan and Switzer 1990, Yamagata and Ichishima 1995, Seizen and Leunissen 1997 and Tsuchiya *et al.*, 1997) including *Bacillus anthracis* (Read *et al.*, 2003) and *Clostridium difficile* (Sebaihia *et al.*, 2006). The ISPs are considered to be the main component of the bacilli degradome making up approximately 80% of the intracellular protease activity (Orrego *et al.*, 1973 and Burnett *et al.*, 1986). Despite their central role in proteolysis, little is known about their exact physiological role, mechanism of regulation and 3D structure. It was originally postulated that the ISPs played a major role in sporulation but gene knock out studies revealed that ISP deficient strains still sporulated (Band *et al.*, 1987). One of the only reports into ISP activity was performed by Lee *et al.*, (2004) who incubated ISP with the cellular contents of ISP deficient Bacilli. They studied the decrease in protein bands visualised using two-dimensional electrophoresis. This showed that ISP degraded proteins with a variety of functions including DNA packing, genetic competence and protein secretion. Interestingly ISP did not act to degrade proteins non-specifically, which would be expected from the ESPs, suggesting that there is either a regulating factor in Bacillus or that the specificity of ISPs differ from the ESPs. It is now thought that they may play a role in protein processing when cells enter the late stationary phase (Lee *et al.*, 2004). The ISPs have been isolated in both full length and a processed form (considered to be from the N-terminus) and a lag in activation has been reported of approximately 4 hours. However, it is not known what is responsible for any post-translational modifications and whether this processing is relevant to function or just an artefact of purification (Strongin *et al.*, 1978, Sheehan and Switzer 1990 and Tsuchiya *et al.*, 1997).

1.5 Primary structure of subtilases

The ESPs, ISPs and proprotein convertases are distinct in terms of their primary sequence architecture (Figure 1.6). The mature domains all house the catalytic triad residues. The ESPs and the proprotein convertases have signal sequences that target the proteins to the extracellular and periplasmic locations, respectively. They also have a prodomain immediately N-terminal to the mature region that acts as a chaperone to correctly fold the mature domain to its active conformation. The proprotein convertases contain an additional P domain C-terminal to the mature region involved in protein folding and the stabilisation of the mature domain as well as controlling calcium and pH dependant stability. This domain is also involved in intracellular compartmentalisation and is unique to each protein (Seidah *et al.*, 1999). The ISPs lack a common prodomain or P-domain suggesting that they are not required for the maturation of ISP. There are only sequence similarities surrounding the active site residues when comparing the proprotein convertases and the ESPs and ISPs; there is only 8-10 percent sequence similarity between the mature regions. The ISPs and ESPs are much more closely related. Comparing the primary structures of the ISP and ESPs (Figure 1.7) reveals high homology. For example the ISP show approximately 41 and 47 percent sequence identity with subtilisin BPN' and Savinase, respectively. Sequence alignment of ISP with the N-terminal ESP's prodomain is not shown as no homology is observed. The N-termini of the ISPs shows homology especially at a conserved LIPY(F) sequence. The function of the ISP's N-terminal sequence is unknown. It is too short to act as a classical prodomain, similar to that found in the ESPs, however conservation of the sequence throughout the ISPs suggests either a structural or functional role. The highest homology is observed at the active site residues and importantly the residues immediately adjacent to the active site. This highlights the importance of these residues in the controlling the topology and orientation of the active site residues. There is also homology at the high affinity calcium-binding site found in the ESPs (Figure 1.7, highlighted red) suggesting that the ISPs may bind calcium at a similar position.

1.6 Tertiary structure of subtilisins

The subtilisins are monomers which fold with the aid of a prodomain to form a seven stranded parallel β-sheet core surrounded by nine α-helices (Figure 1.8).

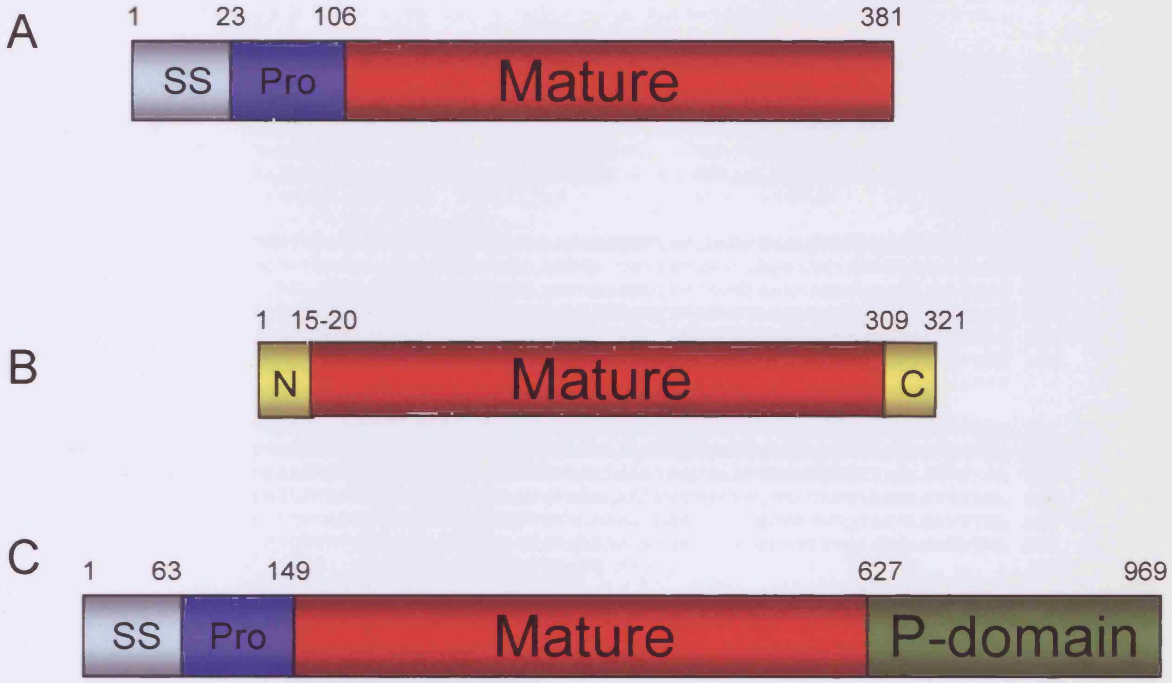


Figure 1.6 Schematic representation of (A) ESP, (B) ISP and (C) Proprotein convertases primary structure. The signal sequence region is coloured grey and labeled SS. The prodomain is coloured purple and labeled Pro and the mature domain is coloured red. The additional P-domain present in the proprotein convertases is coloured green. For ISPs the N- and C-terminal conserved amongst the ISPs are coloured yellow. The representatives for numbering purposes were: Subtilisin from *Bacillus subtilis (*P35835), Intracellular alkaline protease from *Bacillus clausii*, (P29140) and Proprotein convertase subtilisin/Kexin type from *Homo sapiens* P29122 (Protein Sequence Records from SWISS-PROT).

```
ISP-B.clausii        --------MRKFRLIPYKQVDKVSALSEVPMGVEIVEAPAVWKASAKGAGQIIGVIDTGC 52
ISP-B.cereus         MNSSTANKQKGIQLIPFTVDKVVEQVNEIPPGVQLIHAPQVWEKSAKGKDIVVAVLDTGC 60
ISP-B.anthracis      MNSSTANKQKGIQLIPFTVDKVVEQVNEIPPGVQLIHAPQVWEKSAKGKDIVVAVLDTGC 60
ISP-B.subtilis       -------MNGEIRLIPYVTNEQIMDVNELPEGIKVIKAPEMWAKGVKGKNIKVAVLDTGC 53
ESP-BPN              ------------------------AQSVPYGVSQIKAPALHSQGYTGSNVKVAVIDSGI 35
ESP-Savinase         ------------------------AQSVPWGISRVQAPAAHNRGLTGSGVKVAVLDTGI 35
                                            ..:* *:. :.**     . .* .  :.*:*:*

ISP-B.clausii        QVDHPDLAERIIGGVNLTTDYGGVETNFSDNNGHGTHVAGTVAAAETGSGVVGVAPKADL 112
ISP-B.cereus         DMNHIDLKDRIIGGRNFTKDYEGDPNIYLDNNGHGTHVAGTIAATENGVGVLGVAPLAKM 120
ISP-B.anthracis      DMNHIDLKDRIIGGRNFTKDYEGDPNIYLDNNGHGTHVAGTIAATENGVGVLGVAPLAKM 120
ISP-B.subtilis       DTSHPDLKNQIIGGKNFTDDDGGKEDAISDYNGHGTHVAGTIAANDSNGGIAGVAPEASL 113
ESP-BPN              DSSHPDLK--VAGGASMVP---SETNPFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASL 90
ESP-Savinase         S-THPDLN--IRGGASFVP---GEPS-TQDGNGHGTHVAGTIAALNNSIGVLGVAPSAEL 88
                     .  * **   : ** .:.    .      * *.********:** :.. *: **** *.:

ISP-B.clausii        FIIKALSGD-GSGEMGWIAKAIRYAVDWRGPKGEQMRIITMSLGGPTDSEELHDAVKYAV 171
ISP-B.cereus         LVLKVLAGD-GSGSYEQIIEAIHYAVNWRGPNKERVRVISMSLGGPQDVPELHEAIQNAV 179
ISP-B.anthracis      LVLKVLAGD-GSGSYEQIIEAIHYAVNWRGPNKERVRVISMSLGGPQDVPELHEAIQNAV 179
ISP-B.subtilis       LIVKVLGGENGSGQYEWIINGINYAVEQK------VDIISMSLGGPSDVPELKEAVKNAV 167
ESP-BPN              YAVKVLGAD-GSGQYSWIINGIEWAIANN------MDVINMSLGGPSGSAALKAAVDKAV 143
ESP-Savinase         YAVKVLGAS-GSGSVSSIAQGLEWAGNNG------MHVANLSLGSPSPSATLEQAVNSAT 141
                       :*.*... ***.   * :.:.:*          : : .:***.*     *. *:. *.

ISP-B.clausii        SNNVSVVCAAGNEGDGREDTNEFAYPAAYNEVIAVGAVDFDLRLSDFPNTNEEIDIVAPG 231
ISP-B.cereus         KQDVLVVCAAGNNGDCNDNTEELDFPGAYSEVIEVGAVNLERKIACFSNSNQEIDLVAPG 239
ISP-B.anthracis      KQDVLVVCAAGNNGDCNDNTEELDFPGAYSEVIEVGAVNLERKIACFSNSNQEIDLVAPG 239
ISP-B.subtilis       KNGVLVVCAAGNEGDGDERTEELSYPAAYNEVIAVGSVSVARELSEFSNANKEIDLVAPG 227
ESP-BPN              ASGVVVVAAAGNEGTSG-SSSTVGYPGKYPSVIAVGAVDSSNQRASFSSVGPELDVMAPG 202
ESP-Savinase         SRGVLVVAASGNSGAG-----SISYPARYANAMAVGATDQNNNRASFSQYGAGLDIVAPG 196
                       .* **.*:**.*        . :*. * ..: **:..   . : *.. .  :*::***

ISP-B.clausii        VGIKSTYLDSGYAELSGTSMAAPHVAGALALIINLAKDAFKRTLSETEICAQLVRRATPI 291
ISP-B.cereus         DEILSTYPEGKYAVLSGTSMATPHVAGALALLIKQCEREYGRKLSEPEIYAQLIKRTVPL 299
ISP-B.anthracis      DEILSTYPEGKYAVLSGTSMATPHVAGALALLIKQCEREYGRKLSEPEIYAQLIKRTVPL 299
ISP-B.subtilis       ENILSTLPNKKYGKLTGTSMAAPHVSGALALIKSYEEESFQRKLSESEVFAQLIRRTLPL 287
ESP-BPN              VSIQSTLPGNKYGAYNGTSMASPHVAGAAALILSK-----HPNWTNTQVRSSLENTTTKL 257
ESP-Savinase         VNVQSTYPGSTYASLNGTSMATPHVAGAAALVKQK-----NPSWSNVQIRNHLKNTATSL 251
                       : **     *.  .*****:***:** **: .        . :: ::   * . :  :

ISP-B.clausii        GFTAQDKGNGFLTLGLVERITGQFTEKGKK-- 321
ISP-B.cereus         GYERTSEGNGLIDL-LKE-------------- 316
ISP-B.anthracis      GYERTSEGNGLIDL-LKE-------------- 316
ISP-B.subtilis       DIAKTLAGNGFLYLTAPDELAEKAEQSHLLTL 319
ESP-BPN              G-DSFYYGKGLINVQAAAQ------------- 275
ESP-Savinase         G-STNLYGSGLVNAEAATR------------- 269
                     .      *.*::
```

Figure 1.7 Amino acid sequence alignment of various ISP and ESPs. Amino acid sequence alignment of several ISPs from *Bacillus clausii*, *Bacillus cereus*, *Bacillus anthracis* and *Bacillus subtilis* together with two ESPs (BPN' from *Bacillus amyloliquefaciens* and Savinase from *Bacillus lentus)*. Active site residues are highlighted yellow and the main ESP metal ion binding residues are coloured red. The conserved ISP LIPY(F) sequence is coloured green. Sequences were aligned using ClustalW program (www.ebi.ac.uk/Tools/clustalw2). Alignment symbols are used for gaps (-), conserved (*), conserved substitutions (.) and semi-conserved substitutions (:).

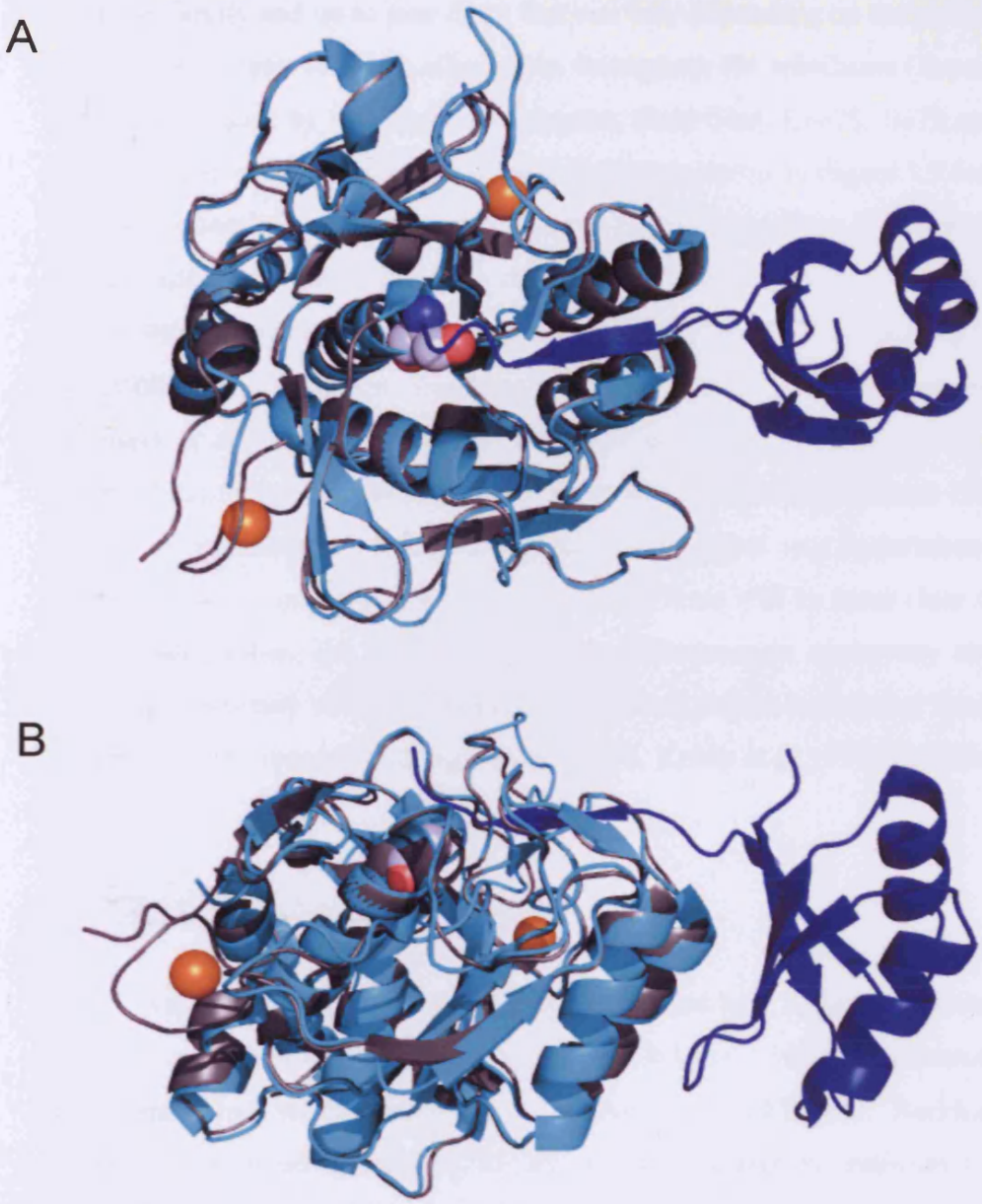


Figure 1.8 Structure of subtilisins. The structure of the savinase from *Bacillus lentus* (PDB 1SVN, coloured grey) and subtilisin BPN' from *Bacillus amyloliquefaciens* (PDB 3CNQ, coloured light blue) in the (A) top and (B) side orientations. The pro domain is coloured blue, the active site serine is shown as spheres and the calcium binding sites are coloured orange.

Structural studies have identified one main calcium-binding site conserved amongst the subtilisin family and up to four more that can vary depending on the protein. The main site is a well-conserved high affinity site throughout the subtilisins (Bryan 2000). This site is coordinated by four carbonyl oxygens, from Gln4, Leu75, Ile79 and Val81, and two side chain oxygens, from Asp41 and Asn77, as shown in Figure 1.9 for BPN'. Four of the coordinating residues form the loop comprising residues 75- 83 and Gln4 is able to coordinate only upon prodomain cleavage. Calcium binding is essential for subtilisin kinetic stability adding approximately 10 kcal/mol to the ΔG of unfolding (Pantoliano *et al.*, 1988). While some studies have suggested that ISP activity is calcium dependent (Strongin *et al.*, 1978, Tsuchiya *et al.*, 1997 and Subbian *et al.*, 2004), others have implied calcium is not critical for these enzymes (Sheehan and Switzer 1990). There is no known structure of an ISP at the time this project was undertaken. While the differences in the primary structure indicate that there will be some clear differences as well as similarities, the may also be some differences in quaternary structure. The quaternary structure of the ISPs is still a matter of debate but several lines of evidence suggest they are dimeric (Strongin *et al.*, 1978, Koide *et al.*, 1976 and Tsuchiya *et al.*, 1997).

1.7 Substrate recognition by subtilisins

Substrate binding in the subtilisins is defined by a series of protein pockets (S), bind the side chains of the peptide substrate (P) (Figure 1.10). The substrate target bond is the centre from which numbering occurs and is labelled P1-P1'. Residues N-terminal to the cleaved bond are termed P1, P2, P3 and P4 and the residues C-terminal are termed P1', P2', P3' and P4' (Schechter and Berger, 1967).

The ESPs are secreted into the surrounding environment to degrade extracellular protein as a defence and nutritional mechanism; they therefore have low substrate specificity. While much is known about the proteolytic activity of the ESPs, very little is known about the ISPs. The subtilisins substrate site normally has sites for P1' to P4 however the main substrate binding sites are S1 and S4, which bind P1 and P4, respectively. In the majority of subtilisins the S1 site forms a large hydrophobic pocket accommodating non-polar residues at P1 (Markland and Smith 1971, Poulos *et al.*, 1972). The S2 site is normally a smaller pocket that may accommodate smaller residues such as alanine or proline. The S3 site is not normally as distinct as the P3

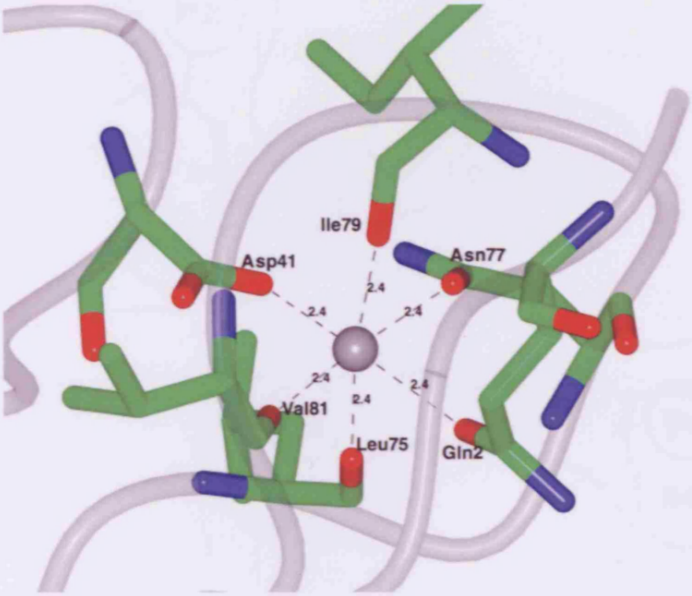


Figure 1.9 Calcium binding to the high affinity site conserved amongst the subtilases. The calcium binding site in BPN' (PDB 1TO2). The coordinating residues are shown in ball and stick. Bonds distances are in Angströms.

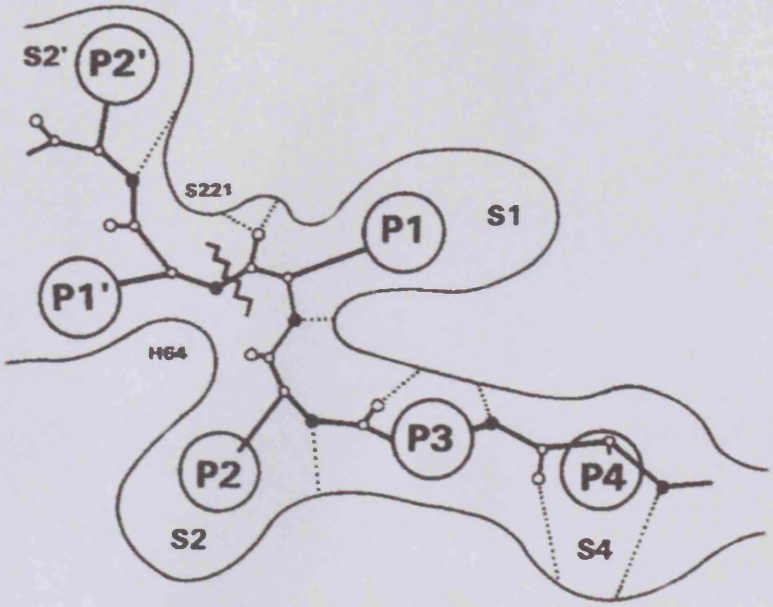


Figure 1.10 Schematic representation of enzymatic hydrolysis nomenclature for subtilases as suggested by Schechter and Berger (1967). Both the enzyme subsites (S) labelled S1, S2, S3 and S4 and substrate peptide (P) labelled P1, P2, P3 and P4 binding sites. These represent residues or subsites N-terminal to the scissile bond, those C-terminal are labelled P1', P2', S1' and S2' (Taken from Siezen and Leuunissen 1997)

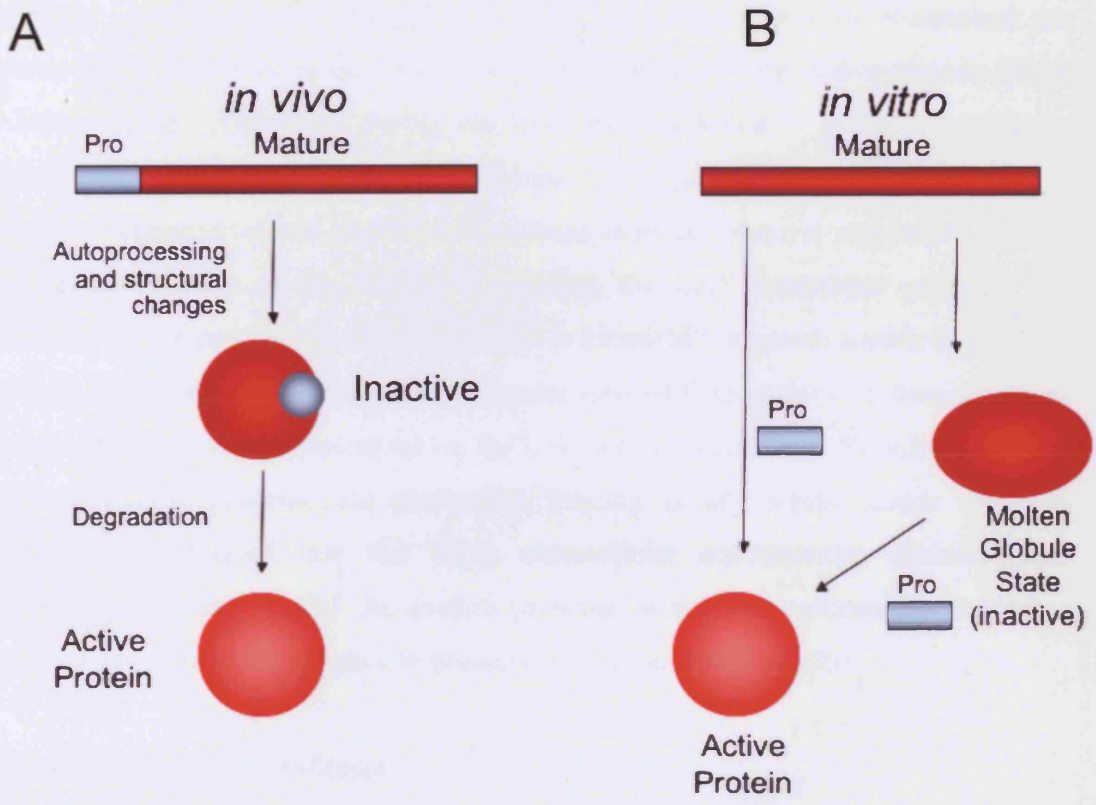


Figure 1.11 Schematic representation of ESP prodomain dependent maturation. Two pathways are shown (A) *In vivo* maturation, where the prodomain and mature domain are expressed together, leading to active protein. (B) *In vitro* where the prodomain is exogenously added to the mature domain. In this pathway the prodmomain can be added when the mature domain is first expressed (mature) or when the mature domain is trapped in partially folded state. Both possibilities have been shown to lead to an active protease. (Adapted from Shinde and Inouye 2000)

ESPs intermediate and native states are unstable without metal ion binding. Since the prodomain is degraded following processing the ESPs are kinetically trapped unable to refold if denatured.

The majority of proteases are expressed as inactive precursors to avoid untimely proteolysis. Inactivation of zymogens is normally performed by an N-terminal pro sequence that is processed in the conversion of the zymogen to the active enzyme (Khan and James 1998). Therefore, during the time the prodomain is associated with the mature domain the protease is directly inhibited. The tight binding of the prodomain to ESP before release is critical for efficient maturation of the protease and causes a delay in activation (Yabuta *et al.*, 2001). Following the ESP maturation pathway and degradation of the prodomain the mature ESP is kinetically trapped, unable to refold if denatured. It has been suggested that prodomain-assisted folding evolved through direct environmental selective pressures set on the ESP during maturation. Regulation of this potent scavenging protease via prodomain binding is key whilst inside the cell, however, once secreted into the harsh extracellular environment, removal and, therefore, kinetic trapping of the mature protease in native conformation increases proteolytic life span and resistance to proteolysis (Subbian *et al.*, 2005).

1.8.2 Inhibition of the subtilisins

The general serine protease inhibitors phenylmethylsulphonyl fluoride (PMSF) and diisopropyl fluorophosphate (DFP) also inhibit subtilisins. PMSF binds to the hydroxyl group of the active site serine forming an inactive sulphonyl enzyme and DFP is partially hydrolysed by the protease but causes inhibition by leaving a phosphate group bound to the active site serine. As the activity of many subtilisins is dependent on calcium binding, they are also inactivated by the addition of the metal ion chelating agents, such as ethylenediaminetetraacetic acid (EDTA). The ESPs are mainly scavenging proteases secreted by Bacilli for non-specific protein degradation. Therefore nature has evolved mechanisms to restrict this activity. Most obvious is the subtilisins own prodomain, as described earlier, which binds tightly in the nanomolar range. Many organisms have evolved proteinacous inhibitors of serine proteases as a defence mechanism. The chymotrypsin inhibitor-2 (CI-2) isolated from barley (Williamson *et al.*, 1987) and Eglin from the leech *Hirudo medicinalis* (Bode *et al.*, 1986), which belong to the CI-2 superfamily of inhibitors (Figure 1.12). These are

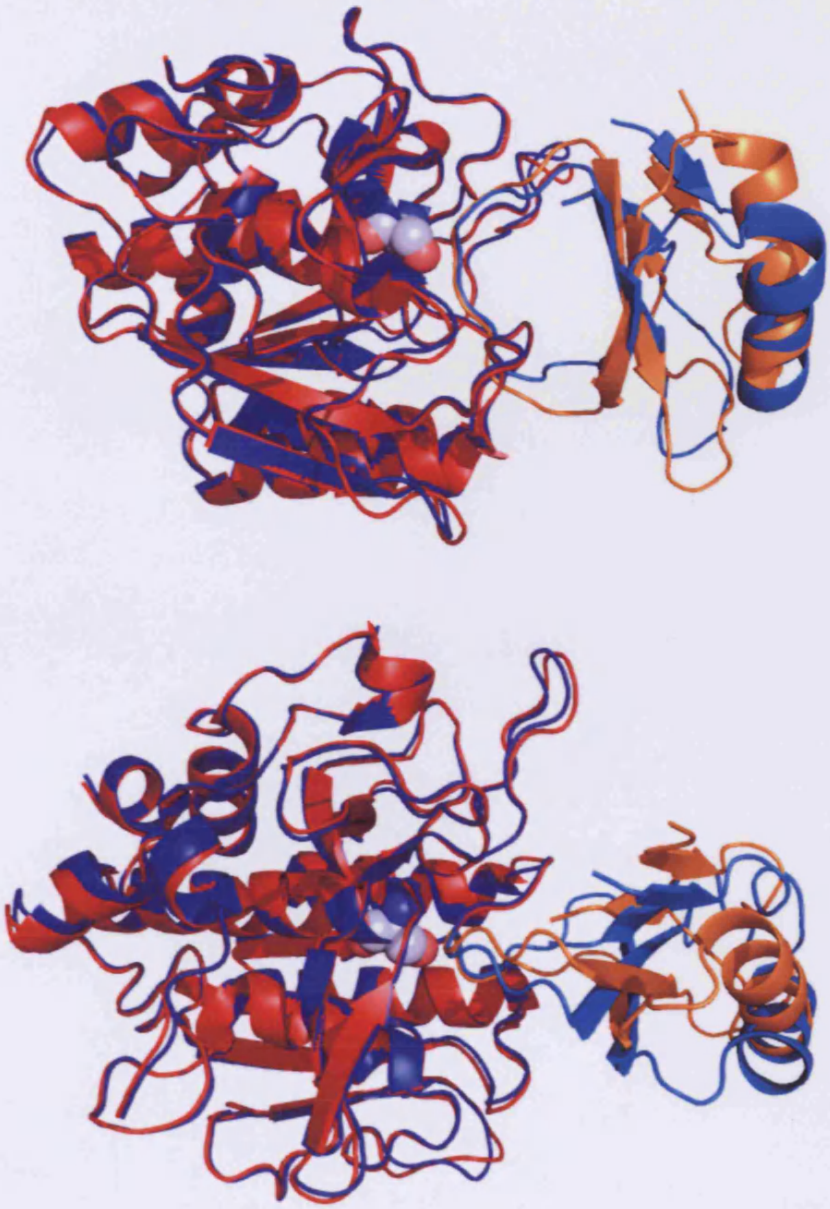


Figure 1.12 Structure of subtilisins with bound inhibitors. The structure of the subtilisin Carlsberg from *Bacillus licheniformis* (blue) with eglin-c from the leech *Hirudo medicinalis* (light blue) (PDB 2SEC) and BPN' from *Bacillus amyloliquefaciens* (red) with CI-2 from barely (orange) (PDB 2SNI). The active site serine of the subtilisins is shown as spheres.

relatively small protein inhibitors (60-90 residues) that contain a binding loop of approximately nine residues that bind directly with subtilisins active site and remain bound upon cleavage. The inhibitors bind to the enzyme with an association constant of up to 10^{14} however their turnover is decreased by between 10^{6}-10^{10} compared to normal protein substrates. It is thought that a network of hydrogen bonds anchors the inhibitor to the enzyme. The inhibitor is cleaved by the protease forming an acyl intermediate that is on pathway but the reaction is halted at this stage of hydrolysis. This is due to the retention of the amine leaving group; the hydrogen-bonding network orientates the peptide towards the nucleophilic attack on the acyl-intermediate and blocks the entrance of a water molecule stopping its deprotonation and use as a nucleophile (Radisky and Koshland 2002).

1.9 Experimental objectives

The aim of this project is to characterise the structure, function and folding of the little studied intracellular subtilisin proteases (ISPs). This will allow a full comparison to be made with their closest relatives, the bacterial ESPs, to elucidate the differences that must exist between these two homologous classes of enzymes that belong to the important subtilisin family of proteases. The ISPs have relatively high homology (40-50 % identity) with the ESPs and are therefore likely to share a common structural fold. However, based on the primary sequence and functional location within the cell it is hypothesised that the activity, structure and folding of ISPs vary from those of the ESPs. Therefore using the ISPs isolated from *Bacillus clausii* as a representative, features of the ISPs will be investigated. One key difference between the ISPs and ESPs is the absence of a classical prodomain. Furthermore the quaternary state of ISP is also still debated; therefore the native molecular weights of all forms of the ISP from *Bacillus clausii* will be determined to elucidate the quaternary structure.

The requirement for calcium, shown with the majority of subtilisins, appears at odds with its intracellular location, as free calcium levels are normally very low within the cell (Herbaud *et al.*, 1998). Therefore structural and functional investigation will be undertaken to assess the importance of calcium to the ISPs.

The exact function of the ISPs is also unknown, despite constituting over 80% of the intracellular proteolytic activity of bacilli bacteria. A non-specific, active protease is likely to cause damage inside the cell by degrading vital components. With respect to

the activity of ISP there are two main facets that need to be addressed: (1) regulation through post-translational modification and (2) regulation through substrate specificity. To address (1) the role of inherent elements of ISP such as the N-terminal extension will be investigated. They do have a short 15-20 amino acid N-terminal extension that is in all known ISPs suggesting it has a pivotal but as yet unknown role. The ISPs are commonly isolated with the N-terminal extension removed but it is unknown whether this usually imprecise processing event is relevant to function or is just an artefact of protein purification. To address (2) proteolytic activity towards various substrates will be investigated and the enzyme kinetics will be determined *in vitro* using a variety of synthetic chromogenic para-nitroanilide (pNA) peptide substrates and compared to existing data available for the bacterial ESPs.

Differences in the primary structure between the ISPs and ESPs, especially the absence of the prodomain in ISPs, suggest that the two homologous proteins will have very different folding properties, contrary to existing hypotheses that state proteins overall structural topology rather than fine sequence detail dictates protein folding pathways and mechanisms. To address this issue, the protein folding properties of ISP will be investigated.

Finally, the 3D structure of ISP is currently unknown. Such structural information will be invaluable in determining how the novel sequence elements of ISP, unique amongst the subtilisins, translate in tertiary and quaternary structures. The full structural detail will also provide a molecular basis behind the structural, functional and folding differences between ISPs and ESPs.

1.10 Hypotheses

I. Novel sequence elements of ISP translate into important structural, functional and folding features, unique amongst the subtilases

II. Proteolytic activity can be regulated post-translationally.

III. ISPs are involved in protein processing via proteolysis of hydrophobic regions of unfolded proteins.

IV. The ISP sequence contains the required information to fold in the absence of a pro-domain.

1.11 Experimental Objectives – the following objectives will be performed using ISP from *Bacillus clausii* as a model ISP

(1) Construction of ISP clones, such as wild-type ISP and an inactivating serine 250 to alanine mutant.

(2) Characteristics of ISP, such as quaternary structure, activity and substrate specificity.

(3) Folding properties of ISP.

(4) Generate material to allow 3D structure of ISP to be determined by X-ray crystallography by the Wilson group at the University of York.

(5) Factors involved in post-translational control of ISP protease activity, such as N-terminal processing.

(6) Structural, functional and folding differences between the ISP and the ESP by comparing the 3D structures, protease characteristics and folding properties.

Chapter 2
Materials and Methods

2. Materials and Methods

2.1 Materials

2.1.1 *Escherichia coli* bacterial host strains.

Escherichia coli (*E. coli*) host strain DH5α (Novagen, Darmstadt, Germany) were used during gene cloning and mutagenesis. Protein expression using the T7 based promoter plasmid system was performed in *E. coli* host strain BL21(DE3) Gold cells (Stratagene, La Jolla, California, USA). The genotypes are shown in Table 2.1

Strain	Genotype
DH5α	F'phi80dlacZ delta(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (rk-, m k+) phoA supE44 lambda-thi-1 gyrA96 relA1/F' proAB+ lacIqZdeltaM15 Tn10(tetr)
BL21(DE3) Gold	F^-, ompT, $hsdS_\beta$ ($r_\beta^- m_\beta^-$), gal, dcm, λDE3

Table 2.1 Genotypes of *E. coli* DH5α and BL21(DE3) Gold bacterial strains.

2.1.2 Media

Luria-Bertani (LB) broth medium and agar plates were prepared according to Sambrook (1989) using reagents from Sigma-Aldrich, Poole, Dorset, UK. LB agar consisted of tryptone (10 g/l), yeast extract (5 g/l), sodium chloride (5 g/l) and agar (15 g/l). LB broth consisted of tryptone (10 g/l), yeast extract (5 g/l), sodium chloride (5 g/l) Reagents were dissolved distilled water and autoclaved for 20 minutes at 121°C and stored at 4°C.

SOC medium (Sigma-Aldrich, Poole, Dorset, UK) contained tryptone (20 g/l), yeast extract (5 g/l) and sodium chloride (0.5 g/l). The solution was autoclaved for 20 minutes at 121°C and sterile glucose (Sigma-Aldrich, Poole, Dorset, UK) added to 1% (w/v).

The antibiotic ampicillin (Sigma-Aldrich, Poole, Dorset, UK) was prepared as a stock (100 mg/ml), filtered using a sterile syringe filter unit (0.22 μm) (Fisher scientific, Loughborough, Leicestershire, UK) and stored at -20°C

2.1.3 Molecular weight standard markers.

The 1 kb DNA molecular standard and protein prestained broad range standard were obtained from New England Biolabs (Hitchen, Hertfordshire, UK). The protein marker was boiled for 5 minutes before use and both protein and DNA markers were stored at -20°C.

A

Protein	MW (Da)
MBP-β-galactosidase	175,000
MBP-paramyosin	83,000
MBP-CBD	62,000
Aldolase	47,500
Triosephosphate isomerase	32,500
β-Lactoglobulin A	25,000
Lysozyme	16,500
Aprotinin	6,500

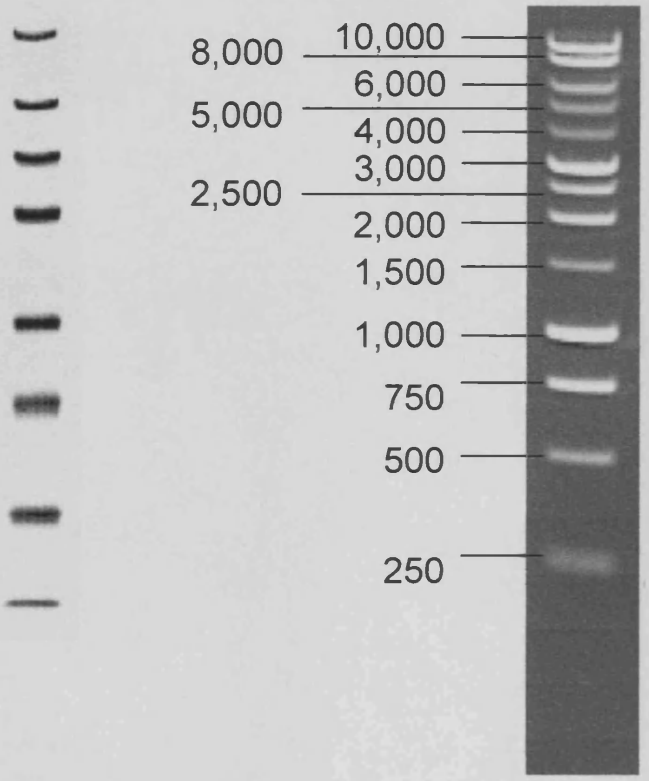


Figure 2.1. DNA and Protein molecular weight standards. (**A**) Prestained broad range protein marker (New England Biolabs). (**B**) 1 kb DNA ladder (New England Biolabs).

2.1.4 Antibodies

The primary antibody mouse anti-Penta Histidine IgG and secondary antibody alkaline phosphatase conjugated rabbit anti-mouse IgG were purchased from Qiagen, West Sussex, UK. The alkaline phosphatase detection kit was from Bio-Rad, Hemel Hempstead, Hertfordshire, UK.

2.1.5 Peptides

The chromogenic *para*-nitroanilide peptides (FAAF, AAPF, AAPNle, AAPNva, AAPL, AAPM, AAVA, AAPK, AAPOrn, AAPE and YVAD) used for protease activity assays were purchased from Sigma-Aldrich, Poole, Dorset, UK or Bachem, Bubendorf, Switzerland. Peptides were dissolved in either water or DMF and stored at -20°C.

The ISP N-terminal 17 amino-acid peptide was synthesised by Activotec, Comberton, Cambridge, UK. The purity was >90% determined by HPLC. The peptide was dissolved in water and stored at -20°C.

2.1.6 DNA manipulation reagents and kits

The following enzymes were supplied by New England Biolabs, Hitchin, Hertfordshire, UK; Phusion high fidelity DNA polymerase, the restriction endonuclease enzymes *Nde*I and *Xho*I and the Quick T4 ligase enzyme. The DNA polymerase *Taq* DNA polymerase was from Promega Ltd, Southampton, UK. The Qiaquick PCR purification and gel extraction kits were from Qiagen, West Sussex, UK.

2.1.7 Chromatography and electrophoresis reagents

The reagents used for SDS-PAGE, Acylamide/bisacrylamide, SDS and ammonium persulphate, TEMED (Tetramethylethylenediamine) were purchased from Sigma-Aldrich (Poole, Dorset, UK) or Melford (Ipswich, Suffolk, UK).

The chromatography columns Superdex 200 10/300 GL, 5 mL HisTrap HP and His SpinTrap were purchased from GE Healthcare (Buckinghamshire, UK).

2.1.8 Chemicals

Distilled or MilliQ water was used to resuspend all reagents and in all buffers. General-purpose reagents were obtained from Sigma-Aldrich (Poole, Dorset, UK) or Melford (Ipswich, Suffolk, UK) and were of analytical grade or better.

2.2 Molecular biology and recombinant DNA Methods

2.2.1 DNA amplification using the polymerase chain reaction (PCR).

Target DNA was amplified by PCR using Phusion high fidelity DNA polymerase (1 U) (New England Biolabs, Hitchin, Hertfordshire, UK) in buffer containing 10 mM Tris-HCl, pH 8.8, 50 mM KCl, 2 mM $MgCl_2$ and 0.1% (v/v) Triton®X-100. Each reaction also contained 0.2 mM dNTPs and 2 μM of each primer, made up to a total volume of 50 μl using distilled water. The thermo cycling was performed using a Techne TC-412 Flexigene thermocycler (Techne, Staffordshire, UK).

Analysis of clones by colony PCR was performed using bacterial cells as the source of template DNA. PCR was performed using *Taq* DNA polymerase (Promega Ltd, Southampton, UK) (2.5 U) (in buffer containing 50 mM KCl, 10 mM Tris-HCl pH 9.0 at 25 °C, 0.1% (v/v) Triton X-100 and 1.5 mM $MgCl_2$) as described above except the initial denaturation occurred at 95°C for 5 minutes.

2.2.2 Submarine agarose gel electrophoresis.

DNA separation was performed using submarine agarose gel electrophoresis. Agarose (1-2% w/v) was suspended in TAE buffer (40 mM Tris-acetate pH 8.5 and 2 mM EDTA) supplemented with ethidium bromide (0.5 μg/ml). Loading buffer (2.5% (v/v) Ficoll®400, 0.01% (w/v) bromophenol blue, 0.01% (w/v) xylene cyanol FF and 5 mM EDTA) was added to each sample and the DNA fragments were separated by electrophoresis in TAE buffer at 100-140 V for 30-50 minutes. The DNA banding pattern was analysed using an uv-transilluminator and approximate molecular weights were determined using DNA molecular size standards (Figure 2.1).

When required relevant DNA or plasmid bands were excised from the gel using a sharp scalpel and purified using a QiaQuick gel extraction kit (Qiagen, West Sussex, UK) according to the manufacturer's instructions. DNA was eluted from the spin column using 50 μl of elution buffer (10 mM Tris-HCl pH 8.5).

2.2.3 Cloning in pET vector

Template DNA and target plasmid were digested using the restriction endonuclease enzymes *Nde*I and *Xho*I. Approximately 50-150 ng of template DNA was

digested using 20 U of the appropriate enzyme in a reaction buffer containing 20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate and 1 mM dithiothreitol (pH 7.9 at 25°C) supplemented with 0.25 ng of bovine serum albumin. The reaction was incubated at 37°C for 2 hours. Digested products were visualised using agarose gel electrophoresis and relevant bands excised from the gel and purified (Section 2.2.2).

Ligations were performed using the Quick T4 ligase enzyme, according to the manufacturers guidelines. Briefly, insert DNA and plasmid digested with the appropriate restriction enzymes, as described above, were added to the ligation reaction mixture in a 1:3 ratio. The ligation reaction mixture contained 2× Quick ligase reaction buffer consisting of 132 mM Tris-HCl, 20 mM $MgCl_2$, 2 mM dithiothreitol, 2 mM ATP, 15% polyethylene glycol (PEG 6000) pH 7.6 at 25°C and diluted to a final volume of 10µl in distilled water. Quick T4 DNA Ligase enzyme (1 U) was added to the reaction and mixed thoroughly. The reaction was briefly centrifuged and incubated at room temperature for 10 minutes. The ligation was then used to transform bacterial cells (Section 2.2.6).

2.2.4 DNA sequencing

DNA samples (10µl) generated by PCR were submitted for sequencing at a concentration of approximately 5 ng/µl. DNA sequencing was performed at the DNA Sequencing Core of the Molecular Biology Unit at Cardiff University. Sequencing was performed using an ABI Prism 3100 16 capillary genetic analyzer with Big Dye Terminator version 3.1 cycle sequencing kit (ABI, Foster City, California, USA) and pellet paint (Novogen Ltd, Windsor, Berkshire, UK) for co-precipitation of DNA.

2.2.5 Preparation of electro-competent *E. coli* cells.

Electro-competent *E. coli* bacterial strains BL21(DE3) Gold and DH5α were made electrocompetent using the following procedure: A 50 ml starter culture was inoculated using a single colony picked from an appropriate agar plate and was grown aerobically at 37°C overnight. The following day the starter culture was used to inoculate 2 × 500 ml of pre-warmed Luria-Bertani (LB) broth medium (Section 2.2.2) and grown at 37 °C to an attenuance of 0.4. The cells were cooled, transferred to pre-

cooled centrifuge bottles and harvested by centrifugation (1500 × g for 15 min at 4°C). The supernatant was discarded and the cells resuspended in 500 ml of ice-cold water and centrifuged (1500 × *g* for 20 min at 4°C). The cells were resuspended in 250 ml of ice-cold 10% glycerol and centrifuged as before (1500 × g for 20 min at 4°C). Supernatant was discarded and the cells were resuspended in 10 ml ice-cold 10% glycerol. The two preparations were pooled and the cells harvested by centrifugation (1500 × *g* for 20 min at 4°C). Supernatant was removed and the cells resuspended in 1 ml SOC (Section 2.1.2). To ensure high transformation efficiency the cells were diluted to a concentration of between 2×10^{10} to 3×10^{10} cfu/ml and flash frozen in liquid nitrogen in 40 µl fractions and stored at -80°C.

2.2.6 Transformation of *E. coli* cells by electroporation.

A fraction (40 µl) of appropriate electrocompetent *E.coli* cells were gently thawed on ice and ~1 ng of plasmid or ligation mixture was added, gently mixed and transferred to a pre-cooled electroporation cuvette (Thermo Fisher Scientific, Loughborough, Leicestershire, UK). The cells were pulsed with 2.5 kV, with a capacitance of 25 µF and a resistance of 200 Ω, using a gene pulser (Bio-Rad, Hertfordshire, UK). The cells were rescued by adding 500 µl SOC medium and were incubated at 37°C for 1 hour with shaking at 225 rpm. Following incubation, various volumes (from 5-250 µl) were spread onto warmed agar plates, supplemented with ampicillin (100 µg/ml) and grown overnight at 37°C.

2.3 Protein preparation and analysis

2.3.1 Expression of recombinant proteins

An LB broth starter culture (25 ml) supplemented with ampicillin (100 ug/ml) was inoculated from a single colony and grown overnight at 37°C. The starter culture was used to inoculate 1 l of LB broth supplemented with ampicillin (100 ug/ml), which was subsequently grown at 37°C with shaking to an attenuance of 0.4-0.6. Protein expression was induced by addition of 1 mM IPTG and the culture grown for a further 3 h at 37 °C. Aliquots (1 ml) were taken at one-hour intervals after induction and

protein production monitored using SDS-PAGE (Section 2.3.4). The aliquots were centrifuged to harvest the cells (5000 × *g* for 5 min) and the pellets resuspended in SDS loading buffer to an apparent attenuance of 10 to determine increase of protein production independent of bacterial growth.

After 3 hours, the cells were harvested by centrifugation (5000 × *g* for 15 min) and resuspended in 20 ml of cell resuspension buffer (50 mM Tris-HCl pH 7.5, 0.5 M ammonium sulphate, and 1 mM calcium chloride). Cells were lysed using a French press (Section 2.3.2).

2.3.2 Native lysis of *E. coli* by French press

Bacterial *E. coli* cells were lysed using a French Press (American Instrument Company, Silver spring, Maryland, U.S.A.). The pressure cell was pre-cooled to 4°C and bacterial cells lysed by two passages through the cell at a breaking pressure of 20,000 psi in fractions of 20 ml. The lysate was clarified via centrifugation (48,000 × *g* for 20 min at 4°C) to pellet cell debris.

2.3.3 Nickel affinity purification of His tagged recombinant protein

Recombinant proteins with a hexa-histidine tag at the C-terminus were purified via immobilised metal affinity chromatography (IMAC) using a 5 ml HiTrap IMAC HP Column (GE Healthcare, Buckinghamshire, UK) connected to an Äkta FPLC purifier (GE Healthcare, Buckinghamshire, UK). Up to 10 column volumes (CV) of binding buffer (20 mM sodium phosphate pH 8.0, 0.5 M NaCl and 10 mM imidazole) was used to pre-equilibrate the column. Protein lysate (approximately 5 ml) was loaded onto the column and the column washed with a further 10 CV of binding buffer. Contaminating protein was washed from the column with 20 mM sodium phosphate pH 8.0, 0.5 M NaCl and an imidazole gradient ranging from 2-150 mM applied over 4 CV. Proteins with a hexa-histidine tag were eluted using elution buffer (20 mM sodium phosphate pH 8.0, 0.5 M NaCl and 0.5 M imidazole) over 2.5 CV. Eluted fractions were dialysed against protein buffer (50 mM sodium phosphate pH 8.0, 0.5 M ammonium sulphate, 1 mM calcium chloride) and stored at -20°C. Aliquots representing whole cell lysate, column flow-though (unbound protein), wash stages and final elution were visualised by SDS-PAGE (Section 2.3.4).

2.3.4 Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed, as described by Laemmli (1970), using a mini-PROTEAN III electrophoresis system (Bio-Rad, Hertfordshire, UK). The composition of the resolving and stacking gels are shown in Table 2.3. The sample loading buffer was comprised of 2% (w/v) SDS, 0.2 M Tris-HCl pH 6.8, 0.04% (w/v) bromophenol blue, 8% (w/v) glycerol with β-mercaptoethanol added freshly to 10% (v/v). The gel running buffer contained 28.8 g glycine, 8.0 g Tris base made in 2 l of distilled water, with SDS added freshly to 0.1% (w/v). Electrophoresis was performed at a constant voltage of 200 V for 1 hour and stained using Coomassie blue R-250 blue (0.5% w/v) in 50% (v/v) methanol and 10% (v/v) acetic acid. Gels were subsequently de-stained in 40% (v/v) methanol and 10% (v/v) acetic acid. Protein standards (Figure 2.1) were present on every gel for molecular size comparison.

Type of gel	Component	Concentrations
Resolving	Acrylamide/bis-Acrylamide[1]	10-12.5% (w/v)
	Tris-HCl pH 8.8	375 mM
	SDS	0.1% (w/v)
	APS	0.05% (w/v)
	TEMED	0.02% (w/v)
Stacking	Acrylamide/bis-Acrylamide[1]	5% (w/v)
	Tris-HCl pH 6.8	65 mM
	SDS	0.2% (w/v)
	APS	0.1 % (w/v)
	TEMED	0.02 % (w/v)

Table 2.3 Components and concentrations for the SDS-PAGE resolving and stacking gels. [1] Acrylamide and N,N'-methylene *bis*-acrylamide (40% solution containing a monomer (acrylamide) to crosslinker (*bis*-acrylamide) ratio of 37.5:1 (w/w)).

2.3.5 Tris/Tricine SDS-PAGE: Preparation of protein samples for protein N-terminal sequencing

SDS-PAGE (Section 2.3.4) was performed with the following modifications to the gel buffers:

Type of gel	Component	Concentrations
Resolving	Acrylamide/bis-Acrylamide[1]	10% (w/v)
	Tris-HCl pH 8.45	0.5 M
	SDS	0.1% (w/v)
	APS	0.02% (w/v)
	TEMED	3.75% (w/v)
Stacking	Acrylamide/bis-Acrylamide[1]	0.05% (w/v)
	Tris-HCl pH 8.45	375 mM
	SDS	0.075% (w/v)
	APS	0.1 % (w/v)
	TEMED	0.02 % (w/v)

Table 2.4 Components and concentrations for the Tris/tricine SDS-PAGE resolving and stacking gels. [1] – Acrylamide and N,N'-methylene *bis*-acrylamide (40% solution containing a monomer (acrylamide) to crosslinker (*bis*-acrylamide) ratio of 29:1 (w/w)).

Protein samples were prepared in buffer containing 0.5 M Tris-HCl pH 8.45, 25% glycerol, 8% SDS, 0.025% bromophenol blue and 4% β-mercaptoethanol and boiled at 95°C for 5 minutes. The cathode electrolyte buffer contained 0.05 M tricine, 0.05 M Tris base and SDS added freshly to 0.1% (w/v). The anode electrolyte buffer contained 0.1 M Tris-HCl, pH 8.9. Buffers were precooled to 4°C before use.

Electrophoresis was performed at 50-90 V (<35 mA) until the samples travelled through the stacking gel followed by 50 V (<35 mA) for 3-4 hours. Protein standards (Materials 2.3) were present on every gel for molecular size comparison. Samples were transferred onto PVDF Sequi-Blot membrane (0.2 μm) (Bio-Rad, Hertfordshire, UK) using a BioRad mini-PROTEAN III transfer kit (Bio-Rad, Hertfordshire, UK) (Section 2.3.6) and transfer buffer (25 mM Tris-HCl, 192 mM tricine and 20% methanol). Membranes were stained using Ponceau S stain (Sigma-Aldrich, Dorset, UK). The

membranes were then sent for N-terminal sequence analysis, which was performed by Alta Bioscience, Birmingham University, UK.

2.3.6 Western blotting

Protein samples for analysis were initially separated, in duplicate, by SDS-PAGE (Section 2.3.4). The first gel was stained using Coomassie blue R-250. Proteins on the duplicate gel were transferred to Hybond-ECL nitrocellulose membrane (GE Healthcare, Buckinghamshire, UK) using the BioRad mini-PROTEAN III transfer kit (Bio-Rad, Hertfordshire, UK). Transfer was carried out in ice-cold transfer buffer (25 mM Tris-HCl, 192 mM glycine and 20% methanol) at 100 volts for 1 hour.

The membrane was blocked in powdered milk (5% w/v) dissolved in Tween Tris-buffered saline (TTBS; 50 mM Tris (pH 7.5), 150 mM sodium chloride and 0.1% (v/v) Tween-20 for 1-2 hours. The membrane was then washed in TTBS for 20 minutes. The primary antibody, mouse anti-penta-histidine IgG (Qiagen, West Sussex, UK), was added at a 1/5000 dilution in TTBS and incubated for 1-2 hours and the membrane washed with TTBS for a further 20 minutes. The membrane was then incubated with the secondary antibody, alkaline phosphatase (AP) conjugated rabbit anti-mouse IgG (Qiagen, West Sussex, UK), diluted to 1/5000 in TTBS, for 1-2 hours and the membrane again washed for 20 minutes with TTBS. The Western blot was developed using the AP conjugate substrate kit (Bio-Rad, Hertfordshire, UK). Washing the membrane in distilled water terminated the reaction.

2.3.7 MALDI-TOF Mass spectroscopy

Mass spectroscopy was performed using a MALDI-TOF micro MX (Waters, Milford, Massachusetts, USA) mass spectrometer. Samples were exchanged into water using an Amicon centrifugal filter unit with a 10 kDa molecular weight cut off (Millpore, Watford, Hertfordshire, UK). Each sample was diluted to ~20 μM and 1 μl added to 1 μl of matrix containing sinapinic acid (20mg/mL) resuspended in 70% (v/v) acetonitrile in distilled water with 0.1% (v/v) TFA. This mixture was pipetted onto a MALDI 96-well plate (Waters, Milford, Massachusetts, USA). Samples were analysed in linear mode with positive polarity.

2.3.8 Determination of enzyme kinetic parameters using synthetic chromogenic peptides.

The kinetic parameters K_m, k_{cat} and k_{cat}/K_m were determined using the succinyl-XXXX-*para*-nitroanilide as substrates. The X designation represents the following amino acids sets: FAAF, AAPF, AAPNle, AAPNva, AAPL, AAPM, AAVA, AAPK, AAPOrn, AAPE and YVAD. The concentrations of ISP used for each peptide range are given in Table 2.5.

Initial rates (v) were measured using different concentrations of peptide substrate [S]. All reactions were performed in 50 mM Tris-HCl pH 8.0, 0.5 M ammonium sulphate, 1 mM $CaCl_2$ and 0.025% (v/v) triton X-100. The peptide substrate and reaction buffer were mixed in a quartz cuvette. This was equilibrated to reaction temperature (25°C) before protease was added. Enzymatic activity was determined by measuring the increase in absorbance at 405 nm due to hydrolysis of the peptide bond N-terminal to the pNA moiety using a Varian Cary 50 Bio uv/vis spectrophotometer.

Succinyl-XXXX-pNA	ISP (μM)
FAAF	0.02
AAPF	0.2
AAP Nle, AAP Nva, AAPL, AAPM, AAVA, AAPK, AAPE, AAP Orn and YVAD.	2

Table 2.5 ISP concentrations used for each peptide substrate. All amino acids are represented by the standard single letter abbreviation except for Nle (norleucine), Nva (norvaline) and Orn (ornithine). The X represents various amino acids shown as variables in column one.

The Michaelis-Menten (equation 1), given below, was used to calculate the kinetic parameters K_M and V_{max}. The parameter k_{cat} was calculated using equation 2.

$$v = \frac{V_{max}[S]}{K_M + [S]} \qquad \text{Eq 1}$$

$$k_{cat}=\frac{V_{max}}{[E]} \qquad \text{Eq 2}$$

The initial rates collected as absorbance change per minute were converted into velocity ($\mu M^{-1}min^{-1}$) for each substrate concentration using the molar absorbance coefficient for pNA (9800 $M^{-1}cm^{-1}$ at 405 nm). Data were analysed using GraphPadPrism software.

2.3.9 pH dependent activity assay

Enzyme activity was measured at different pH values from 2-12 using the synthetic chromogenic peptide substrate Suc-FAAF-pNA (100μM). Reaction conditions were altered to buffer in the relevant pH range: pH 2-6 (sodium phosphate/citric acid), 6-7 (sodium phosphate), 8-9 (Tris HCl), 10 (sodium borate/NaOH) and 11-12 (sodium phosphate/NaOH).

2.3.10 Proteolytic activity assay

The proteolytic digestion patterns of ISP-BC and Savinase, a representative of the extracellular serine protease family, were analysed using bovine serum albumin (BSA), malate dehydrogenase (MDH) and lactate dehydrogenase (LDH) (all from Sigma-Aldrich, Poole Dorset, UK) as substrates. The substrate protein was assayed before and after heat treatment at 95°C for 5 minutes. Protease (2 μM) was added to a 38 μg of each substrate in buffer containing 50 mM Tris-HCl pH 8.0, 1 mM calcium chloride and 0.5 M ammonium sulphate. Aliquots were taken at time intervals of 0, 1, 5, 10, 30 and 60 minutes and digestion analysed by SDS-PAGE (Section 2.3.4).

2.3.11 Size exclusion chromatography

Analytical size exclusion chromatography (SEC) was performed using a superdex 200 10/300 GL column (GE Healthcare, Buckinghamshire, UK) connected to an Äkta FPLC purifier (GE Healthcare, Buckinghamshire, UK). Protein elution was determined using the UV 900 detector by measuring the absorbance at 280 nm.

The column was calibrated, at a flow rate of 0.5 ml/min, using the following molecular standards: ribonuclease A (13.7 kDa), chymotrypsinogen A (25 kDa), ovalbumin (43 kDa), albumin (67 kDa), transferrin (81 kDa) and gamma globulin (150 kDa) (Sigma-Aldrich, Dorset, UK). Stocks for each molecular standard (5 mg/ml) were prepared in 50 mM Tris-HCl pH 8.0, 0.2 M ammonium sulphate. Each standard (0.5 mg, 100µl) was added to the column and the elution volume (Ve) was measured from the absorbance peak at 280 nm. The void volume (Vo) and total column volume were determined in a similar manner using Blue dextran (2000 kDa) (0.5 mg) or tryptophan (204 Da) (0.5 mg) respectively.

The partition coefficient (Kav) for each standard was calculated using the following equation:

$$Kav = \frac{(Ve - Vo)}{(Vt - Vo)}$$

A standard curve was generated for the size exclusion column by plotting the partition coefficient (Kav) for each molecular standard against the $\log_{10}$ of the molecular weight. This curve was used to determine the apparent molecular weight of the different ISP samples.

2.3.12 Circular dichroism spectroscopy

Circular dichroism (CD) analysis was performed in a 1 mm quartz cuvette using a Chirascan circular dichroism spectrometer with a PCS.3 single cell peltier temperature controller (Applied photophysics Ltd, Surrey, UK). Protein was diluted to 10 µM in buffer containing 20 mM sodium phosphate pH 8.0, 0.2 M ammonium chloride and 1 mM calcium chloride and centrifuged at 15,000 × g for 10 minutes in a microfuge. Spectra were recorded from 260-190 nm at 1 nm intervals at 25°C.

Secondary structure was deconvoluted using the CDSSTR program Johnson, 1999, http://biochem.science.oregonstate.edu/people/w-curtis-johnson).

2.3.13 Fluorescence spectroscopy

Fluorescence spectroscopy was performed in a 5 mm quartz cuvette using a Cary Eclipse fluorescence spectrophotometer attached to a PCB-150 peltier water bath (Varian Inc. Scientific instruments, Yarnton, Oxford, UK). Emission spectra were measured using an excitation wavelength of either 280 nm or 295 nm. The corresponding emission spectra were monitored between 280-425 nm and 295-425 nm, respectively. Spectra were recorded at 15-30 nm/min with slits widths set at 5 nm and 10 nm (excitation and emission respectively). Protein was diluted to 2 μM in buffer containing 20 mM sodium phosphate pH 8.0, 0.2 M ammonium chloride and 1 mM calcium chloride. Spectra were recorded at 25°C.

2.3.14 Equilibrium unfolding and refolding experiments

The assays were performed in a thermostatically controlled cuvette at 25°C in 20 mM sodium phosphate pH 8.0, 0.2 M ammonium chloride. For urea and guanidinium chloride (GdnHCl) denaturation curves stock solutions of 9 M urea and 8 M GdnHCl were prepared in buffer. Fractions were set up to give final urea and GdnHCl concentrations between 0.5-7 M and 0.1-6 M, respectively. To this ISP^{S250A} was added to a final concentration of either 10 μM or 1 μM depending on analysis by CD or fluorescence, respectively. For urea and guanidinium chloride (GdnHCl) refolding curves stock solutions of urea and GdnHCl were set as previously indicated however second stock was prepared of ISP^{S250A} protein unfolded in either 9 M urea or 8 M GdnHCl. Fractions were set up to give final urea and GdnHCl concentrations between 0.5-7 M and 0.1-6 M, respectively. All fractions were incubated at 5°C over night before analysis. Spectroscopic measurements were taken as described in Section 2.3.12 and 2.3.13.

Data analysis

The denaturation curves were fitted to a two-state model, where applicable, according to Walters *et al.*, 2009. This model assumes that the protein exists only as a native dimer (N_2) and unfolded monomers (2U). (See appendix for summary). Fits of the data sets to the two-state models were performed using KaleidaGraph v3.6.2 (Synergy Software).

Chapter 3
Production of ISP Protein Variants

3. Production of ISP protein variants

3.1 Introduction

The study of the subtilase family has been dominated by the extracellular proteases, which have had a large role in the understanding of key areas of protein structure and function. The subtilisin BPN' was one of the first structures solved by X-ray crystallography (Wright, 1969) and since then, the subtilase family has been used as model systems for studying enzyme mechanisms such as catalytic action, protein structure-function relationships and protein folding (Bryan 2000; Dodson, 1998; Fisher 2007; Hedstrom, 2002). Modern recombinant protein engineering has enabled mutation of over 50% of the amino acids of subtilisin giving insight into regions involved in substrate binding and structural stability (Bryan 2000). In contrast, the intracellular subtilisins, which share approximately 40-50% sequence homology with the ESPs, have received far less interest with no known structure or function.

Therefore modern genetic engineering will be used to generate ISP variants in order to study the structure, function and folding of this novel class of subtilases. The two ISP variants studied in this project will be the wild type ISP from *Bacillus clausii*, used as a representative of ISPs, and a deactivated variant. This ISP protein contains a mutation of the active site serine to an alanine (S250A) and will be used in structural and folding assays where proteolysis may interfere (Figure 3.1).

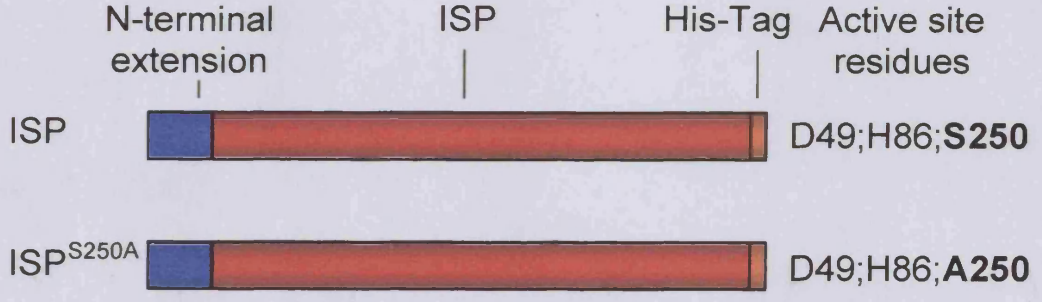


Figure 3.1 Schematic representation of ISP primary structure. The two variants used in this project are shown with the location of the N-terminal (blue), mature ISP domain (red) and HisTag (orange). The active site residues are shown with the deactivating mutation of the serine to alanine at position 250.

3.2 Cloning of ISP variants

The gene encoding ISP from *Bacillus clausii* (strain NN010181) was previously amplified from genomic DNA in the laboratory of Dr Dafydd Jones at Cardiff University. The Expand High Fidelity PCR system (Roche Applied Science) was used with sense primer DDJisp002
(5' GAGGGAAACCATATGAAGAAACCGTTGGGGAAAATTGTCGC 3') and antisense primer DDJisp001
(5' TTTTATCTCGAGTTTTTTGCCTTTCTCAGTAAACTGGC 3'). These primers introduced the restriction sites *Nde*I and *Xho*I shown underlined, respectively. The primers were designed based on the EBI entry D10730 and SwissProt entry P29140. The gene was cloned between the *Nde*I and *Xho*I sites of pET11c (Merck Biosciences). A mutant, termed ISP^{S250A}, was created that substituted the essential catalytic serine to an alanine using splice-by-overlap mutagenesis using the mutagenic primers DDJisp003 (5' GAGTTGTCAGGCACTGCGATGGCCGCGCCCCACGTC 3') and DDJisp004 (5' GTGGGGCGCGGCCATCGCAGTGCCTGACAACTCGGC 3') in conjunction with DDJisp001 or DDJisp002, respectively. The mutation sites are underlined. The resulting DNA fragments where then spliced together by PCR using primers DDJisp001 and DDJisp002. PCR was performed using the Pfu polymerase system (Stratagene).

For this project, ISP and ISP^{S250A} genes were cloned from pET11c into pET22b to incorporate a C-terminal hexa-histidine affinity purification tag to aid purification. The genes were amplified using Phusion high fidelity DNA polymerase (NEB) (Section 2.2.1) with primers that flank the multiple cloning site of pET22b (pET-F 5'ATGCGTCCGGCGTAGAGGA3' and
pET-R 5'CGTAGTTATTGCTCAGCGGTG3'). The resultant PCR products observed (Figure 3.2 A) were of the correct approximate length, calculated to be 1098 bp from the ISP gene sequence. Restriction digestion of the PCR products using the restriction enzymes *Nde*I and *Xho*I removed 132 bp, resulting in PCR digests of 966 bp shown in Figure 3.2 B. The pET22b vector was also digested with the restriction enzymes *Nde*I and *Xho*I to allow ligation of the ISP PCR digests and to remove the pelB leader signal peptidase sequence. Ligation of the PCR digests and vector was performed using Quick T4 ligase enzyme (Section 2.2.3). *E.coli* DH5α bacterial cells were transformed, using electroporation (Section 2.2.6), with plasmids pET22b-ISP and $pET22b\text{-}ISP^{S250A}$. The resultant colonies were screened using colony PCR (Section 2.2.1) using the primers

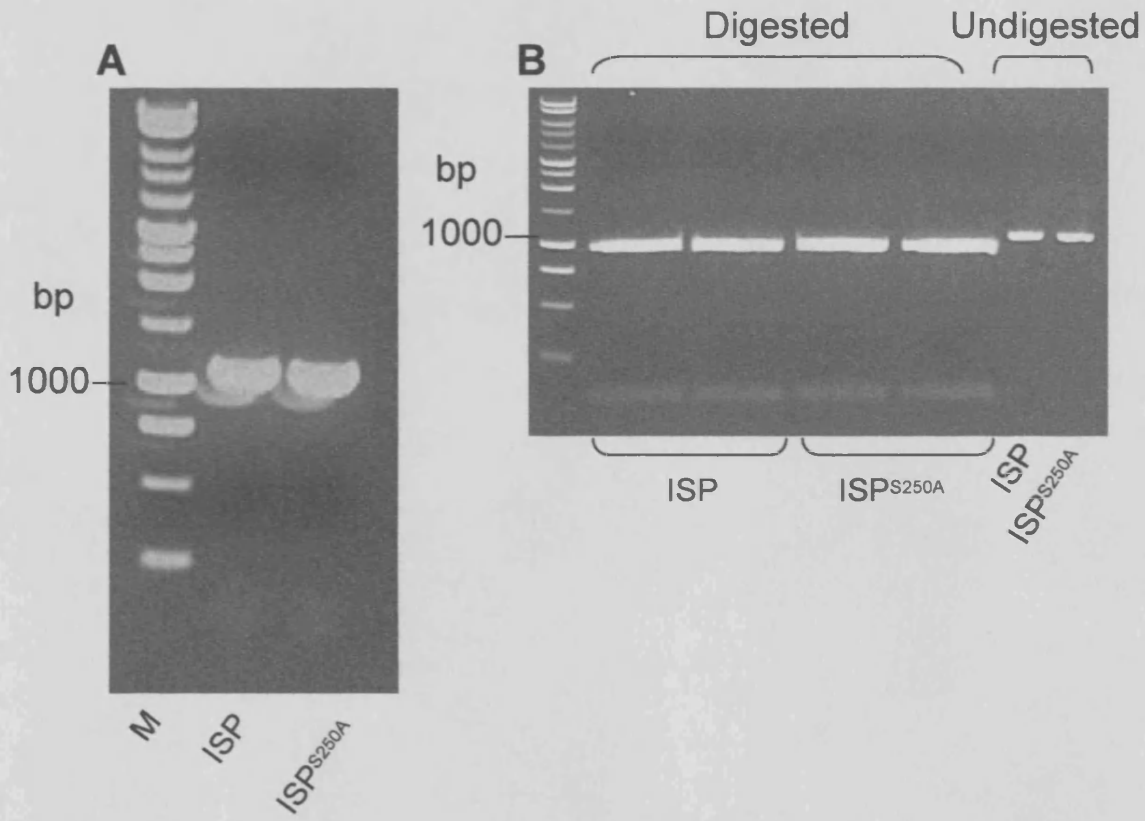


Figure 3.2 Analysis of ISP and ISP^{S250A} PCR amplification and restriction digestion. *(***A)** ISP and ISP^{S250A} genes were amplified from pET11cISP templates. PCR products of the expected size were observed (1098 bp). **(B)** ISP and ISP^{S250A} PCR products were digested with *Nde*I and *Xho*I restriction endonucleases. Restriction digestion removed 132bp from the PCR products to give PCR product of the expected size (966 bp). DNA products were analysed using agarose gel electrophoresis (1% w/v) and for size comparison samples were analysed against the 1 kb DNA marker (M)

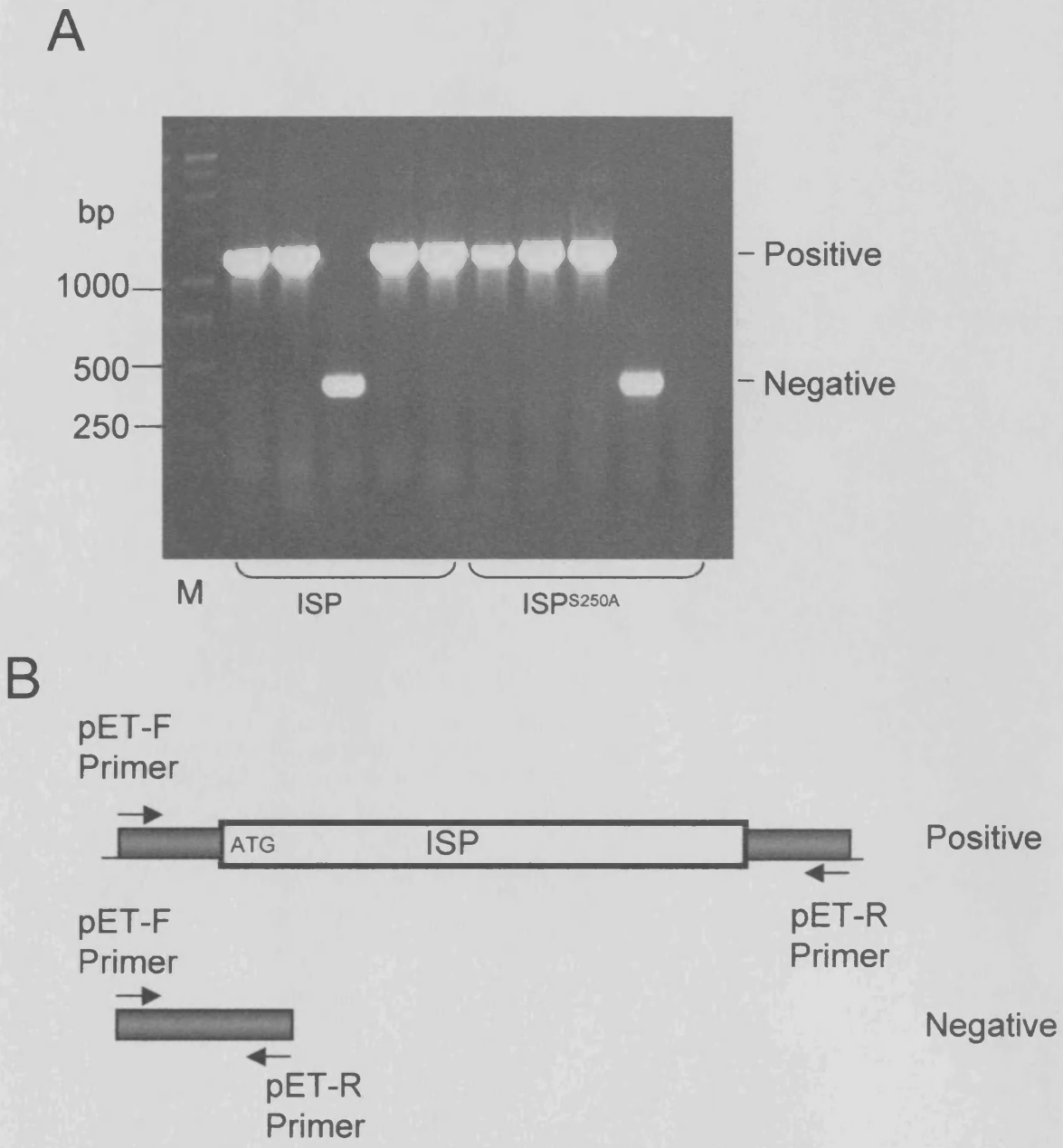


Figure 3.3 Analysis of ISP and ISP^{S250A} DNA insertion into the pET22b vector. *E.coli* DH5α cells, transformed with pET22b-ISP and pET22b-ISP^{S250A}, were screened for the presence of ISP gene inserts by PCR using primers (pET-F and pET-R) flanking the pET vector multiple cloning site. PCR products were analysed using agarose gel electrophoresis (1% w/v) (**A**) and positive inserts had the expected size 1317 bp and pET vectors without an insert showed a PCR product of 351 bp. (**B**) Schematic of the expected PCR products.

pET-F and pET-R flanking the multiple cloning site of the pET vector. Positive clones, i.e those containing a DNA sequence inserted within the multiple cloning site, were of the expected 1317 bp length, whereas no insert was observed with a 351 bp PCR product shown in Figure 3.3. To ensure correct sequence was cloned, positive PCR products were sequenced as shown in (Section 2.2.3).

3.3 Production of ISP proteins

The pET expression system has been specifically designed for the controllable and rapid production of large quantities of protein. This system has adapted the *E.coli lac* operon normally involved in the metabolism of lactose. For this system the host bacteria, normally *E.coli* BL21(DE3), genome is genetically modified supplementing genes encoding the T7 RNA polymerase and *lac* promoter and operator. This allows the user to insert a gene of interest into multiple cloning site (MCS) within a pET expression vector that is controlled by the T7 promoter. The expression vector also encodes the *lac*I repressor protein, to reduce unwanted protein expression, and an ampicillin resistance cassette to allow only bacteria transformed with the vector to survive when the antibiotic is added to the media. Protein expression is induced by the addition of the lactose analogue isopropyl β-D-1-thiogalactopyranoside (IPTG), which displaces the *lac* repressor protein allowing expression of the T7 RNA polymerase in the host genome. The T7 RNA polymerase is then available to transcribe the gene within the MCS, which is then translated into protein (Studier and Moffatt 1986, Studier *et al.*, 1990, Dubendorff *et al.*, 1991).

The *E.coli* cells BL21(DE3)gold were made electro-competent as shown in Section 2.2.5. These cells were transformed with pET22b-ISP and pET22b-ISP^{S250A} plasmids using electroporation (Section 2.2.6) and recombinant protein encoded in the MCS produced (Section 2.3.1). The primary amino acid sequence and percentage composition are shown in Figure 3.4. A molecular weight of 34906.1 Da and a theoretical pI of 4.96 were calculated using the ProtParam software (http://www.expasy.ch/cgi-bin/protparam). Protein expression was analysed using SDS-PAGE (Figure 3.5), which revealed an increase in protein level at approximately 40 kDa. These proteins were present at the first time point (0 hours) but at low levels and increased to constituted approximately 20-40% of cellular protein suggesting that

A

```
        10          20          30          40          50          60
MRKFRLIPYK QVDKVSALSE VPMGVEIVEA PAVWRASAKG AGQIIGVIDT GCQVDHPDLA
        70          80          90         100         110         120
ERIIGGVNLT TDYGGDETNF SDNNGHGTHV AGTVAAAETG SGVVGVAPKA DLFIIKALSG
       130         140         150         160         170         180
DGSGEMGWIA KAIRYAVDWR GPKGEQMRII TMSLGGPTDS EELHDAVKYA VSNNVSVVCA
       190         200         210         220         230         240
AGNEGDGRED TNEFAYPAAY NEVIAVGAVD FDLRLSDFTN TNEEIDIVAP GVGIKSTYLD
       250         260         270         280         290         300
SGYAELSGTA MAAPHVAGAL ALIINLAEDA FKRSLSETEI YAQLVRRATP IGFTAQAEGN
       310         320
GFLTLDLVER ITGQFTEKGK KLEHHHHHH
```

B

Ala	(A)	40	12.2%
Arg	(R)	13	4.0%
Asn	(N)	13	4.0%
Asp	(D)	21	6.4%
Cys	(C)	2	0.6%
Gln	(Q)	7	2.1%
Glu	(E)	24	7.3%
Gly	(G)	37	11.2%
His	(H)	11	3.3%
Ile	(I)	22	6.7%
Leu	(L)	21	6.4%
Lys	(K)	14	4.3%
Met	(M)	6	1.8%
Phe	(F)	10	3.0%
Pro	(P)	11	3.3%
Ser	(S)	17	5.2%
Thr	(T)	20	6.1%
Trp	(W)	3	0.9%
Tyr	(Y)	9	2.7%
Val	(V)	28	8.5%

Figure 3.4. Sequence characteristics of recombinant ISP with HisTag. (A) Primary sequence of ISP^{S250A}. (B) Amino acid composition of ISP^{S250A} generated using ProtParam software (http://www.expasy.ch/cgi-bin/protparam). Also calculated, a molecular weight of 34906.1 Da and a theoretical pI of 4.96.

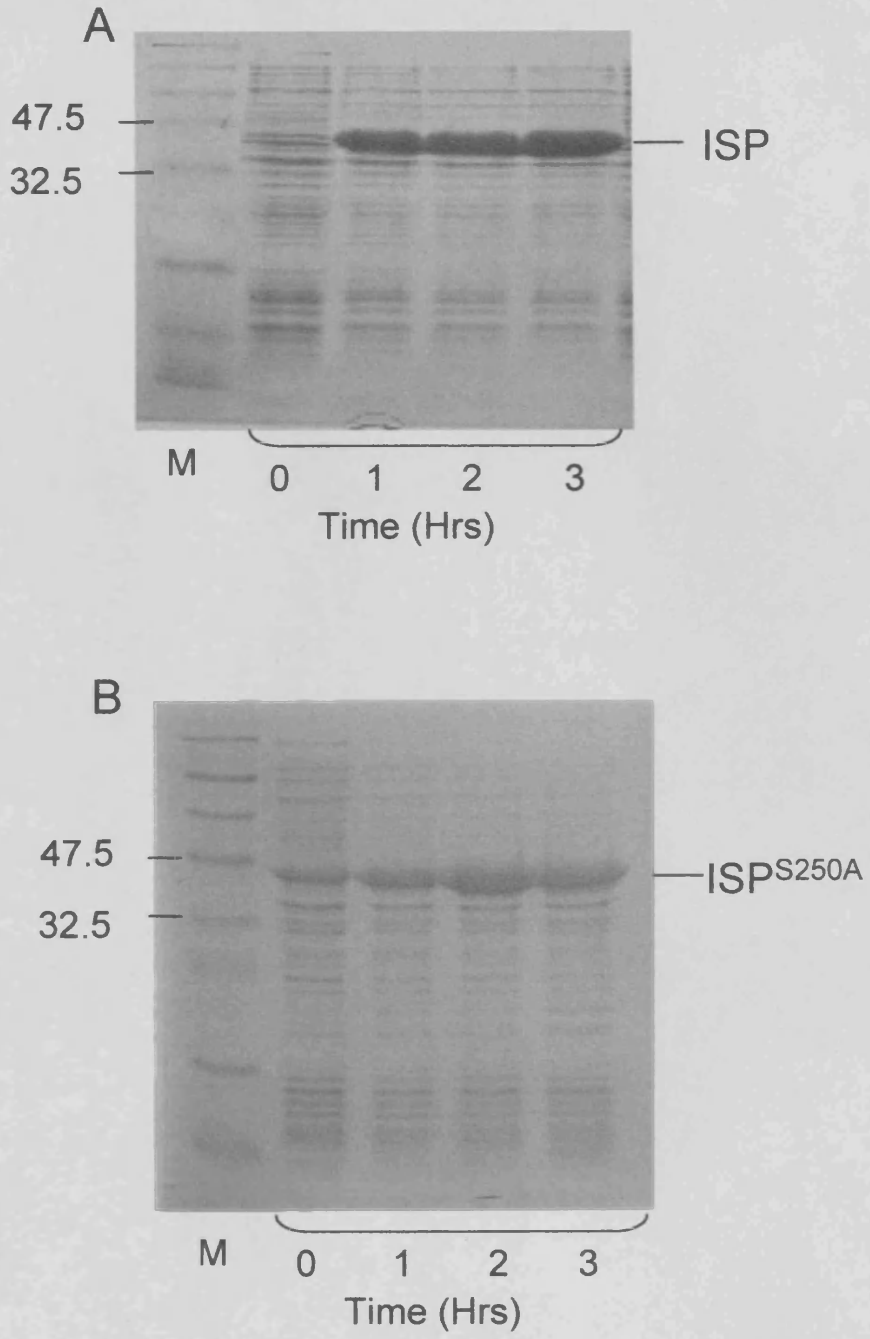


Figure 3.5 SDS-PAGE showing the over-expression of (A) ISP and (B) ISP^{S250A}. *E.coli* BL21 (DE3) bacterial cells were transformed with the pET22b-ISP or pET22b-ISP^{S250A} plasmid and induced with IPTG (1 mM). Fractions were taken at time intervals of 1-3 hours.

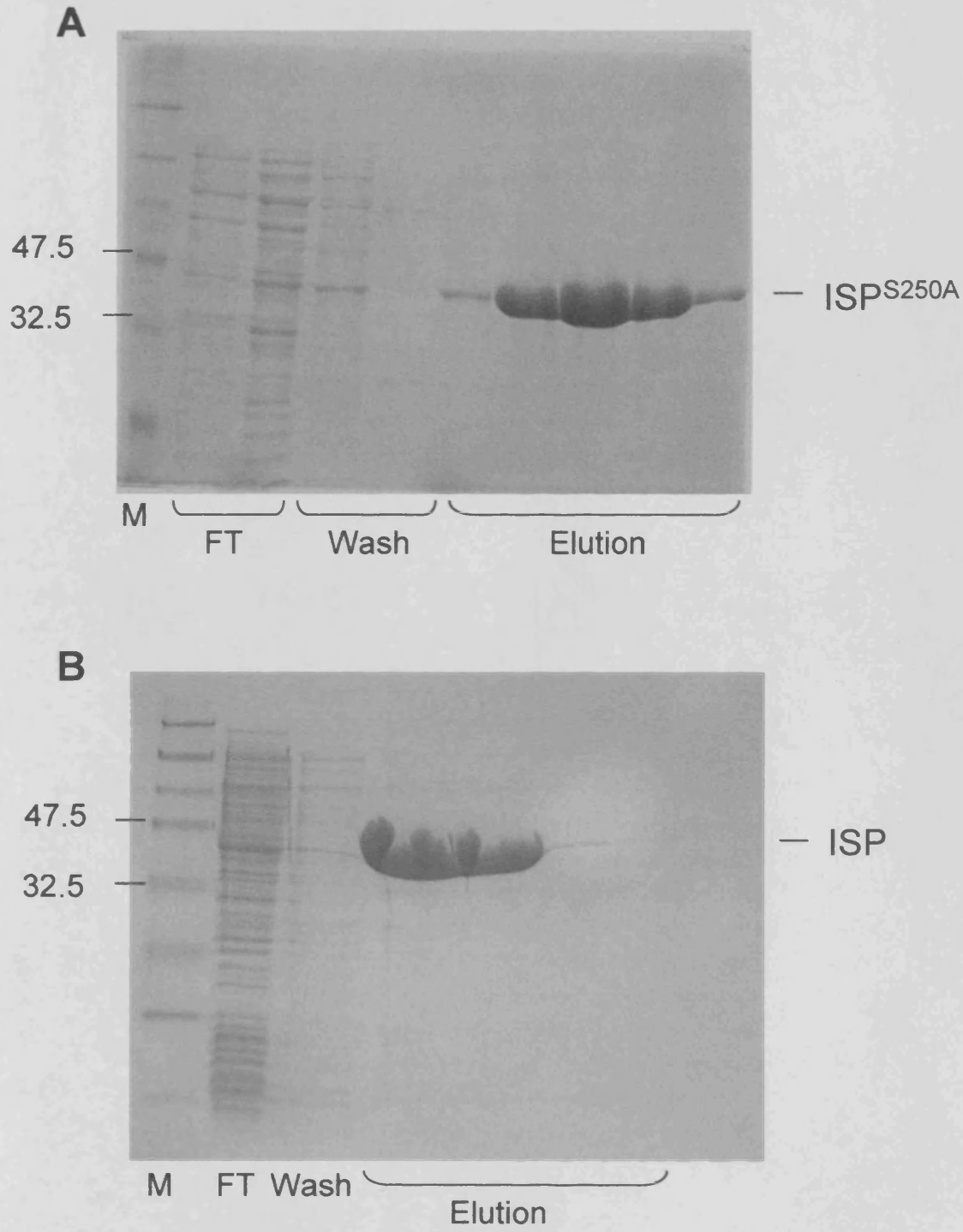


Figure 3.6 Affinity purification of ISP variants. ISP (**A**) and ISP^{S250A} (**B**) proteins containing an integrated C-terminal His Tag were purified using a HisTrap HP column attached to an AKTA chromatography system.

they were induced by the IPTG. The protein bands were larger than the expected molecular weights for the ISP variants (35 kDa), however this anomaly with ISP has been shown previously (Yang *et al.,* 1984). The *E.coli* cells were harvested by centrifugation (Section 2.3.1), lysed using a French Press and clarified via centrifugation (Section 2.3.2).

Recombinant protein was purified via immobilised metal affinity chromatography (IMAC) using a 5 ml HiTrap IMAC HP Column (Section 2.3.3). Fractions were taken during the purification procedure for analysis using SDS-PAGE. As shown in Figure 3.5, the purifications of ISP and ISP^{S250A} were similar. The larger band shown in the expression was highly reduced in the flow-through (FT) fraction after going through the column, suggesting the majority bound to the column. Protein was eluted from the column using imidazole as outlined in Section 2.3.3, which revealed the purity of the proteins to be greater than 95% by visual inspection. MALDI-TOF mass spectroscopy of the ISP^{S250A} purified product revealed that the measured Mw was within 7 Da of the expected Mw (34,782 and 34,775 Da, respectively) suggesting that the product purified was ISP^{S250A}. MALDI-TOF mass spectroscopy of the ISP purified product revealed that the product was 3230 Da smaller that the expected Mw (31,561 and 34,791, respectively). This is examined in detail within Chapter 5. ISP^{S250A} were stored at -20°C.

3.4 Discussion

The T7-promoter based pET expression system is a highly studied and widely used resource tool within the biotechnology industry and academic research (Studier *et al.*, 1990). Its ease of use and rapid production of protein makes it easier to generate large quantities of protein to utilize their function such as enzymes in food production (Rastall, 2007, Bankar *et al.*, 2009) or for study of the proteins in their own right.

The ISP and ISP^{S250A} proteins were produced and highly purified (>95%) aided by the addition of a C-terminal HisTag. This is an important result showing that ISP, especially the active variant, can be expressed in *E.coli* without damage to the host cell. This is in contrast to the ESPs, where intracellular expression is known to be detrimental to the host cell (Subbian, 2004). Therefore, this suggests that ISP must either have higher substrate specificity than the ESPs or tight regulation following expression. This will be explored in Chapter 5.

Problems arising from the addition of Histidine tags to proteins have been shown, such as contamination by histidine rich proteins (Hengen, 1995) or the generation of insoluble protein by altering the structure of the protein (Kneusel *et al.*, 1998). However, this was not the case for either ISP variant. As shown here ISP was expressed and purified to greater than 95% (Figure 3.5 and 3.6).

The ISP variants were found to be larger than expected when analysed using SDS-PAGE. This has been shown with ISP protein before (Yang *et al.*, 1984) and also with proteins that have a high negative charge (Matagne *et al.*, 1991) and ISP contains 20 aspartic acid and 22 glutamic acid residues compared to 12 arginine and 17 lysine residues, suggesting that this is most likely an anomaly of SDS-PAGE. Western blotting (using an antibody targeting the C-terminal HisTag), N-terminal sequencing and MALDI-TOF mass spectroscopy shown in Chapter 5, all showed that the ISP protein species were correct.

Therefore we were able to use modern molecular techniques to generate ISP and ISP^{S250A} proteins in relatively large quantities to study the structure, function and folding of this little studied intracellular protease.

Chapter 4

Structural Analysis of ISP from *B. clausii*

4. Structural analysis of ISP from *B. clausii*

4.1 Introduction

The bacterial intracellular subtilisin proteases (ISPs) form an important group within the subtilisin family constituting the majority of intracellular protease activity (Burnett *et al.*, 1986; Orrego *et al.*, 1973). However, key molecular details, most critically the 3D structure of the ISPs were unknown. In contrast, BPN' was one of the first structures solved by X-ray crystallography (Wright, 1969) and much is now known about the structure of many different ESPs including mutant forms (Wells *et al.*, 1988, Almog *et al.*, 1998, Tanaka *et al.*, 2007, Takeuchi *et al.*, 2009) and complexes with prodomains (Gallagher *et al.*, 1995, Ruan *et al.*, 2008) and inhibitors (McPhalen and James 1988, Radisky *et al.*, 2005).

In this chapter structural features of ISP are investigated by a variety of biochemical and biophysical approaches. The quaternary state of ISP has been disputed, however the majority of literature suggests that N-terminal processed ISPs are dimeric. Therefore, size exclusion chromatography (SEC) and analytical ultracentrifugation (AUC) were used to analyse the quaternary structure of ISP. The structural dependence of metal ion binding, which is fundamental amongst the subtilisins, is not fully understood within the ISPs. Primary sequence alignment between ESPs and ISP suggests that these sites are highly conserved (Figure 1.7, Section 1.5). Therefore, biophysical spectroscopic techniques were used to probe any structural changes to ISP and ISP^{S250A} upon the removal of bound metal ions using the divalent metal ion chelator EDTA.

This chapter will also include the first 3D structure of an ISP solved by X-ray crystallography. The structure was solved by Professor Keith Wilson's team at the Structural Biology Laboratory, University of York in collaboration with the Jones group, most notably myself who supplied the protein. The ISP^{S250A} protein produced during this project was chosen as the target because it shows no protease activity to complicate analysis and may provide structural and mechanistic rationale for inhibition by the N-terminal extension.

4.1.1 Introduction of the spectroscopic techniques CD and Fluorescence

Circular dichroism (CD) and fluorescence have become valuable tools for investigating the structural properties of proteins. CD spectroscopy measures the differential absorption of anti-clockwise (left handed) and clockwise (right handed) polarised light of a chiral sample. CD analysis is split into two wavelength regions: Near-UV (250-350 nm) and far-UV (190-250 nm). Signal within the near-UV is normally associated with aspects of tertiary structure, such as the chromophores of the aromatic amino acids as well as disulphide bond linkage. CD signal in the far-UV is mainly due to the orientation and rotation of the peptide backbone and is independent of the side chain transitions. The rotation around the peptide backbone (Phi and Psi dihedral angles) in theory would allow proteins to fold into many secondary structure conformations. However, steric crowding from non-bonded atoms results in only a few structural conformations being possible, each with well-defined CD signal. Therefore CD spectroscopy, within the far-UV, is widely used to study protein secondary structure. Furthermore, a range of computer programs have been developed that compare the experimental CD spectrum with a library of known structures and CD spectra and calculate the percentage of each different secondary structure conformations. In this project the program developed by WC Johnson, CDSSTR, will be used (Johnson, 1999). One of the most common secondary structures in proteins is the α-helix. The α-helix structure is normally made up of approximately 10 residues stabilised by hydrogen bonds between the backbone NH and the C=O from 4 residues previously, which results in a rotation of 100° between residues with 3.6 residues making up one turn of the helix. The CD spectrum has characteristic negative signal maxima at 222 nm and 208 nm. The second most accessible and well-organised secondary structure conformation is the β-sheet, with hydrogen bonds forming between rows of parallel or anti-parallel strands. The CD signal observed for β-sheet structure is generally characterised by a negative maxima at about 216 nm and a comparable positive band at 196 nm. The β-turn is normally associated with the rotation allowing the formation of an anti-parallel β-sheet but can also include all turns that occur in the polypeptide chain. This structure is stabilised by hydrogen bonds between one residue and the n+3 residue forming a turn. The CD signal associated with this structure has a negative band near 225 nm and a positive band between 200 and 205 nm. An

alternative helix-type secondary structure conformation includes the 3_{10} helix, which forms a turn of 120° with 3 residues making one complete turn. This is a tighter helix and is normally associated with shorter helices and often found at the termini of α-helices. The nomenclature of this type of helix uses the number of residues per turn and the number of atoms in the ring formed by the hydrogen bond. In comparison, using the same nomenclature, the alpha helix can be called a 3.6_{13} helix. The loss of secondary structure due to unfolding, especially the main conformations (α-helix, β-sheet and turns), results in the generation of unfolded protein termed random coil. The CD spectrum of unfolded protein is characterised by a weak positive peak at 218 nm and a larger negative band just below 200 nm.

Fluorescence spectroscopy takes advantage of the intrinsic fluorescent nature of predominantly the tryptophan residues. Excitation of the aromatic chromophore at 280 nm results in emission of a photon in the range 300-350 nm depending on the polarity of the surrounding solvent. Normally aromatic residues are located within the non-polar environment found within the hydrophobic core of a protein. This results in an emission maximum below 350 nm. If the chromophore is exposed to polar aqueous environment, for example during protein unfolding, a shift in the emission maxima as well as a decrease in fluorescence intensity is observed due to quenching. Therefore both CD and fluorescence are highly complementary in the study of proteins and are different probes to detect changes in protein structure.

4.2 Overview of ISP primary structure features.

The ISPs are closely related to the ESPs and share approximately 40-50% sequence identity. Amino acid sequence alignment of several ISPs and the mature domain of two ESPs from various Bacilli species (Figure 1.7, Section 1.5) suggest that the ISPs share a common core structure and general function with the ESPs. High homology is observed surrounding the active site residues D49, H86 and S250 suggesting that these residues are important for the structure orientation of the active site. Sequence homology is also observed around the potential metal ion binding sites D58 and A97-V103, suggesting ISPs bind calcium. Neither the N-terminal signal sequence nor the prodomain present in the ESPs are observed in the ISPs. Algorithms that predict whether a protein contains a signal sequence for secretion (e.g. SignalP 3.0, www.cbs.dtu.dk/services/SignalP) predicted that the ISP was a non-secretory protein.

This lack of a signal sequence was expected due to the cellular location of the ISPs. However, the lack of a folding chaperone resembling the ESP's prodomain is more intriguing and suggests an alternative mechanism for ISP folding. The ISPs have an N-terminal extension (15-20 residues) with a conserved LIPY(F) sequence which is much shorter than the 70-90 residue ESP's prodomain. The structural and functional role of this region is discussed later in Chapter 5, *Processing and Activation of ISP*. There is also lack of sequence homology observed at the C-terminus past the core helix housing the active site serine. Beyond a structural role, the C-terminal end has no functional role in the ESPs.

4.3 Crystal structure of ISP from *B. clausii*

During the later stages of this project the structure of the *B. clausii* ISP^{S250A} variant, as outlined in Section 3.3, was determined by collaborators at the University of York by X-ray crystallography. Two independent structures from different crystallisation conditions were determined; Crystal 1 at a resolution of 2.7 Å and refined to an R and Rfree of 16.7 and 25.0%, respectively; Crystal 2 at 2.5 Å to an R and Rfree of 18.8 and 26.9%, respectively. In both cases the structure were determined using molecular replacement with the BPN' structure (1YJC) used as the main refinement model.

The details of the 3D structure of ISP^{S250A} will be presented in each of the following sections in relation to the biochemical and biophysical investigations. The general structure of ISP is consistent with the α/β hydrolase fold found in the majority of the mature domains within bacterial subtilisins (Tonaka *et al.*, 2004). This fold is consists of 8 α-helices and 9–11 β-strands that forms a stable scaffold that correctly orientates the active site residues (Figure 4.1). The sequence similarity is generally low between these proteases however the overall fold is comparable. This fold is most often linked with serine, cysteine or aspartic acid proteases (Nardini and Dijkstra., 2000). The residue that provides the nucleophile is always located in a sharp turn, termed the 'nucleophile elbow'. This allows easy access for the substrate and the hydrolytic water molecule. The differences in the structure are discussed in detail in later chapters. The role of other structural features such as the N-terminal and substrate-binding site will be discussed in more detail in Chapters 5 (*Processing and Activation of ISP*) and 6 (*Characterisation of the Enzymatic Activity of ISP*), respectively. However, some of the

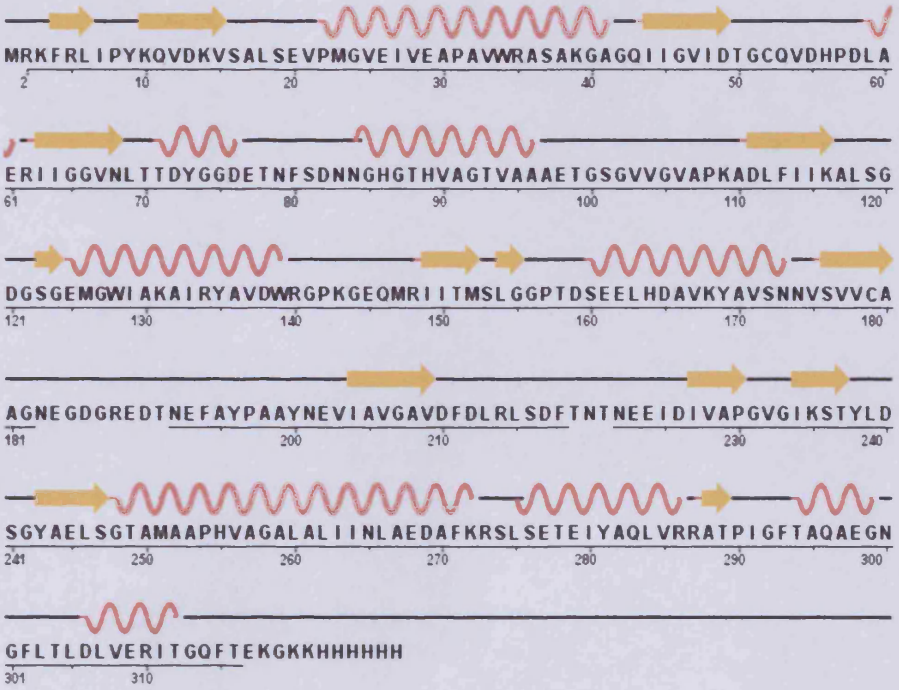


Figure 4.1 The primary amino acid sequence and secondary structure elements of ISP. The α-helices and β-sheets are coloured red and yellow, respectively. Image taken from PDB website (http://www.pdb.org), ISP from *Bacillus clausii* (2WV7).

features of the crystal structure are outlined here. These crystals produced a unit cell comprised of six highly similar protomers arranged as three independent dimers. The r.m.s.d. of the Cα atomic coordinates was approximately 0.4-0.5 Å. The structure of each protomer included residues K3-E317. No density was observed for the first two amino acids at the N-terminus (Met and Arg) or the final four residues naturally occupying the ISP^{S250A} C-terminus (KGKK) and the adjoining hexa-His Tag (LEHHHHHH). In addition, two regions are either disordered or show substantial differences between protomers in part reflecting their different environments within the crystal: residues 183-193 and 216-223. The equivalent residues in the ESPs have well defined density and are part of an elaborate hydrogen-bonding network contributing to the S1 binding pocket. There was no interpretable density in any of the protomers for residues 183-193 even though there is some degree of conservation amongst the subtilisins (Figure 1.7, Section 1.5). Residues 216-223, which are also conserved in the subtilisins, can be modelled in chains B, D and F. In the other three protomers there is uninterpretable difference density in this region, suggesting the presence of multiple conformations rather than complete disorder.

4.4 Quaternary structure of ISP.

Normally subtilisins are monomeric but it has been suggested that ISP are dimeric. To investigate the quaternary structure of ISPs, the native molecular weights of ISP^{S250A} was investigated by size exclusion chromatography (SEC) and analytical ultracentrifugation (AUC).

ISP^{S250A} eluted from the Superdex 200 column (GE Healthcare) at 14.9 ml at the three ISP^{S250A} concentrations assayed (10, 25 and 50 μM) Figure 4.2. The apparent molecular weights of protein were measured by SEC, which was calibrated with proteins of known molecular weights (Section 2.3.11) (Figure 4.3). The apparent molecular weight of ISP^{S250A} was 55 kDa. This is larger than the expected molecular weight for the monomeric ISP^{S250A} (35 kDa) and closer to the estimated molecular weight of an ISP^{S250A} dimer (68 kDa) (Table 4.1). Therefore the hydrodynamic volume occupied by ISP^{S250A} suggests that ISP either forms a partially unfolded monomer or a compact dimer. CD spectroscopy, which is used to probe the secondary structure of proteins, indicated that ISP^{S250A} was folded with characteristic troughs at 208 and 222 nm suggesting a mixed α-helix and β-sheet conformation (Figure 4.4). The ISP^{S250A} CD

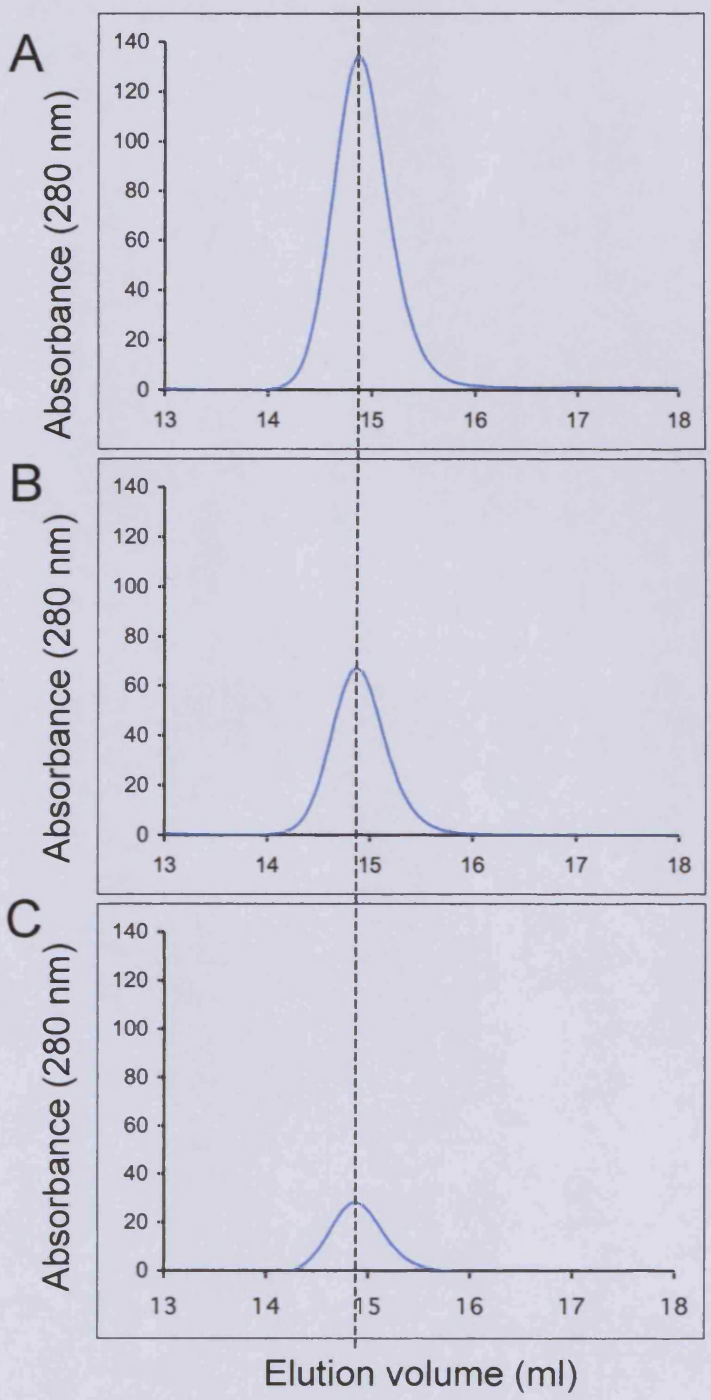


Figure 4.2. Analysis of ISP^{S250A} quaternary structure using SEC. SEC was performed in buffer (50 mM sodium phosphate, pH 7.5, 0.2 M ammonium sulphate and 1 mM calcium chloride at different ISP^{S250A} concentrations (**A**) 50 µM, (**B**) 25 µm and (**C**) 10 µM.

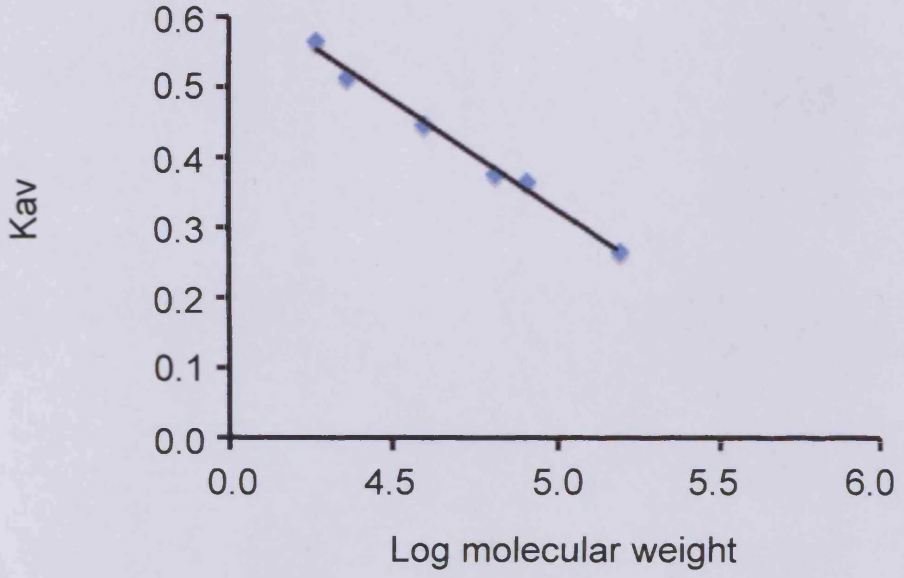


Figure 4.3 Standard curve for SEC. An analytical Superdex 200 column (GE Healthcare) was calibrated using the protein standards ribonuclease A (13.7 kDa, Kav 0.568), chymotrypsinogen A (25 kDa Kav 0.514), ovalbumin (43 kDa Kav 0.448), albumin (67 kDa Kav 0.377), transferrin (81 kDa Kav 0.368) and gamma globulin (150 kDa Kav 0.268). The void and total volumes was measure using blue dextran (2000 kDa) and tryptophan (204 Da), respectively.

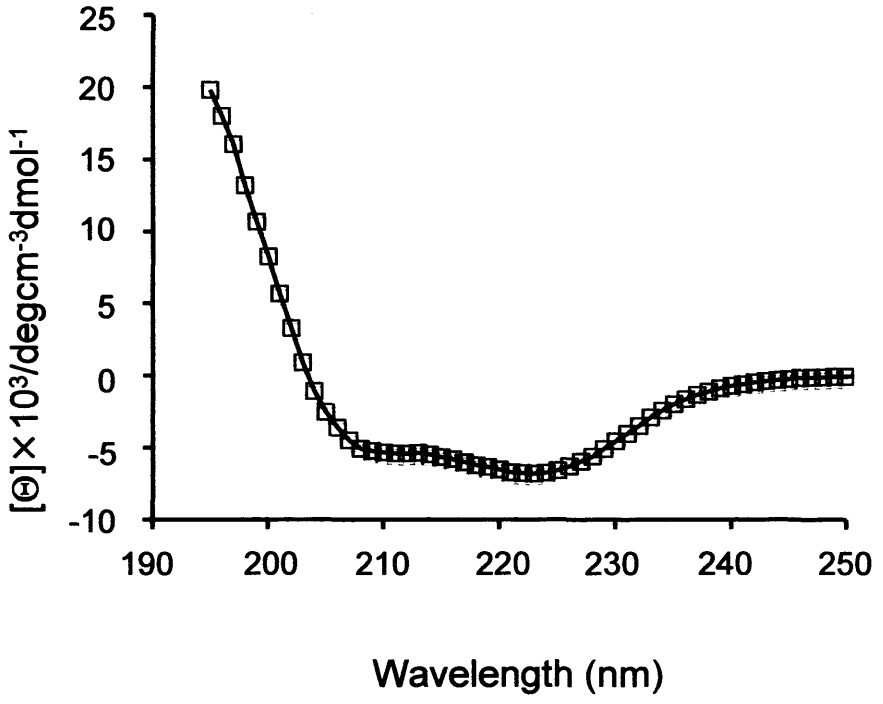


Figure 4.4 CD spectrum of ISP^{S250A}. The CD spectrum of 5 mM ISP^{S250A} in 20 mM sodium phosphate, pH 7.5, 0.1 M ammonium sulphate..

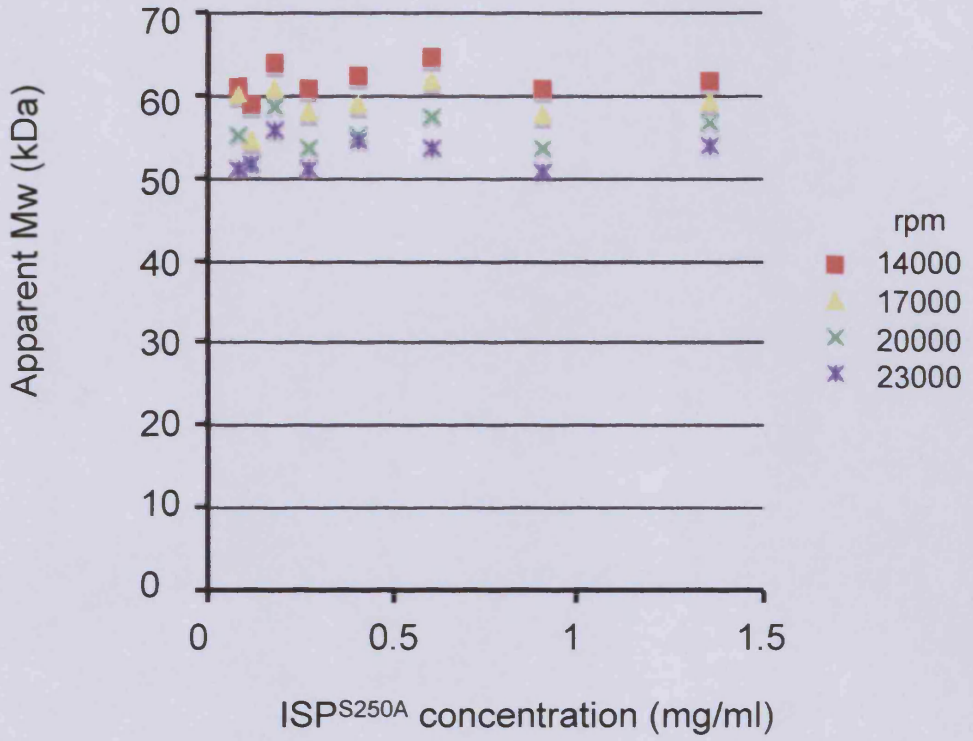


Figure 4.5 AUC data of ISP^{S250A}. Sedimentation equilibrium experiments were conducted on a Beckman Optima XL/I analytical ultracentrifuge in 50 mM sodium phosphate 200 mM ammonium sulphate and 1 mM TCEP.

spectrum almost perfectly matched the spectrum measured for the active ISP, which in most proteases requires a stable folded conformation for activity (Figure 5.7, Section 5.4). There was no observed shift in the elution volume at the concentrations assayed. This suggests that ISP^{S250A} remains a dimer at concentrations ≥10 μM.

AUC data, performed by the Technology Facility, Department of Biology, at York University, confirmed that the SEC data for ISP^{S250A}, with a buoyant molecular weight calculated to be 57 kDa, therefore closer to the molecular weight calculated for the dimer than the monomer (Figure 4.5). There was some decrease of Mw on increasing speed that may indicate the presence of a small amount of monomer or other low molecular weight species, but this is apparently not in equilibrium with the dimer as there is no change in molecular weight across the concentration range used. This suggests that ISP^{S250A} is more likely to be in a dimeric rather than a monomer quaternary state.

Table 4.1 Native molecular weights of ISP

	Calculated Mw (kDa)		Measured Mw (kDa)
	Monomer	Dimer	
SEC	35	70	57± 4
AUC	35	70	55± 5

ISPs contain three cysteine residues and disulphide bonds have been identified within subtilisins, such as subtilisin S41 from *Bacillus TA41* (Davail *et al.*, 1994). To determine if dimerisation was the result of the formation of disulphide bonds the SEC elution profile of ISP^{S250A} was determined in the presence of the disulphide bond reducing agent dithiothreitol (DTT). The addition of DTT (1 mM) did not alter the elution profile of ISP^{S250A} with protein eluting at 14.9 ml, the same as in the absence of DTT (Figure 4.6). This suggests that disulphide bond formation is not required for dimerisation.

The new structure of ISP^{S250A}, including the location of the catalytic triad, the N-terminal extension and putative metal binding sites, is shown in Figure 4.7A. Importantly, this confirms that ISP^{S250A} is a dimer with interactions mainly in the C-terminal. The core of the protein containing the mature domain (residues 19-305) has the same overall fold as the homologous mature ESPs (Figure 4.7B). Superposition of

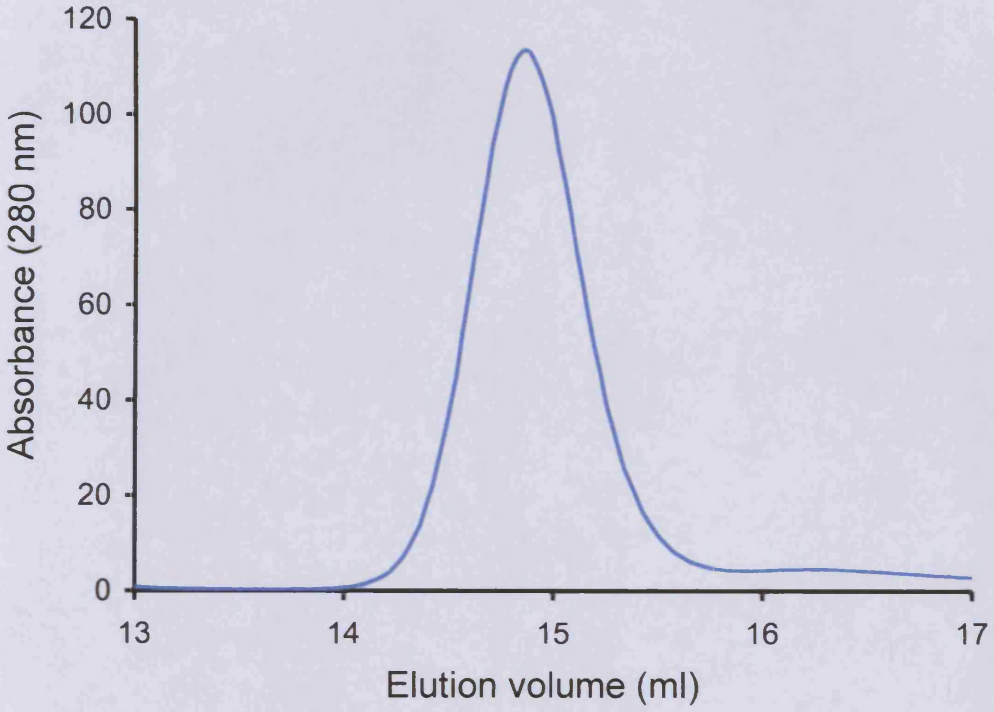


Figure 4.6 Effect of dithriothreitol (DTT) on the elution profile of ISP^{S250A}. The assay was performed using 50 μM ISP^{S250A} in buffer (50 mM sodium phosphate, pH 7.5, 0.5 M ammonium sulphate) supplemented with 1 mM DTT. For comparison the elution profile of ISP^{S250A} in the absence of DTT is shown in Figure 4.1.

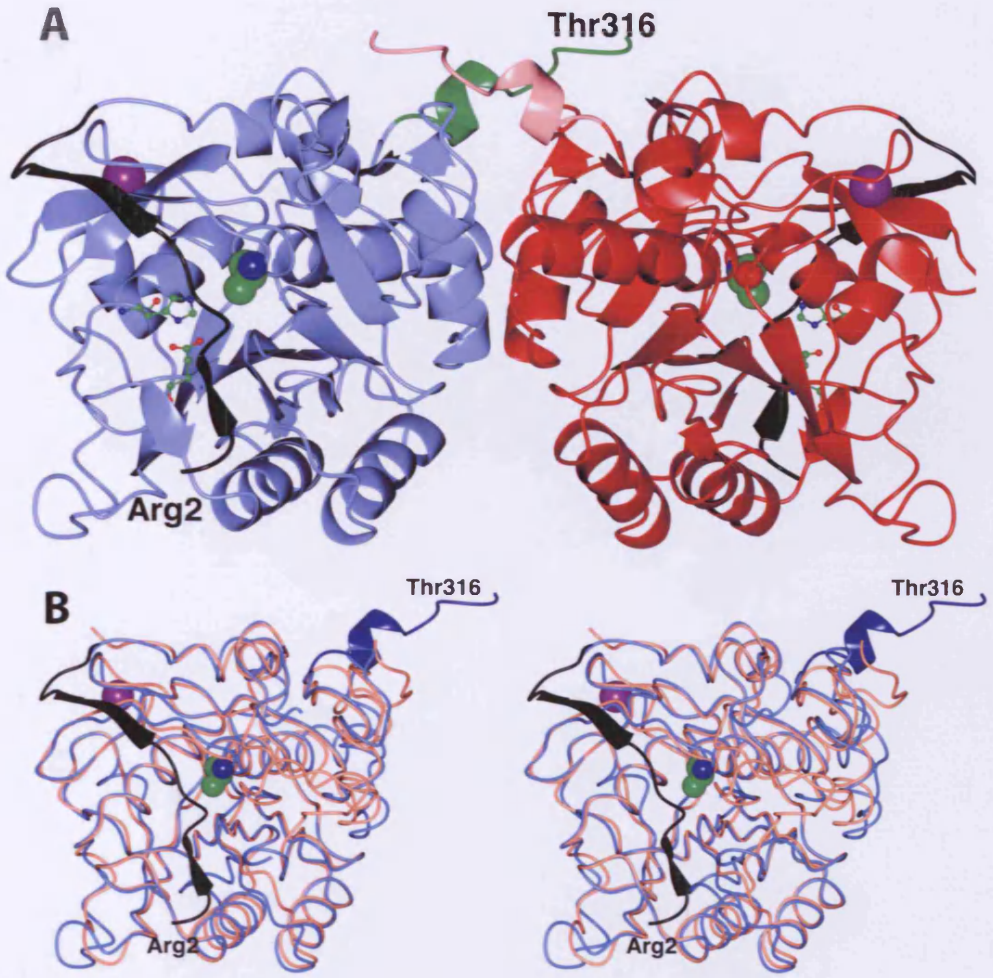


Figure 4.7 The 3D structure of ISP^{S250A}. (A) Ribbon representation of the AB dimer, with subunits A (blue) and B (red). The N-terminal extension, residues 2-18, is coloured in black. The C-terminal extension involved in the dimer interface is coloured green and pink in subunits A and B, respectively. The catalytic Ala250 is shown as spheres (where the carbon, nitrogen and oxygen atoms are coloured green, blue and red, respectively), with the other residues of the catalytic triad, His86 and Asp49 in ball-and-stick. N- and C- terminal residues are labelled in subunit A (blue). (B) Stereo view of the superimposition of 261 structurally equivalent residues of protomer A of ISP (blue worm) on subtilisin BPN' (coral, PDB 1TO2). Ala250 is shown as spheres. The N-terminal extension is shown as a black ribbon and the C-terminal as a dark blue ribbon and metal ions are shown as purple spheres. The Cα rmsd was 1.27 Å. Molecular structures were generated using CCP4mg using default fitting parameters (Potterton *et al.*, 2004).

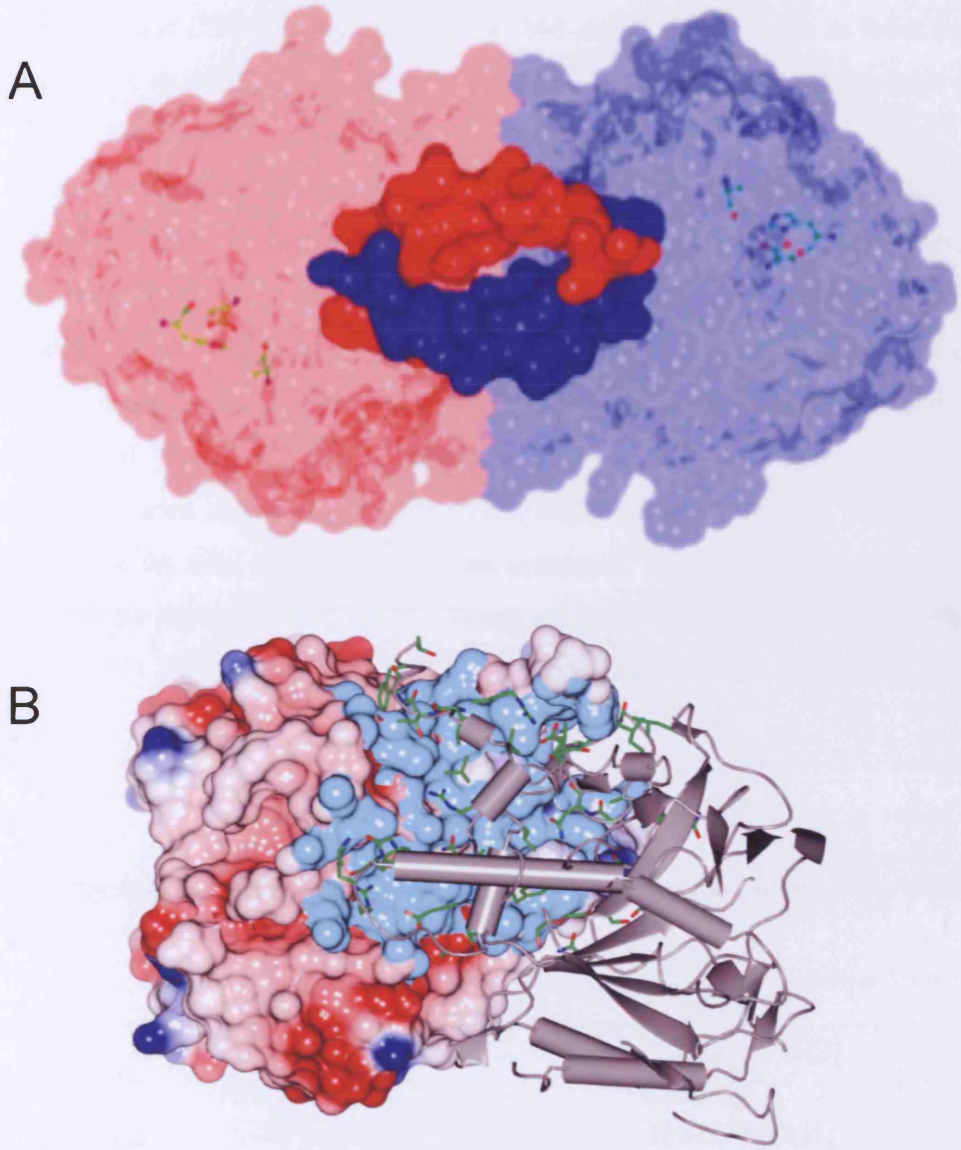


Figure 4.8 The dimer interface of ISP^{S250A}. (A) The two subunits shown as surfaces, with the C-terminal extensions crossing over to make substantial contributions to the interface. (B) Subunit A as an electrostatic surface and B as worms and tubes. The side chains for residues in Chain B involved in dimer interactions are in ball-and- stick and the surface of protomer A which is buried in the interface is shaded cyan.

ISP^{S250A} with a representative structure of subtilisin BPN' (PDB 1TO2; (Radisky *et al.*, 2004)), gave a Cα r.m.s.d. of 1.27 Å over the 261 structurally equivalent residues. There are differences in some surface loops, in part reflecting the insertions/deletions between ESPs and ISPs around residues 74, 144 and 270. Variation in these regions is not unexpected, as other ESPs, for example from the archeon T. kodakarensis (Tanaka *et al.*, 2007), have several larger loops.

The 3D structure reveals the C-terminal sequences cross over and interact with the adjacent subunit (Figure 4.8). Also evident in this figure is the location of the catalytic triad, which is relatively distant from the dimer interface suggesting dimerisation is not required to form the active site. Furthermore, the N-terminal extension unique to the ISPs does not appear to play a role in dimerisation. PISA (Krissinel and Henrick, 2007; http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) estimates the buried surface area within each dimer to be ~1500 Å2, which equates to around 12% of the total surface area of the protomer. There are 14 hydrogen bonds at the interface as shown in Table 4.2. Seven of these involve the side chains of the residues Arg273, Gln283, Arg286 and Asn300 and eight involve main chain CO or NH groups.

Table 4.2 Hydrogen bonds present in the ISP^{S250A} dimer interface. The distances correspond to those seen in the AB dimer.

	Subunit A	Dist.[Å]	Subunit B
1	ILE225[O]	3.2	ARG273[NH2]
2	ARG273[NH1]	3.1	ILE225[O]
3	THR278[OG1]	3.4	THR278[OG1]
4	GLN283[NE2]	2.7	ASN300[OD1]
5	VAL285[O]	3.5	ARG286[NH1]
6	ARG286[NH1]	2.9	ASN300[O]
7	ALA288[O]	3.2	ARG286[NH1]
8	PRO290[O]	2.8	GLY313[N]
9	GLN296[N]	3.3	ALA270[O]
10	ASN300[OD1]	2.7	GLN283[NE2]
11	ASN300[OD1]	3.2	ARG286[NE]
12	ASN300[O]	3.0	ARG286[NH2]
13	ARG310[NH2]	3.4	THR316[OG1]
14	GLY313[N]	2.7	PRO290[O]

4.5 Metal ions and ISP structure.

Calcium is essential within the subtilisins for structural stability and hence activity with the majority of subtilisins (Bryan 2000) having one or more calcium binding sites. The main site, termed site A, is an high affinity binding site coordinated in a loop consisting amino acids 75-83, within subtilisin BPN', and is highly conserved. Sequence alignment (Figure 4.9) of ISPs from various *Bacillus* species and two ESPs (BPN' from *Bacillus amyloliquifaciens* and Savinase from *Bacillus lentus*) suggest that the high affinity binding site found in the ESPs is also present within the ISPs with amino acids that coordinate calcium being conserved. However, the role of calcium binding within the ISPs remains uncertain (Kurotsu *et al.*, 1982, Koide *et al.*, 1986, Tsuchiya *et al.*, 1997, Subbian *et al.*, 2004).

4.5.1 Spectroscopic analysis of the metal ion dependency of ISP

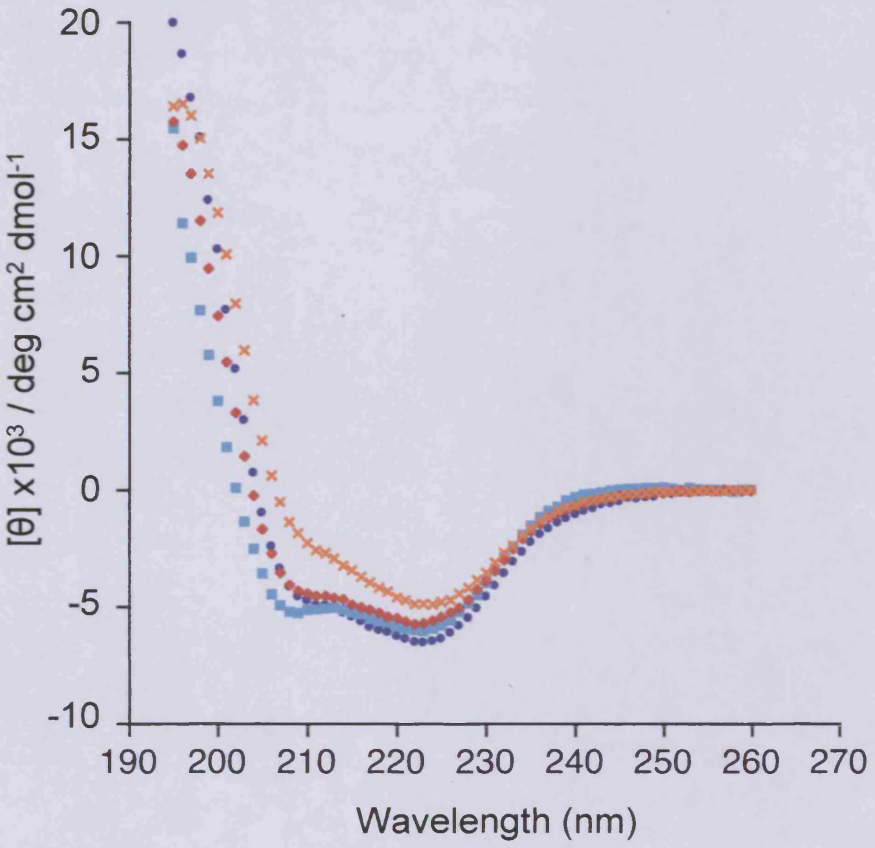
To investigate any effect the removal of divalent metal ions, such as calcium, have on the structure of ISP^{S250A}, ethylenedinitrilotetraacetic acid (EDTA) was added to the protein and CD and fluorescence spectroscopy were used to monitor structural changes. EDTA is a polyamino carboxylic acid and binds divalent metal cations through its two amines and four carboxylates. The CD spectrum of ISP^{S250A} suggests that the protein is undergoing a structural change in the presence of EDTA (Figure 4.10A). ISP^{S250A} in buffer supplemented with 1 mM calcium showed a characteristic spectrum for a mixed α-helical and β-sheet-containing protein and suggests that the protein is folded. Deconvolution of this spectrum using CDSSTR software revealed ISP^{S250A} to have 18% α-helical and 24% β-strand content. This is consistent with approximately 19% α-helical and 26% β-strand calculated from the ISP^{S250A} structure. The addition of 10 mM calcium caused a small change in the ellipticity at 208 nm and 222 nm compared to the spectrum of ISP^{S250A} in 1 mM calcium. The addition of 0.1 mM EDTA caused a small perturbation in the spectrum of ISP^{S250A} compared to that in the presence of 1 mM $CaCl_2$. However, at a higher concentration of EDTA (1 mM) these changes were exacerbated with less negative ellipticity observed at 208 nm and 222 nm; CDSSTR analysis revealed an apparent 10% decrease in the helical content of ISP^{S250A} and an apparent 11% increase in the β-sheet (Figure 4.10B). Deconvoltion of the CD depends highly on the reference proteins used to generate the percentages of secondary

```
ISP-B.clausii      --------MRKFRLIPYKQVDKVSALSEVPMGVEIVEAPAVWKASAKGAGQIIGVIDTGC 52
ISP-B.cereus       MNSSTANKQKGIQLIPFTVDKVVEQVNEIPPGVQLIHAPQVWEKSAKGKDIVVAVLDTGC 60
ISP-B.anthracis    MNSSTANKQKGIQLIPFTVDKVVEQVNEIPPGVQLIHAPQVWEKSAKGKDIVVAVLDTGC 60
ISP-B.subtilis     -------MNGEIRLIPYVTNEQIMDVNELPEGIKVIKAPEMWAKGVKGKNIKVAVLDTGC 53
ESP-BPN            ------------------------AQSVPYGVSQIKAPALHSQGYTGSNVKVAVIDSGI 35
ESP-Savinase       ------------------------AQSVPWGISRVQAPAAHNRGLTGSGVKVAVLDTGI 35
                                          ..:* *:. :.**     . .* .  :.*:*:*

ISP-B.clausii      QVDHPDLAERIIGGVNLTTDYGGVETNFSDNNGHGTHVAGTVAAAETGSGVVGVAPKADL 112
ISP-B.cereus       DMNHIDLKDRIIGGRNFTKDYEGDPNIYLDNNGHGTHVAGTIAATENGVGVLGVAPLAKM 120
ISP-B.anthracis    DMNHIDLKDRIIGGRNFTKDYEGDPNIYLDNNGHGTHVAGTIAATENGVGVLGVAPLAKM 120
ISP-B.subtilis     DTSHPDLKNQIIGGKNFTDDDGGKEDAISDYNGHGTHVAGTIAANDSNGGIAGVAPEASL 113
ESP-BPN            DSSHPDLK--VAGGASMVP---SETNPFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASL 90
ESP-Savinase       S-THPDLN--IRGGASFVP---GEPS-TQDGNGHGTHVAGTIAALNNSIGVLGVAPSAEL 88
                   .  * **   : ** .:.    .      * *.********:** :.. *: **** *.:
```

Figure 4.9 Sequence alignment of the subtilisins high-affinity metal binding site. The alignment of various ISPs from *Bacillus* species and two ESPs. The highlighted residues represent the conserved ligand binding site within BPN' (PDB 1TO2): Q2, D41, L75, N77, I79 and V81.

A



B

	%	α-helix	β-strand	β-turns	others
Calcium	1 mM	18	24	10	48
	10 mM	18	25	9	48
EDTA	0.1 mM	17	23	13	47
	1 mM	7	35	11	47

Figure 4.10 Circular dichroism analysis of ISP^{S250A} in the presence of calcium or EDTA. (**A**) The CD spectra for ISP^{S250A} in buffer (20 mM sodium phosphate (pH 8.0) and 0.2 M ammonium sulphate) supplemented with either $CaCl_2$ (1 mM (blue) or 10mM (light blue)) or EDTA (0.1 mM (red) or 1 mM (orange)). (**B**) Deconvolution of the CD spectra was performed using CDSSTR software (Johnson 1999), and the proportion of relative secondary structure elements.

structural elements. Whilst this is feasible for native proteins there is a risk when using deconvolution programs for partially unfolded protein since the protein will most likely be in equilibrium with different structures that cannot be accounted for in the reference set. However, these data suggest that chelation by EDTA led to structural changes to ISP^{S250A}, suggesting that calcium may be important in maintaining the native fold of ISP^{S250A}.

Analysis of ISP^{S250A} intrinsic fluorescence in the presence of 1 mM calcium revealed a wavelength emission maxima (λ_{max}) of 350 nm (Figure 4.11). The λ_{max} of buried tryptophan residues is normally in the range 320-340 nm, suggesting that the tryptophan residues are partially solvent exposed. The crystal structure of ISP^{S250A} confirms this (Figure 4.12). The solvent accessible area (ASA) for each of the tryptophans W34, W128 and W139 was 10, 18 and 13%, respectively. This was calculated using the GETAREA program (http://curie.utmb.edu/getarea.html) where values less than 20% represent a buried amino acid. When compared to fully buried residues, such I150, which has an ASA of 0%, this suggests that all three tryptophan residues are partially buried. The fluorescence emission spectrum of ISP^{S250A} supplemented with 10 mM calcium showed only a minor change, in line with observations using CD. The addition of EDTA has a more significant effect on the fluorescence emission spectrum of ISP^{S250A} (Figure 4.11). Both concentrations of EDTA (0.1 and 1 mM) caused a decrease in intensity, of approximately 80%. This confirms, as suggested by CD, that chelation of a divalent metal ion by EDTA causes a structural change of ISP^{S250A}.

4.5.2 Effects of EDTA on the quaternary structure of ISP^{S250A}

As shown previously, ISP^{S250A} in the presence of calcium (1 mM) had an apparent molecular weight of 55 kDa (Figure 4.5). To analyse the dependence of the dimeric state of ISP^{S250A} on calcium, SEC was performed in the presence of EDTA. The elution volume for ISP^{S250A} in buffer supplemented with EDTA (50 mM) was 15.1 ml (Figure 4.13) equivalent to an apparent molecular weight of 50 kDa. To determine if this change in the elution volume was due to a structural change in ISP^{S250A} and not other events, such as proteolysis, SEC was performed with ISP^{S250A} in EDTA. The fraction corresponding to the main peak that eluted at 15.1 ml was collected and

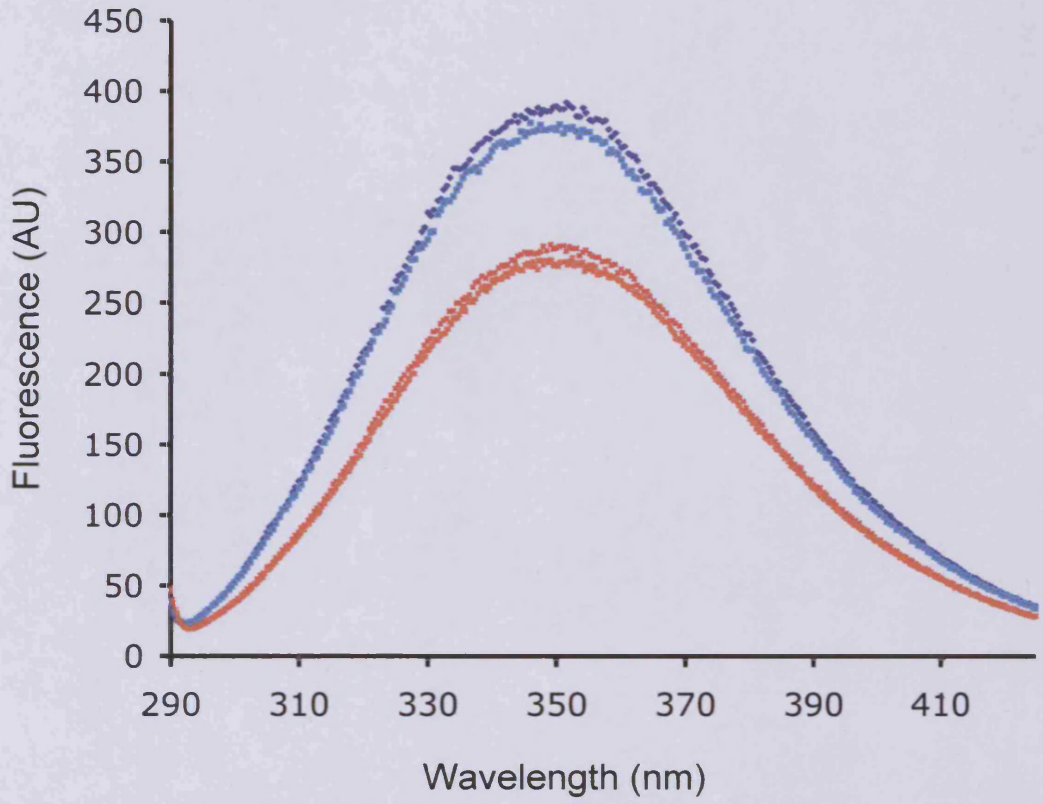


Figure 4.11 Fluorescence emission spectra of full length ISP^{S250A} in the presence of calcium or EDTA. ISP^{S250A} was incubated for one hour in buffer (20 mM sodium phosphate pH 8.0 and 0.1 M ammonium sulphate) supplemented with either $CaCl_2$ (1 mM (blue) or 10mM (light blue)) or EDTA (0.1 mM (red) or 1 mM (orange)). The excitation wavelength was 280 nm.

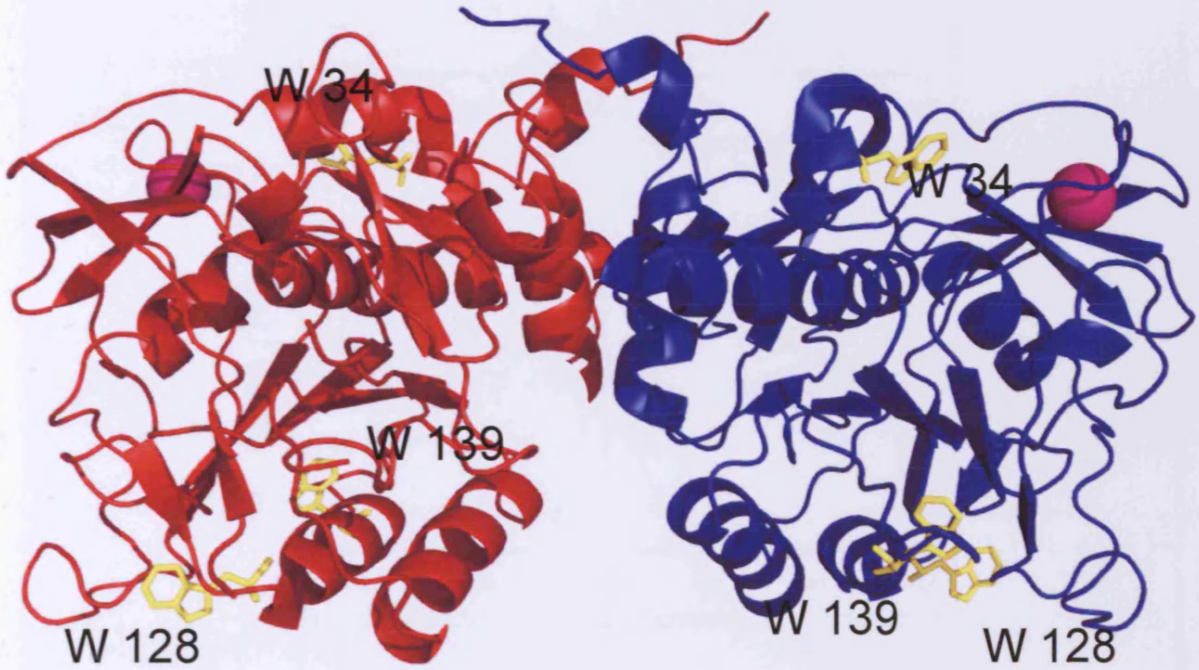


Figure 4.12 Location of tryptophan residues in the structure of ISP^{S250A}. The A (red) and B (blue) dimer are shown as a ribbon representation with the three tryptophan amino acids (W34, W128 and W139) for each protomer shown as sticks (yellow). Also shown are the metal ion binding site (sphere, pink)

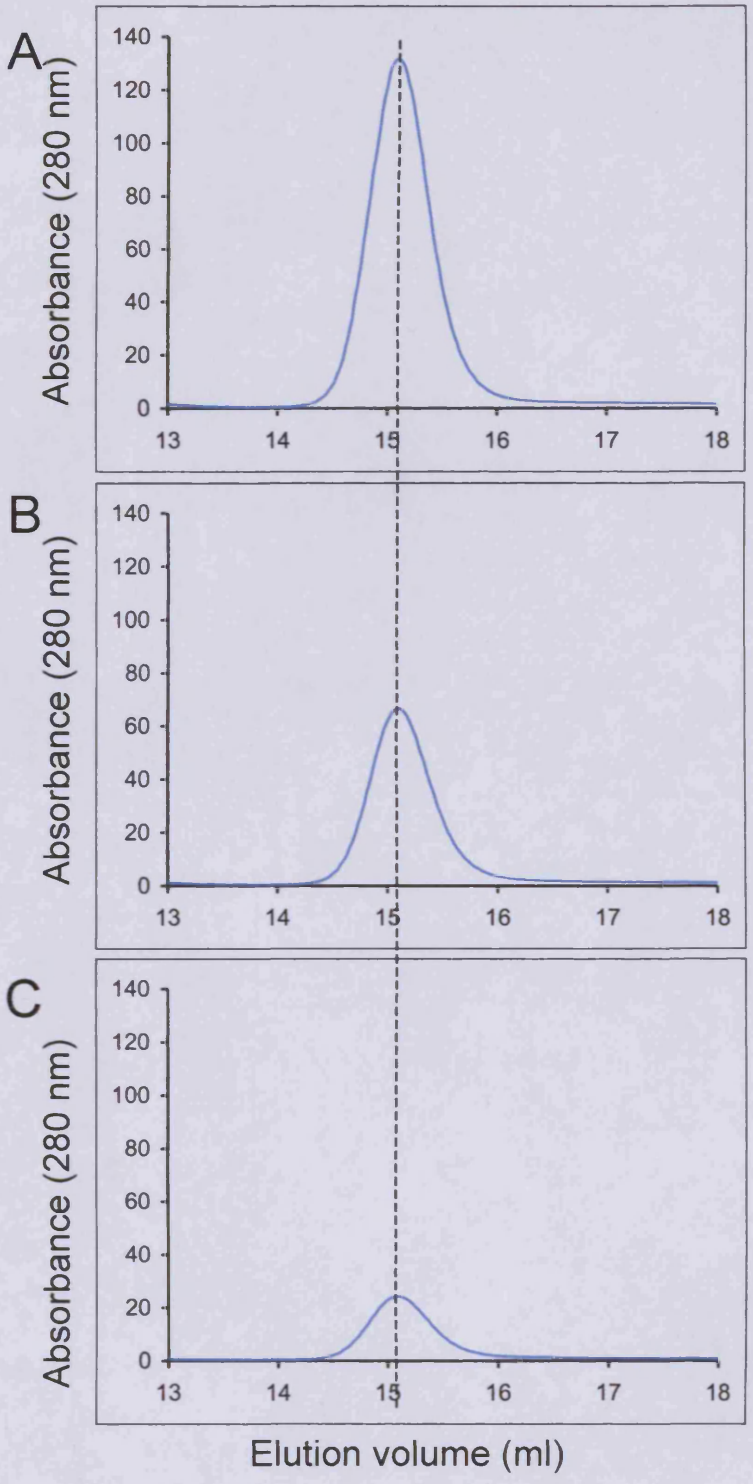


Figure 4.13 The influence of EDTA on ISP^{S250A} SEC profile. The SEC profile ISP^{S250A} was determined at (**A**) 50 μM, (**B**) 25 μm and (**C**) 10 μM of ISP^{S250A} in the presence of 50 mM EDTA. The peak elution volume in each case was 15.1 ml,

exchanged into buffer to remove the majority of EDTA. The buffer was supplemented with calcium and the fraction reanalysed by SEC (Figure 4.14). The result showed that the elution volume for ISP^{S250A} returned to 14.9 ml, the elution volume normally observed in buffer containing calcium.

4.5.3 Metal ions and the 3D structure of ISP^{S250A}

Structure determination of ISP^{S250A} confirms the presence of a metal ion in all six protomers located in an equivalent position to the high affinity calcium site typical within Bacillus ESPs (Figure 4.15). The coordination is similar to that in the ESPs but is less accurately defined in the lower resolution ISP^{S250A} structure than in the high resolution ESP structures. The coordination involves the main chain carbonyl oxygens of residues Ala97, Ser101 and Val103 and the side chain oxygen of Asp58 and Thr99. The sixth site is occupied by a water molecule but may become occupied after N-terminal removal where potentially either residue Ser19 or Glu20 flips around in the processed form to co-ordinate via its side-chain.

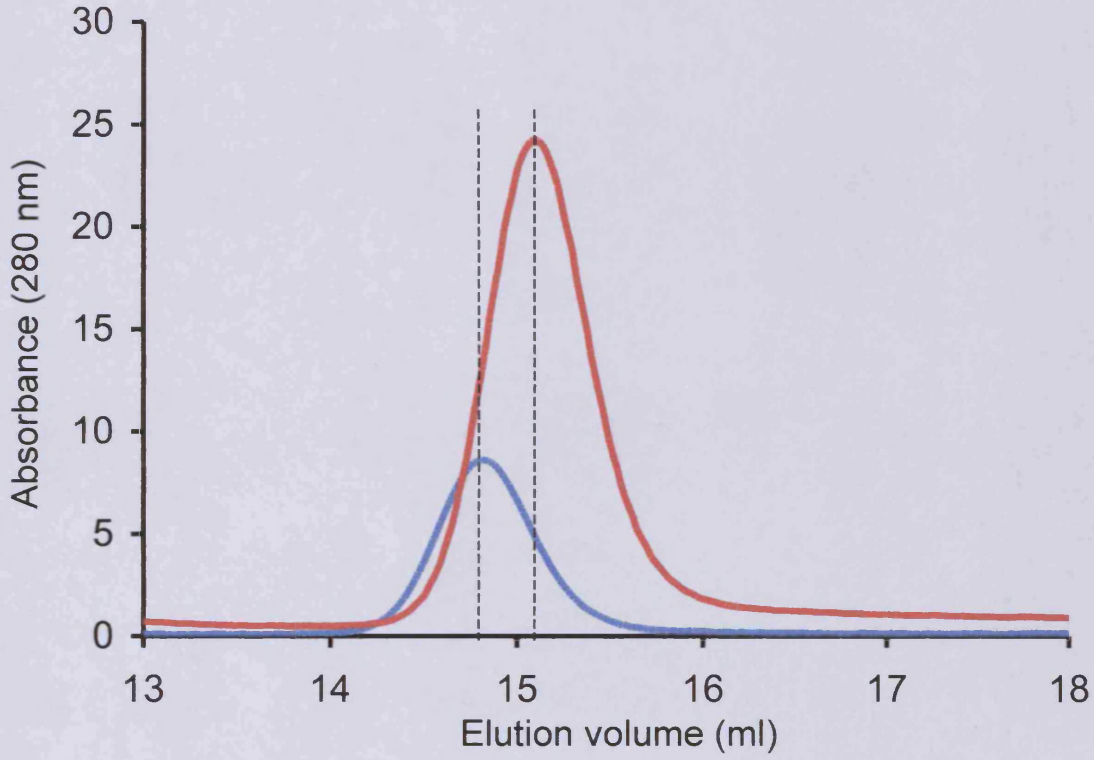


Figure 4.14 Calcium dependent SEC elution profile of ISP^{S250A}. The SEC elution profile of ISP^{S250A} was determined in the presence of EDTA (50 mM) (red). The peak that eluted at 15.1 ml was collected and exchanged into buffer (50 mM Tris-HCl (pH 8.0), 0.2 M ammonium sulphate) supplemented with 1 mM calcium chloride and reanalysed by SEC (blue). The main elution peak shifted to 14.9 ml.

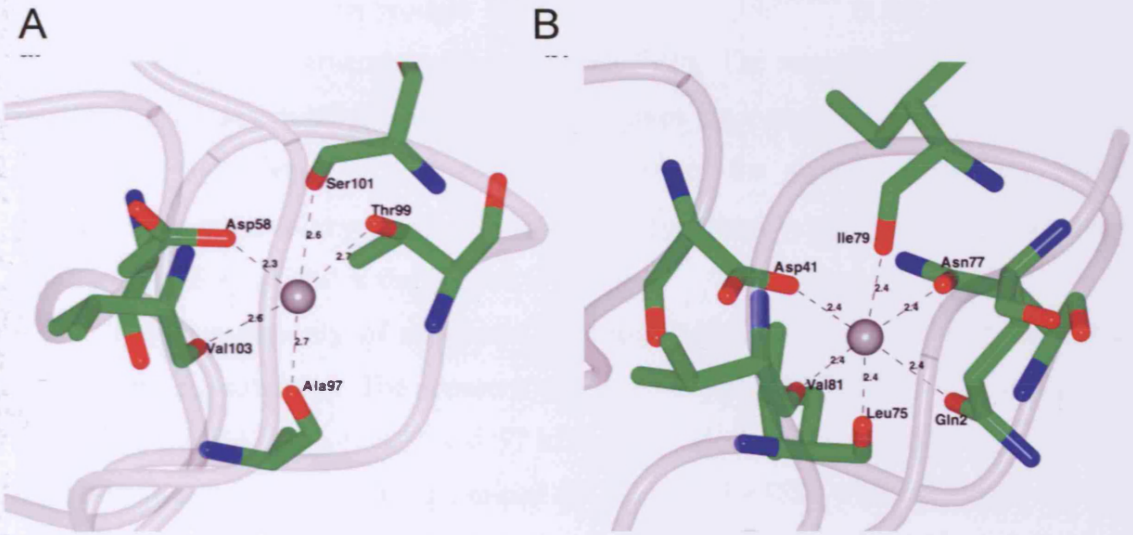


Figure 4.15 Structural representation of the high-affinity metal binding site (site A). (**A**) The calcium binding site for ISP^{S250A} showing the five available coordinating side chains. The sixth ligand position is a water. (**B**) Equivalent calcium binding site in BPN' (PDB 1TO2).

4.6 Discussion

The bacterial ISPs represent distinctive class of subtilisin-like proteases whose properties differ significantly from their closely related extracellular counterparts. Comparison of the amino acid sequence of ISPs with other bacterial ESPs reveal differences in their primary structure arrangement, with the classical prodomain region replaced by a short N-terminal extension and C-terminal directed dimerisation. The new crystal structure allows the study of an ISP structure for the first time. One of the most important revelations brought by the structure of ISP^{S250A} is the similarity of the ISP^{S250A} core protein structure with that of the ESPs. The superposition of ISP^{S250A} on its close ESP structural homologue, BPN' gives an r.m.s.d. of 1.26 Å over 260 structurally equivalent Cα positions in the core of the protein. Whilst the overall structure is highly conserved, the differences are restricted to a number of surface loops and the extensions of ISP at the N- and C-termini.

The vast majority of subtilases are monomeric. However, the ISPs appear to adopt a dimeric structure. The apparent molecular weight of ISP^{S250A} was determined to be 55 kDa by SEC (Figure 4.5) and 57 kDa using AUC analysis. This corresponds to molecular weights previously determined (54-60 kDa) for ISPs within different *Bacillus* species (Strongin *et al.*, 1979, Kurotsu *et al.*, 1982, Burnett *et al.*, 1986, and Koide *et al.*, 1986). The hydrodynamic volume therefore suggests that ISP either forms partially unfolded monomers or more compact dimers. The latter is more likely because the measured Mw is closer to that calculated for the dimer and CD (Figure 4.10) shows that ISP^{S250A} is folded. With the structure of ISP^{S250A} now available, it is clear that ISP^{S250A} does indeed form dimers. The C-terminal region, where sequence homology is low was shown to form the majority of the dimer interface (Figure 4.8). Most notably, the α-helix comprising residues 277-287 of one protomer cross over and interacts with the other protomer. As ISP^{S250A} formed dimers at all the concentrations tested, this suggests that the dissociation constant (K_d) is much lower than 10 μM, the lowest concentration tested.

The location of the dimer interface is separate from the active site regions of the protomers suggesting dimerisation does not play a role in the regulation of ISP activity or in substrate binding. Therefore dimerisation may have a function similar to the prodomain within the ESPs ensuring correct folding. However, instead of being removed similarly to the ESP's prodomain following activation, the ISP dimer interface

remains an integral structure that is thermodynamically stable. In contrast, removal of the prodomain traps the ESPs in a kinetically stable state that counteracts the low thermodynamic stability and prevents refolding upon denaturation.

The metal ion binding, specifically calcium binding is essential for stability and activity amongst the subtilisins (Bryan 2000). The majority of subtilisins bind one or more calcium ions. However, the role of calcium binding within the ISPs was inconclusive (Kurotsu *et al*, 1982, Koide *et al.*, 1986, Tsuchiya *et al.*, 1997 and Subbian *et al.,* 2004). It has been shown here that metal ion binding does play and important role in the structure and hence function of ISP. The X-ray crystal structure suggests that ISP^{S250A} is likely to contain a metal ion-binding site similar to site A found in the ESPs. The sixth coordinating residue conserved in the ESPs (Gln2) (Figure 4.14), is occupied by water in ISP^{S250A}. However as described in Chapter 5, (*Processing and Activation of ISP)*, ISP undergoes processing of the N-terminus close to the ligand binding site. This may allow structural changes and a sixth co-ordinating residue to be located near this site. The most likely residues are Ser19 or Glu20. In the crystal structure this region did not resolve well enough to positively confirm calcium as the ligand at this site. The lack of the sixth coordinating ligand may decrease the strength and therefore affinity of ligand binding. Should the K_d of calcium be high maybe due to the lack of the sixth coordinating ligand, this may move the calcium binding equilibrium towards dissociation rather than binding or dynamic exchange with the solvent. Further analysis is required to substantiate this using higher calcium concentration in the crystallography buffer for ISP^{S250A} and the analysis of a N-terminal truncated ISP variant where the sixth position may now be occupied by an ISP residue.

The decrease in molecular weight of ISP^{S250A} by approximately 5 kDa with the addition of EDTA (Figure 4.13) suggests that the protein has become either a partially unfolded monomer or a more compact dimer. Biophysical analysis using CD suggested that in the presence of EDTA, ISP^{S250A} was partially unfolded, supporting the former. Degradation of ISP^{S250A} was ruled out by analysing the SEC elution profile in the presence of EDTA followed by secondary analysis after EDTA removal and the reintroduction of calcium. The return to the native Mw suggests that calcium dissociation is a reversible process. This is in contrast to the mature ESPs where calcium removal is a major factor in their inability to refold (Bryan *et al.*, 1992).

The biochemical and biophysical data along with the first structure of an ISP allows analysis of the sequence features unique to ISPs amongst the subtilisins. This includes the formation of a unique dimeric structure through a well-defined interface. Determining the functional role of the dimer interface either in activity regulation or more likely folding will be crucial in discovering why the ISPs are dimers when all of the other subtilisins are monomeric. If the role is within folding, does dimerisation act in a similar mechanism to the prodomain. It has also been shown that the removal of a metal ion, most likely calcium, has an effect on secondary and tertiary structure of ISP^{S250A}. It will be important to clarify what metal ion binds ISP. Intracellular calcium ions are normally low or tightly controlled and are essential for the growth of the bacteria. Calcium is also an important in signalling especially during oxidative stress (Herbaud *et al.*, 1998), which may explain why ISP is found in stationary stage of the Bacillus lifecycle. Addressing these questions will bring us closer to deciphering the physiological importance and role of the ISPs.

Chapter 5
Processing and Activation of ISP from *Bacillus clausii*

5. Processing and activation of ISP from *B. clausii.*

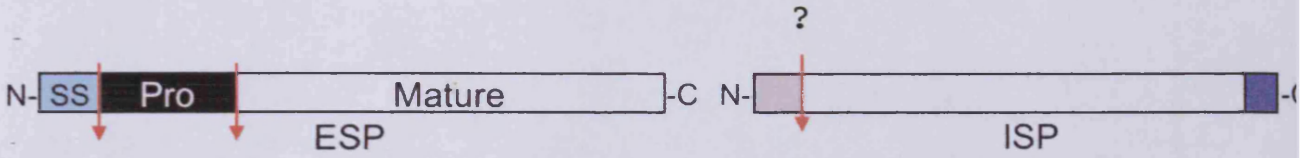
5.1 Introduction

The ISPs have relatively high overall primary sequence homology with the ESPs (approximately 40-50 % sequence identity). However, there is no homology observed between the N-terminal prodomain of the ESPs and N-terminus of the ISPs (Figure 5.1). Proteases are generally synthesised as inactive zymogens that require post-translational modification for activation. In most proteases it is the N-terminal pro-sequence that interacts with and blocks the active site of the protease. This common mechanism allows both temporal and spatial regulation of protease activity post-translationally so that the protease is only active when and where required. Thus unwanted effects of protein degradation are minimised.

The majority of subtilases have a common architecture, as typified by the ESPs from Bacillus species. These are initially synthesised as precursor inactive zymogens consisting of pre-pro-proteins. The pre-sequence (20-30 amino acids) acts as a classical signal sequence facilitating secretion of the protein across the membrane. The signal sequence is then removed by proteases in the membrane (Wong 1986). The pro-sequence (70-80 amino acids) acts as an intramolecular chaperone essential for the folding of ESPs to their mature conformation. The pro sequence subsequently binds back over the active site thus inactivating the enzyme. ESPs lacking a prodomain are able to fold, however the presence of the prodomain increases the rate of folding by a factor of $>10^7$ (Bryan *et al.*, 1995). The active protease is formed following autocatalytic cleavage and degradation of the pro domain. Therefore this pro sequence is essential for the formation of mature active protein and preventing unwanted intracellular protein degradation. The ISPs encode a 15-20 amino acid N-terminal sequence however it is thought to be too short to act as a classical pro-sequence involved in folding and inactivation. It has been suggested that ISP may undergo processing from a precursor protein (Stepanov *et al.*, 1977, Strongin *et al.*, 1978 and Strongin *et al.*, 1979) associated with a lag before the appearance of an active enzyme (Burnett *et al.*, 1986, Ruppen *et al.*, 1988). However, it has also been suggested that no precursor exists and that any processing of ISP may be an unwanted artefact of purification (Sheehan *et al.*, 1990 and Shiga *et al.*, 1993).

Within this chapter the role of the N-terminal extension and its function in modulating the activity of ISP is investigated.

A



B

```
ISP-B.clausii       ------------------------------------------------------------
ISP-B.cereus        ------------------------------------------------------------
ISP-B.anthracis     ------------------------------------------------------------
ISP-B.subtilis      ------------------------------------------------------------
ESP-BPN             MRGKKVWISLLFALALIFTMAFGSTSSAQAAGKSNGEKKYIVGFKQT------MSTMSAA 54
ESP-Savinase        --MKKPLGKIVASTALLISVAF---SSSIASAAEEAKEKYLIGFNEQEAVSEFVEQVEAN 55

ISP-B.clausii       -------------------------------------MRKFRLIPYKQVDKVSALS--EV 21
ISP-B.cereus        ----------------------------MNSSTANKQKGIQLIPFTVDKVVEQVN--EI 29
ISP-B.anthracis     ----------------------------MNSSTANKQKGIQLIPFTVDKVVEQVN--EI 29
ISP-B.subtilis      -------------------------------------MNGEIRLIPYVTNEQIMDVN--EL 22
ESP-BPN             KKKDVISEKG---GKVQKQFKYVDAASATLNEKAVKELKKDPSVAYVEEDHVAHAYAQSV 111
ESP-Savinase        DEVAILSEEEEVEIELLHEFETIPVLSVELSPEDVDALELDPAISYIEEDAEVTTMAQSV 115
                                                          :.:   .       .:
```

Figure 5.1 Primary structure features of ESPs and ISPs. (A) Primary structure schematic of ESP (left) and ISP (right). SS refers to signal sequence, pro to prodomain. The red arrows indicated known proteolytic processing cleavage locations for the ESPs and potential processing location for ISP. (B) Partial sequence alignment of the N-terminal region of ISPs from various *Bacillus* species and two representatives of the ESPs (BPN' from *B. amyloloquifaciens* and Savinase from *B. lentus*). The LIPY(F) sequence is highlighted, red.

5.2 Activation of ISP from *Bacillus clausii*

As indicated in the introduction it is still debated whether ISP undergoes processing. It is therefore important to determine if processing of ISP occurs and whether it relates to the activity of ISP. The cleavage of pro sequences, and therefore activation, occurs via three potential mechanisms: (1) complete self-processing, (2) processing with partial assistance from an additional enzyme and (3) fully assisted processing. In the case of subtilisins complete self-processing of the prodomain occurs. For example, subtilisin BPN' autocatalytically cleaves its prodomain between the tyrosine 77 and alanine 78 releasing the mature enzyme (Shinde and Inouye 2000). Upon release of the prodomain the mature subtilisin undergoes structural rearrangement at the new N-terminus. This completes the high affinity calcium-binding site and creates a protease that is kinetically trapped in the mature conformation and a high resistance to unfolding. This feature most likely evolved because of the harsh environment into which these enzymes are secreted (Siezen and Leunissen 1997).

To assess whether ISP undergoes processing, ISP was expressed and rapidly purified, using a His SpinTrap (GE Healthcare). Following purification, SDS-PAGE was used to visualise any change in Mw of ISP during incubation. This revealed the band corresponding to ISP was processed to a slightly smaller size product in a time dependent manner (Figure 5.2A). The smaller band was relatively stable with no further digestion products observed after 10 hours incubation at 25°C. To assess if activity was linked to the reduction in the size of ISP the aliquots taken at each time point were also assayed with the chromogenic substrate Suc-FAAF-pNA. This showed that the larger band exhibited no activity towards the substrate (Figure 5.2B). However, an increase in hydrolysis of the substrate was observed between 3 and 5 hours, which coincided with the appearance of the smaller band shown in Figure 5.2A. The final processed protein appeared relatively stable with no loss of activity or further degradation even after 10 hr. The two bands were sent for N-terminal sequencing at Alta Bioscience, Birmingham University. The N-terminus of the larger species was determined to be M-R-K-F-R, identical to that expected for the full length ISP (Figure 5.2C). There was a suggestion that there may be multiple forms present: one full length and one with the first methionine absent. However, it is not uncommon for N-terminal methionine to be removed in many proteins and is often important for their stability (Liao *et al.*, 2004).

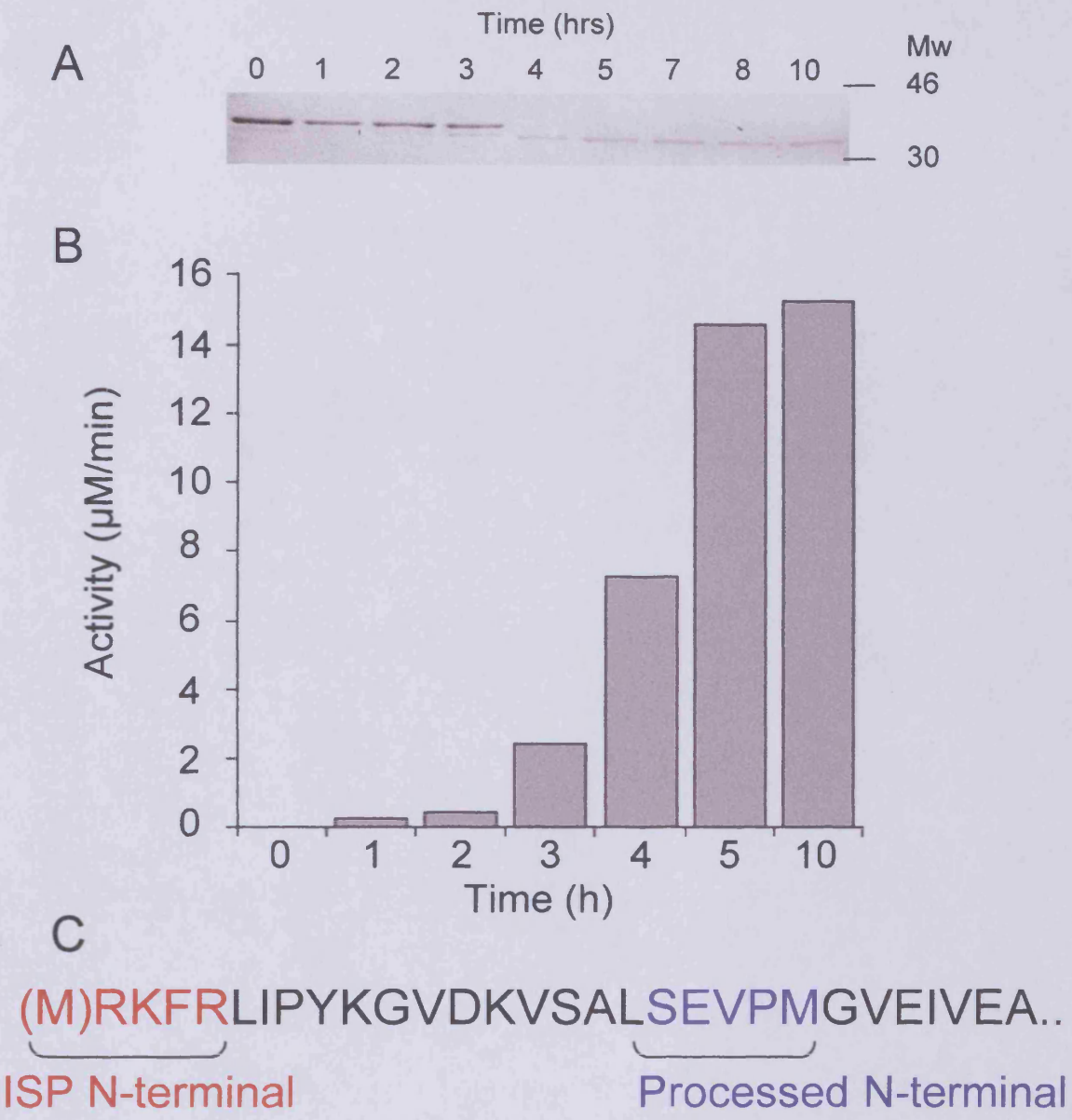


Figure 5.2. Processing and activation of ISP. **(A)** Proteolytic processing of ISP as analysed by SDS-PAGE. **(B)** Activity was measured using the chromogenic substrate Suc-FAAF-pNA over the timeframes shown in the graph. **(C)** N-terminal sequences of the two main products of ISP processing as determined by N-terminal sequencing analysis.

The N-terminal sequence of the smaller stable product was determined to be S-E-V-P-M, which corresponded to removal of the first 18 residues from the N-terminus of ISP. These results suggest that processing of ISP activates the enzyme.

To investigate if ISP itself was responsible for processing of the N-terminal extension, ISP^{S250A} was used as a substrate for the active form of the protease. The ISP^{S250A} substrate was processed to a smaller product by the active ISP (Figure 5.3A). N-terminal sequencing of the larger band matched that expected for ISP^{S250A} (MRKFR). Importantly, the N-terminal sequence determined for the digested form was SEVPM (Figure 5.3B); this matched that observed for the processed ISP. As shown in Figure 5.4, the target bond Leu18-Ser19 is far removed from the active site residues suggesting that processing may be an inter- rather than intra-molecular process.

While N-terminal sequence analysis confirms processing of the N-terminal extension is occurring during activation, it is not clear if there is also removal of residues at the C-terminus. Utilising the hexa-histidine tag present at the C-terminus of ISP variants, the presence of the C-terminus was probed by Western blotting using an anti-HisTag antibody. SDS-PAGE confirmed that there was only one primary product of ISP^{S250A} processing (Figure 5.5A). Western blotting confirmed that the HisTag was present during initial processing but was removed on prolonged incubation with active ISP (Figure 5.5B). The C-terminal regions of ISP and processed ISP were also probed by Western blotting. This revealed that the HisTag was present in the full length ISP (Figure 5.5C) but was removed after processing of the N-terminal extension (Figure 5.5D). The rate of cleavage was appeared comparatively slow compared with the assay presented in Figure 5.3A. However, the ratio of substrate ISP^{S250A} to active ISP enzyme was altered (1:40 to 1:100) due to the sensitivity of Western blotting so that ISP would not appear in the assay.

In order to determine the extent of the C-terminal processing of ISP and ISP^{S250A}, MALDI-TOF mass spectroscopy was used. The calculated Mw for the full length ISP^{S250A} was shown to be within 7 Da of those expected (Table 5.1). Since we have verified that the 18 amino acid N-terminal extension is removed from both processed variants, any reduction in their measured Mw must be due to loss of amino acids at the C-terminus. The calculated Mw for both ISP and ISP^{S250A} processed by ISP were lower that expected with a decrease of 1186 and 1183, respectively. This decrease

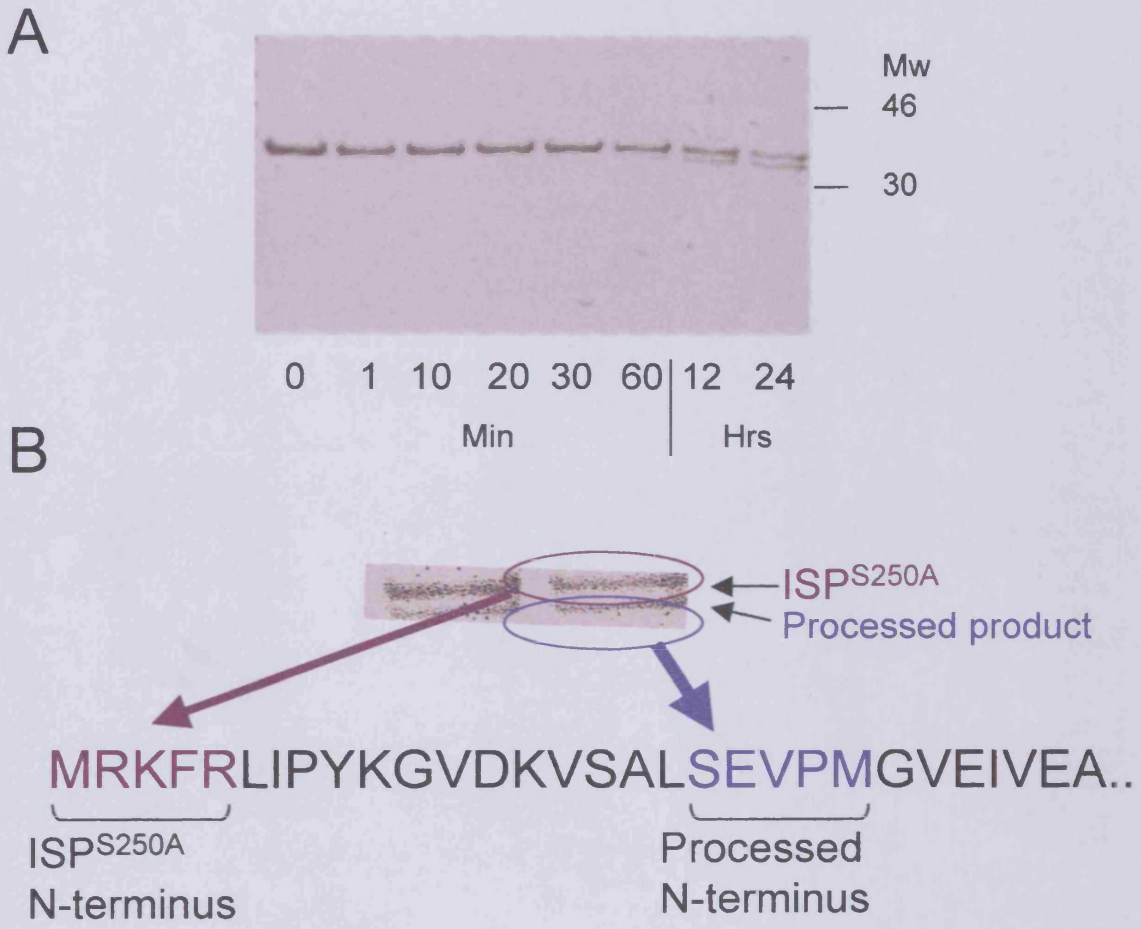


Figure 5.3. Proteolytic processing of ISP^{S250A} by active ISP. **(A)** The processing ISP^{S250A} was monitored by SDS-PAGE over the time scale indicated. **(B)** Schematic representation of the determined N-terminal sequence of two main bands. The N-terminal sequence of the larger product (purple) was MRKFR, which corresponds to the full length ISP^{S250A}. The processed product had the N-terminal sequence SEVPM (blue) indicating 18 amino acids have been removed.

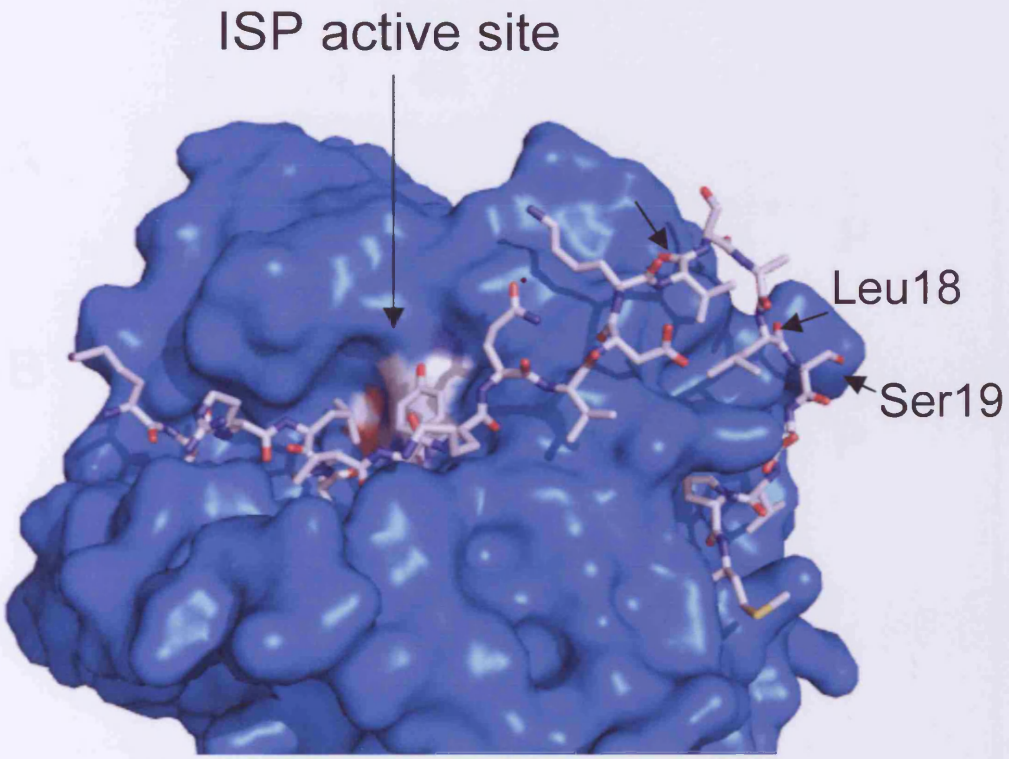


Figure 5.4 Structure of ISP^{S250A} and the location of the N-terminal extension. The surface of ISP^{S250A} (blue) is shown together with the positioning of the N-terminal extension (shown as sticks) and its relation to the active site (coloured CPK). The labeled residues indicate the cleavage point, as determined by N-terminal sequencing.

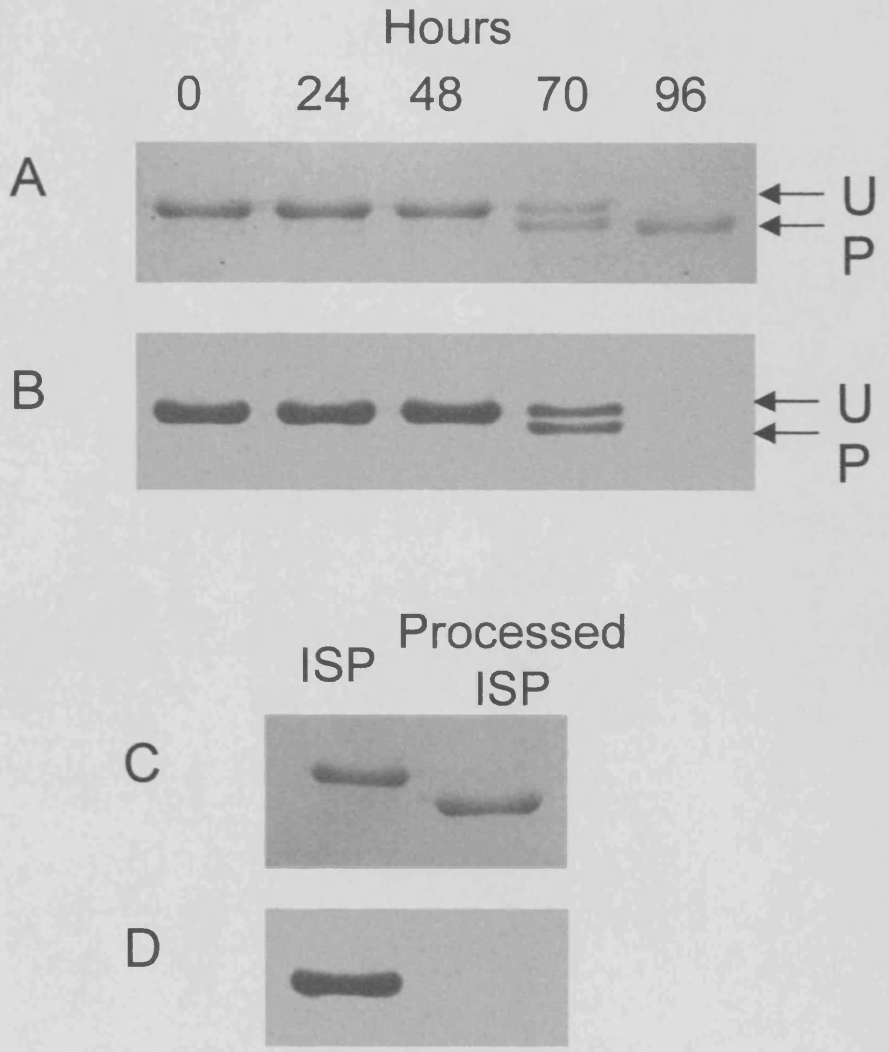


Figure 5.5 Analysis of ISP C-terminal processing. (A) SDS-PAGE showing ISP processing of ISP^{S250A}. (B) Western blot, using a HisTag antibody of ISP processing of ISP^{S250A}. (C) SDS-PAGE showing ISP before and after processing of the N-terminal extension. (D) Western blot, using a HisTag antibody of ISP before and after processing of the N-terminal extension. U and P represent unprocessed and processed products, respectively.

was determined to be the removal of the hexa His Tag and its linker (LE-HHHHHH) (1211 Da) and one lysine residue from the C-terminus.

Table 5.1 Molecular weights of ISP variants.

	Expected Mw (Da)	Measured Mw (Da)
ISP^{S250A}	34,775 /-Met	34,782
ISP^{S250A} processed by ISP	32,731	31,548
Processed ISP	32,747	31,561

The expected Mw of ISP variants are compared with those measured using MALDI-TOF mass spectroscopy. The measured Mw of ISP^{S250A} is within 7 Da, however processing of ISP^{S250A} by ISP and the processed ISP reveals a decrease of 1183 and 1186 Da, respectively, suggesting the loss of KLEHHHHHH from the C-terminus (1211 Da).

5.3 Quaternary structure of processed ISP.

It was previously shown that the unprocessed full length ISP^{S250A} had a native Mw of 55 kDa, suggesting that ISP is dimeric (Section 4.2). This dimeric oligomeric state was confirmed by the recently determined crystal structure of ISP^{S250A}. To investigate if processing affected the quaternary structure of ISP, the native Mw of processed ISP was calculated using SEC. Processed ISP eluted as a single peak at 15.1 ml (Figure 5.6), 0.3 ml later than ISP^{S250A} (Section 4.2.1). The native Mw calculated for processed ISP was 50 kDa. This, similarly to ISP^{S250A}, is larger than the expected molecular weight for the monomeric ISP (31.5 kDa) and smaller than the estimated molecular weight of a processed ISP dimer (63 kDa). This decrease in molecular weight may be due to a conversion to a partially unfolded monomeric state or

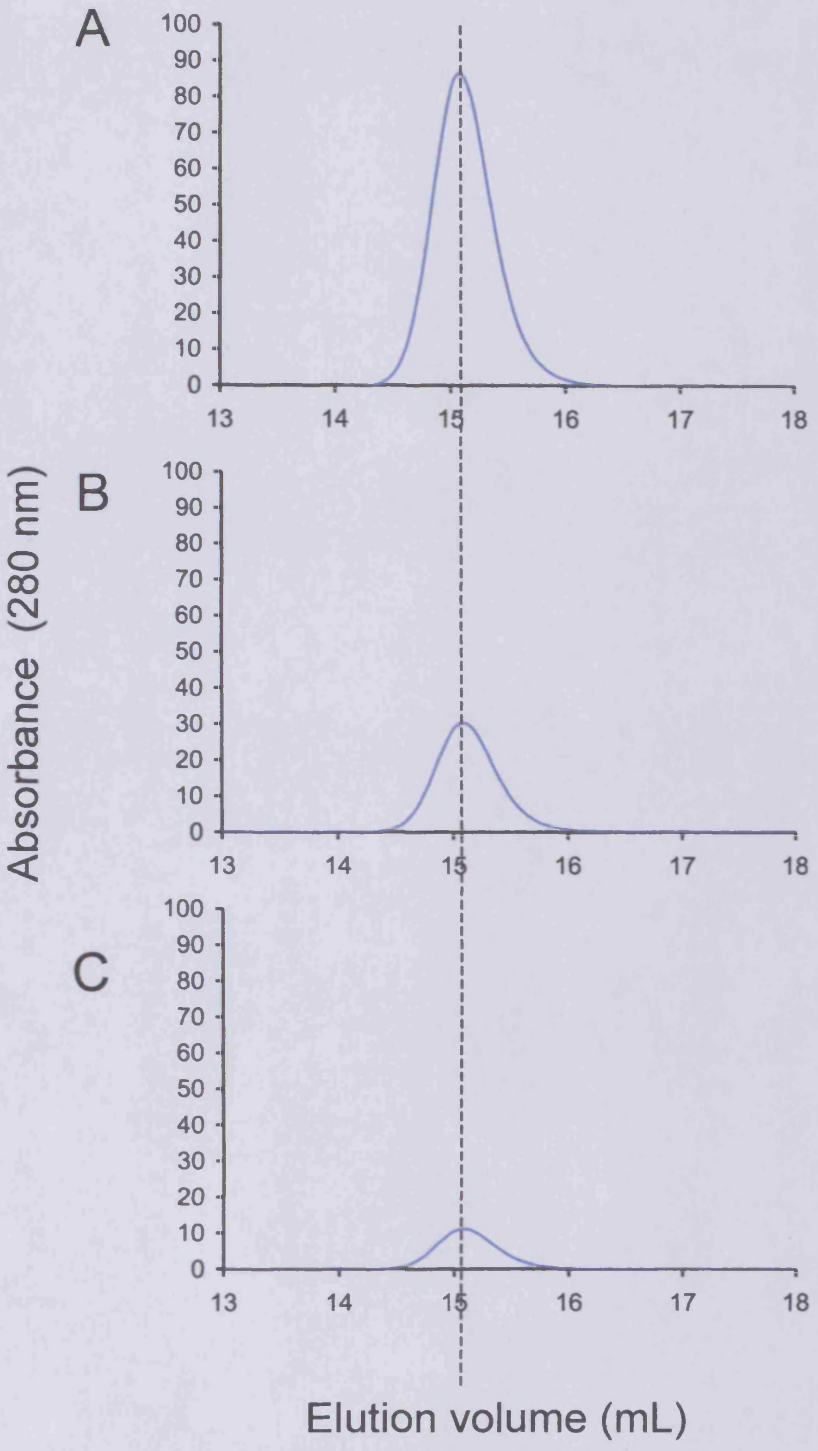


Figure 5.6 Analysis of processed ISP quaternary structure using SEC. SEC was performed in buffer (50 mM sodium phosphate (pH 7.5), 0.5 M ammonium sulphate and 1 mM calcium chloride) at (**A**) 50 µM, (**B**) 25 µm and (**C**) 10 µM concentrations of processed ISP.

simply the result of processing of 18 amino acids from the N-terminal and loss of approximately 2.2 kDa from each protomer.

5.4 Structural changes to ISP on processing

The CD spectrum for processed ISP was almost identical to that for ISP^{S250A} (Figure 5.7A) and characteristic of proteins with a subtilisin like fold (Eder *et al.*, 1993). The troughs at 208 and 222 are consistent with a mixed α-helix and β-sheet conformation. Deconvolution of the CD spectra using the program CDSSTR (Johnson 1999) suggested that ISP and ISP^{S250A} share similar α-helix (17 and 18, respectively) and β-sheet (23 and 24, respectively) percentages. A slight change was observed in the fluorescence intensity for processed ISP compared to ISP^{S250A} (Figure 5.7B) together with a blue shift in wavelength maxima. Whilst this suggests no major change in the overall structure of ISP on processing, confirming the observations with CD spectroscopy, a change in the environment surrounding one of the tryptophan residues is suggested. There may be a change in the environment of Trp128, which is located close to the anti-parallel β-sheet formed between Lys3-Leu6 of the N-terminal extension and main chain Gly122-Gly124. However, the processed ISP structure is required to elucidate what changes occur upon processing of the N-terminal extension.

5.5 Inhibition ISP by the N-terminal extension.

In Section 5.2, the activity of ISP was shown to increase following processing of 18 amino acids from the N-terminus of ISP. It was unclear whether ISP activity was inhibited due to the presence of the N-terminal extension. To investigate this, a peptide was designed constituting the first 17 amino acids (NΔ18) (lacking the N-terminal methionine) of ISP. This was synthesised by Activotec, Comberton, Cambridge, UK. Figure 5.8 shows that the NΔ18 acts as a potent inhibitor of ISP activity. The IC50 of the NΔ18 peptide was calculated to be 1 μM using the chromogenic Suc-AAPF-pNA as a substrate. A more detailed kinetic analysis using the same substrate at difference concentrations (0.5, 1, 2, 3 and 6 mM) revealed that NΔ18 bound with a K_i of 5.0×10^{-7} M. Both Lineweaver-Burk and Dixon plots (Figure 5.9 A and B) indicated a non-competitive type of inhibition with the rates of the reaction decreasing (V_{max} decreased

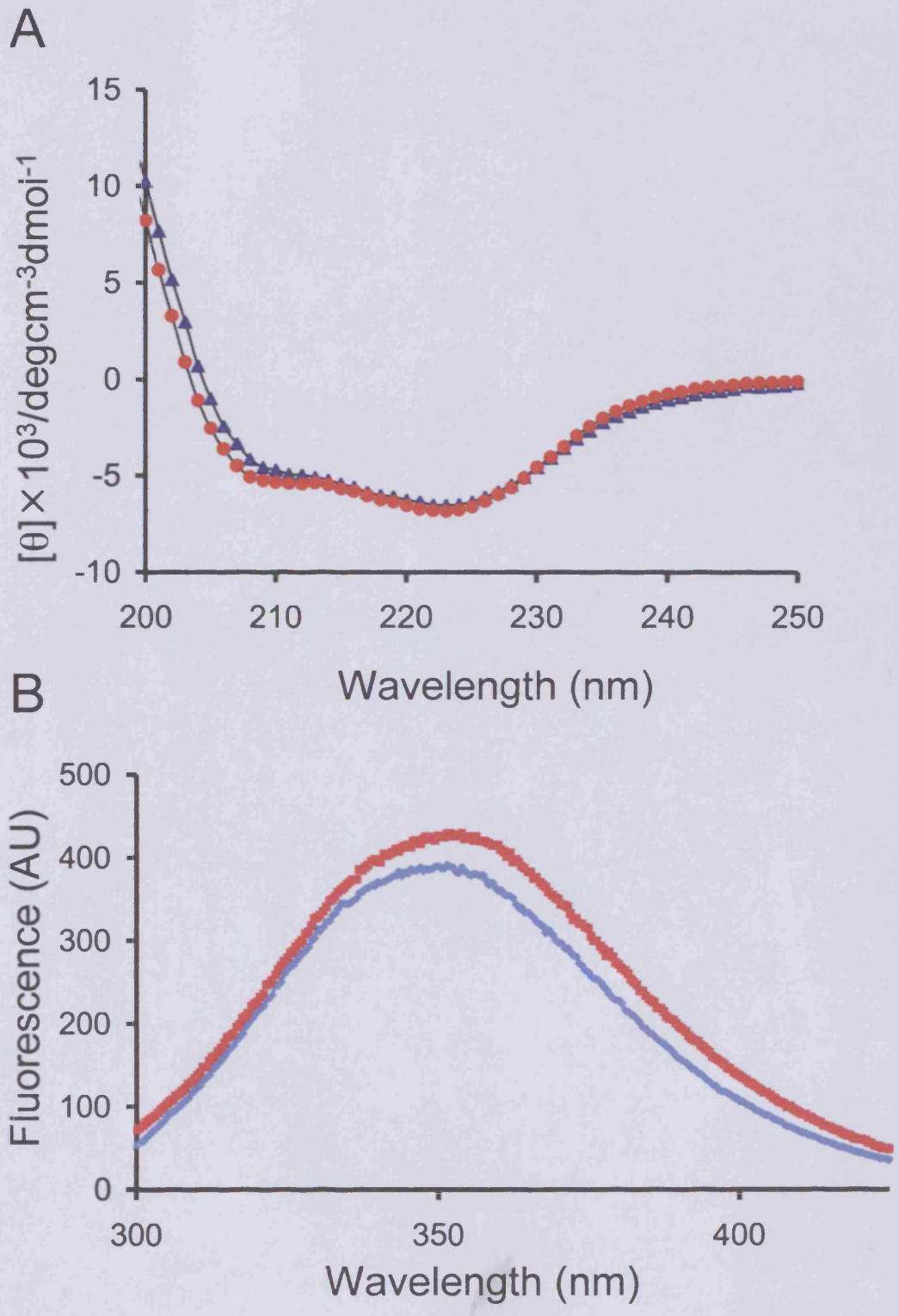


Figure 5.7 Structural changes to ISP after processing. **(A)** CD spectra and (B) fluorescence spectra of full length ISP^{S250A} (red) and processed ISP (blue).

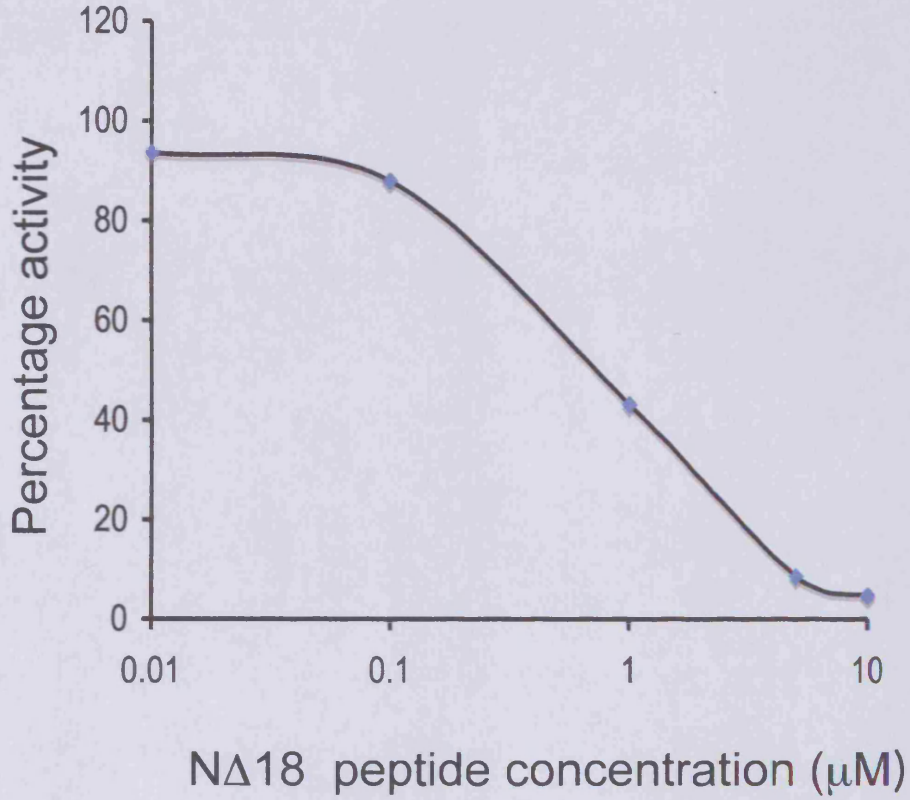


Figure 5.8 Inhibition of processed ISP by the NΔ18 peptide. The activity of processed ISP was determined using the chromogenic substrate Suc-AAPF-pNA (1 mM) in the presence of increasing concentrations of NΔ18 peptide.

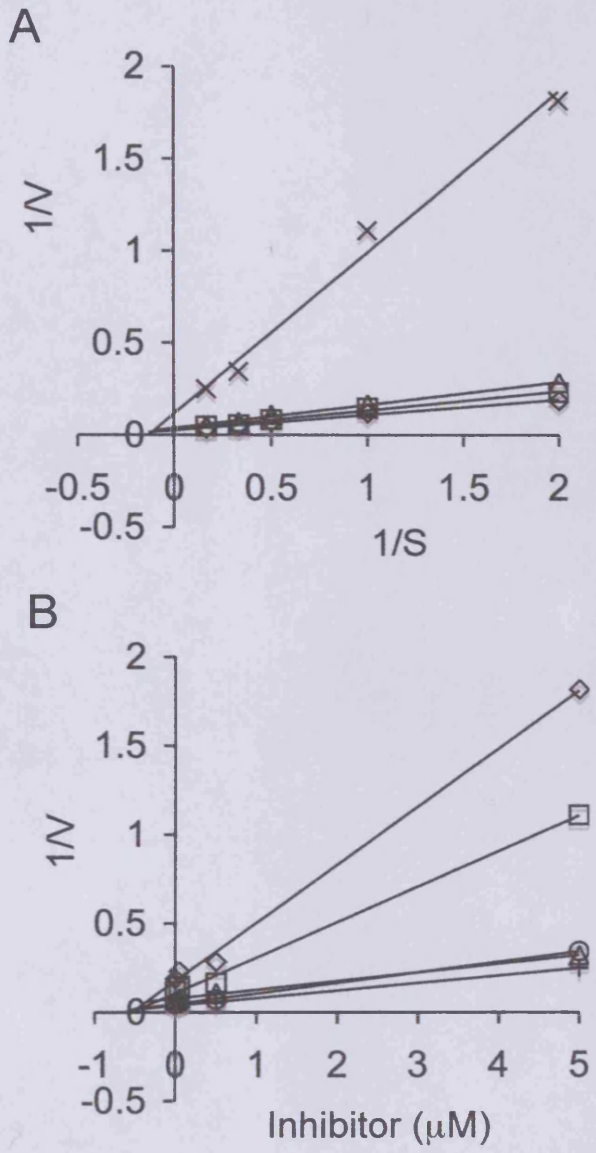


Figure 5.9 Inhibition of ISP by NΔ18 peptide; (A) Lineweaver-Burk plot of reciprocal of initial velocities versus reciprocal of substrate concentration in the absence (◇) and presence of 0.05 μM (□), 0.5 μM (△) and 5 μM (×) NΔ18 peptide. (B) Dixon plot at fixed Suc-AAPF-pNA substrate concentrations, 0.5 mM (◇), 1 mM (□), 2 mM (△), 3 mM (o) and 6 mM (+).

from 60 to 8 mMs^{-1} and k_{cat} decreased from 10 to 1 s^{-1}) with only a small change in K_M (from 6.4 mM without NΔ18 to 5.7 mM at the highest peptide concentration, 5μM).

5.6 Structural basis and mechanism of ISP inhibition by N-terminal extension

Within the new ISP^{S250A} crystal structure there is well-defined density for N-terminal extension. This shows residues 3-20 coiling over the core of the protein with residues Phe4-Leu6 forming the central strand of a 3-stranded antiparallel β-sheet with Gly122-Gly124 and Ser153-Gly155. In addition, residues Lys10-Val15 form a distorted antiparallel β-sheet with Tyr243-Ser247. The LIPY motif, residues 6-9, conserved throughout the ISPs clearly plays a key role in the inhibition process as these residues arch over the active site cleft and fit snugly against the core of the active site (Figure 5.10). The positioning of Pro8 is critical as it points away from the body of the protein and moves its potentially hydrolysable peptide bond well out of the reach of the catalytic serine residue (Figure 5.10). Leu6 and Tyr9 point towards the loop carrying the catalytic serine mutant (S250A) and form van der Waals contacts with His86. Tyr9 appears to contribute to the distortion of the proline bulge forming hydrophobic contacts with Leu246. The side chains of Arg5 and Ile7 point towards the poorly ordered surface loop 183-193. Ile7 corresponds to the traditional P1 position on the precursor and points into the S1 pocket. Residues 10-15 of the N-terminal extension form a β-sheet with residues 243-247 and this helps to disrupt the catalytic triad by displacing the active site serine/alanine by 1.6 Å with respect to the histidine (Figure 5.11). This is a novel mechanism for inhibition of subtilisin activity.

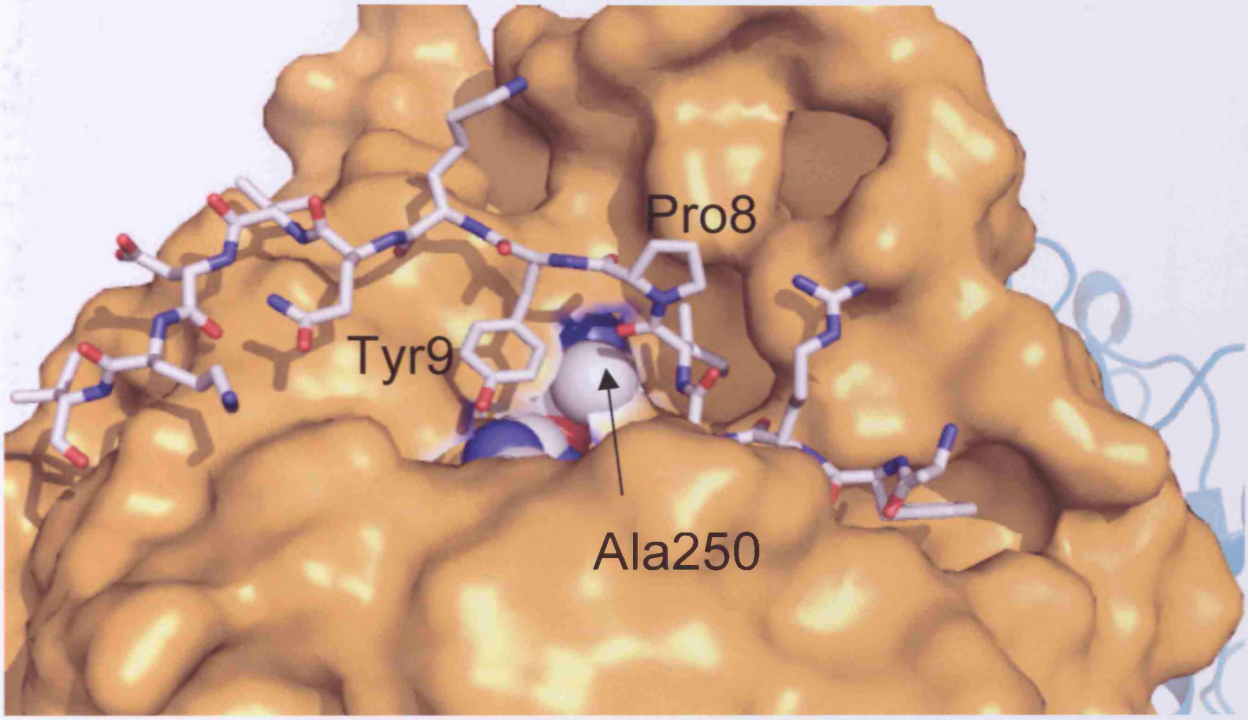


Figure 5.10 Structural role of N-terminal extension of ISP^{S250A}. The N-terminal extension, shown as sticks, tracks across the groove in the surface and passes directly over the active site. The interaction of Tyr9 and Pro8 take the target bond away from the Ala250.

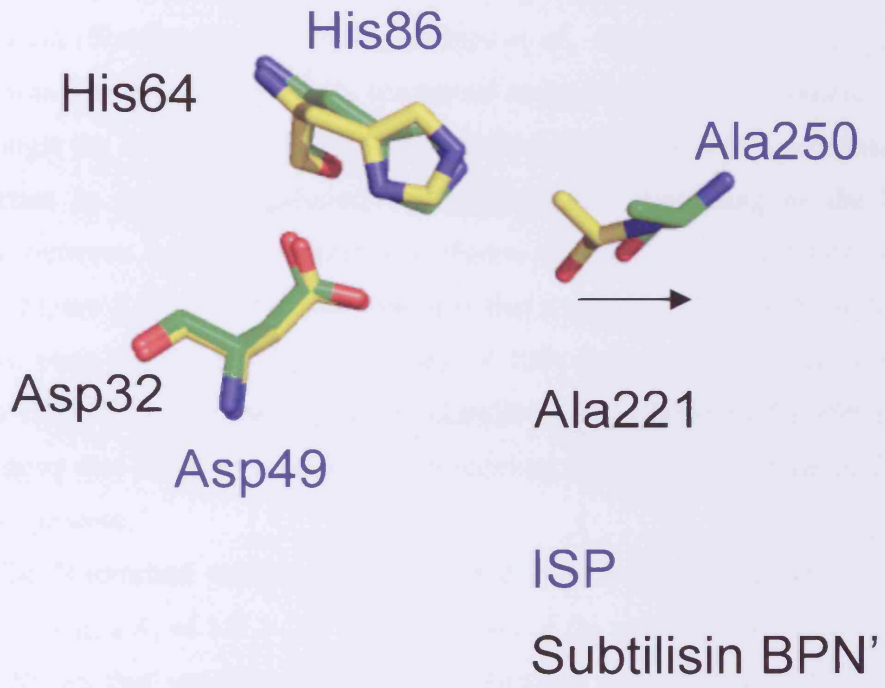


Figure 5.11 Structural comparison of the catalytic sites of ISP^{S250A} and an ESP. The catalytic residues of ISP^{S250A} (blue) and subtilisin BPN'(labelled blue and Black, respectively) are shown as stick representation. The Alanine of ISP^{S250A} is shifted by 1.6 Å away from the histidine residue compared to that of BPN'.

5.7 Discussion

The vast majority of proteases undergo post-translational processing that is vital for their maturation to an active conformation. This process controls the folding, cellular sorting and importantly inhibits the activity of the protease, avoiding untimely protein degradation. It was therefore surprising that previous reports suggest that ISP did not encode a prodomain or obvious folding region, and still folded to an active conformation (Sheehan *et al.*, 1990 and Shiga *et al.*, 1993). The sequence comprising the N-terminal extension is partially conserved suggesting it plays a common functional role amongst the ISPs. Here, it has been established that the ISP N-terminal extension is important in protease regulation and maturation. Processing of the N-terminal extension between Leu18 and Ser19 was shown to result in the formation of an active protease (Figure 5.2). It was initially thought that a contaminating protease from *E. coli* may have been responsible for processing of ISP, however active ISP processed the substrate ISP^{S250A} at the same position (Leu18-Ser19) identified for ISP maturation. This suggests that ISP is responsible for processing and therefore activation by an inter-molecular process.

The N-terminal extension, when added as a synthetic peptide, was shown to inhibit ISP with a K_i of 5.0×10^{-7} M. Analysis of the recently determined structure of ISP^{S250A} shows that residues 3-20 of the N-terminal extension in ISP^{S250A} coils back over the core of the protein restricting access to the active site (Figure 5.5). This also reveals that the conserved LIPY motif, residues 6-9, plays a key role in the inhibition process as these residues arch over the active site region and fit closely against the core of the active site. The positioning of the Pro8 is fundamental as it points away from the protein and directs the scissile bond out of reach of the catalytic serine residue. There is also shift in the position of Ala250 compared to other subtilisins. The main chain of the catalytic Ala250 mutant is displaced for the formation of a catalytic triad. This shift may contribute to inhibition of ISP with the catalytic triad clicking into an active conformation on maturation. This will need to be confirmed with structural studies on the mature enzyme. Therefore, the kink induced by Pro8 of ISP is a novel mechanism for the inhibition of subtilisins, and to our knowledge protein in general.

This is in contrast to the mechanism for binding of the prodomain in the ESPs which retains the target peptide bond in a position to allow autocatalytic processing and subsequent detachment of the prodomain, as highlighted in the structures for

unprocessed ESPs BPN' from *Bacillus amyloliquefaciens* and pro-Tk (Figure 5.12B) from the archea *Thermococcus kodakarensis* (Gallagher *et al.*, 1995 and Tanaka *et al.*, 2007). Furthermore interactions between subtilisins and classical inhibitors such as CI2 (Figure 5.12A) are normally restricted to the active site (Bode *et al.*, 1992 and Radisky *et al.*, 2002) whereas the N-terminal of ISP forms extensive interactions with regions of protein beyond the substrate binding site. These interactions distinct from the active site may prove critical initiating the binding of the N-terminal extension or reinforce those interactions at the active and substrate binding sites. All of these have some common elements in their mode of inhibition, in that a peptide chain lies across the active site, preventing access of potential substrates to the active site. However the nature of the inhibition varies significantly.

Surprisingly, the enzyme kinetics in the presence of the NΔ18 peptide suggested a non-competitive mechanism for inhibition, instead of the expected competitive inhibition. This may suggest that the NΔ18 peptide is not binding at the active site but elsewhere and causing inhibition remotely by altering the structure of ISP substrate binding site. However, the shift of the active site alanine by 1.6 Å away from the histidine residue may account for the non-competitive mechanism of inhibition. The movement of the alanine would interrupt the deprotonation of what would be the active site serine and thus decrease K_{cat} rather than K_M. Therefore the structure of the NΔ18 ISP would be required to determine if the alanine reverts to normal position, more similar to that found in active subtilisins.

Further analysis is required to determine the exact mechanism for the NΔ18 peptide inhibition of ISP. This may be possible by determining the structure of ISP with the synthetic NΔ18 peptide bound for comparison with the full length ISP structure. It would also be of interest to produce synthetic peptides sequentially truncating the NΔ18 to determine which regions are critical for inhibition.

Using the recently determined structure of ISP^{S250A}, the secondary cleavage of the unstructured artificially added C-terminal His Tag and Lys321 residues is unlikely to influence the function of ISP. However it is clear from the biochemical data presented in this chapter in conjunction with the structure of ISP^{S250A} that processing at the N-terminal is crucial for the activation of ISP. The autocatalytic cleavage of the N-terminal extension is unlikely to be the mechanism for ISP activation, as occurs in the ESPs. It is clear that ISP itself is responsible for precise processing of the N-terminal

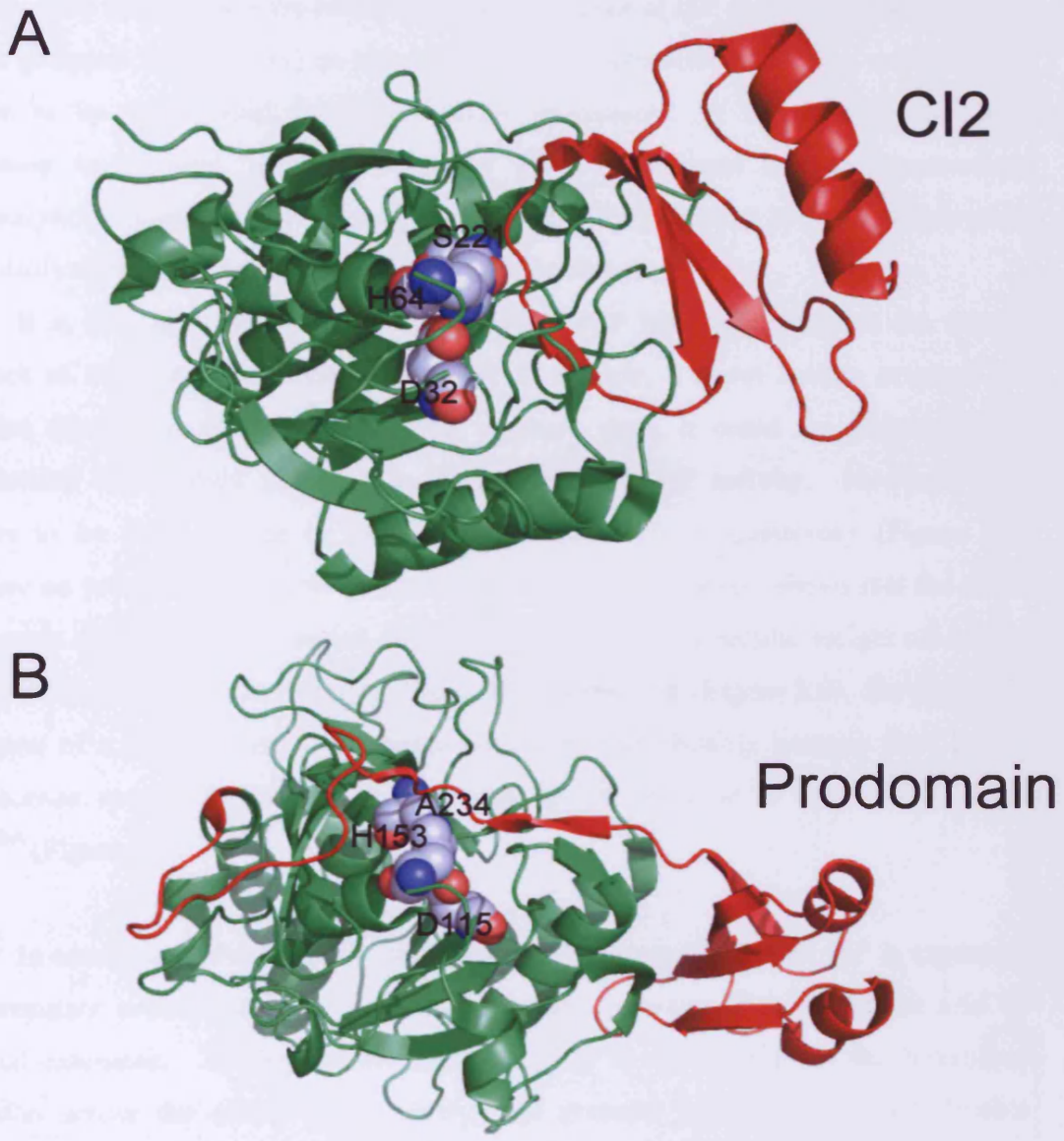


Figure 5.12 Structure of ESP inhibition. (A) The BPN' complexed with CI2 inhibitor (PDB 1TO2). (B) The unprocessed form of the archeal pro-Tk *Thermococcus kodakarensis* subtilisin (PDB 2E1P). Active site residues are shown as spheres.

extension (Figures 5.2 and 5.3) but this must be an inter rather than intra molecular process. This may explain the relatively slow activation of ISP *in vitro* (Figure 5.2) and *in vivo* (Ruppen *et al.,* 1988) as one would expect auto activation via a unimolecular process to be more rapid than bimolecular processing. It is still not clear how processing is initiated. The structure of ISP^{S250A} suggests that an intermediate autocatalytic processing event is unlikely due to the kink induced by Pro8, which takes the hydrolysable bond out of reach of the catalytic serine.

It is also apparent that the processing of ISP has little effect on the overall structure of ISP. As described earlier, ISP is dimeric, a novel feature amongst the subtilase family. In the absence of any contrary data, it could be postulated that dimerisation itself could play a role in regulation of ISP activity. However there appears to be little change in the tertiary (Figure 5.7) or quaternary (Figure 5.6) structure on processing. Furthermore, analysis of the 3D structure shows that the dimer interface is distant from the active site (Figure 4.2). The molecular weight calculated using SEC is larger than that expected for a monomeric ISP (Figure 5.6). However, the formation of a partially unfolded monomeric state was unlikely because the CD and fluorescence spectra for processed ISP were almost identical to that of full length ISP^{S250A} (Figure 5.7).

In conclusion, the biochemical data presented here shows that ISP is expressed as a precursor protein activated via inter molecular cleavage of the 18 amino acid N-terminal extension. The regulation of ISP activity is via binding of the N-terminal extension across the active site inhibiting the protease by taking the hydrolysable peptide bond out of reach of the catalytic mechanism by the formation of a proline bulge. This supports the hypothesis that ISP activity is regulated post-tranlationally.

Chapter 6

Characterisation of the Enzymatic Activity of ISP

6. Characterisation of the enzymatic activity of ISP

6.1 Introduction

The serine proteases are a ubiquitous family of enzymes involved in blood clotting (Davie *et al.*, 1991) and protein digestion for nutrition (Richter *et al.*, 1998). The ISPs are closely related to the highly studied ESPs (40-50% sequence identity). However, their physiological function is currently unknown, despite the fact that they are thought to constitute approximately 80% of the intracellular protease activity in Bacilli (Burnett *et al.*, 1986 and Orrego *et al.*, 1973). Given that ISPs share a catalytic mechanism with the ESPs, their location within the cell suggests that the ISPs are unlikely to share the low substrate specificity of the scavenging ESPs. Intracellular expression of an ESP has been shown to be detrimental to cellular survival (Subbian 2004) whereas, as shown in Chapter 3, intracellular expression of an ISP is well tolerated. It is therefore hypothesised that the ISPs must be controlled post-translationally via a narrow substrate range or by other regulatory factors. The ESP activity is controlled within the cell via an intramolecular chaperone that is only autocatalytically cleaved upon folding of the mature enzyme following secretion from the cell (Shinde and Inouye 1995). As shown in Chapter 4, ISPs may be controlled post-translationally via inhibition by the N-terminal 18 amino acids, which is later processed releasing active enzyme. However this process is not thought to be autocatalytic and occurs via intermolecular cleavage.

Initially, the ISPs were thought to play a role during bacterial sporulation and protein turnover (Hageman *et al.*, 1973). However, mutant Bacillus species lacking the gene encoding ISP were still capable of sporulation (O'Hara *et al.*, 1990). To date, the only report proposing role for ISP suggests ISP is involved in the turnover of regulatory proteins, such as Clp-1 and EF-Tu. These proteins were 'pulled out' of a degradomic assay amongst many proteins that were degraded by ISP. They suggested that ISP did not digest cellular proteins with a broad specificity, but targeted putative recognition sites such as XXAA with small amino acids at P1 (Lee *et al.*, 2004). If the ISPs were involved in the precise processing of proteins through defined proteolysis, a narrow substrate range would be expected. However, if the ISPs were involved in broad ranging protein degradation, similar to the ESPs, such as the turnover of misfolded or denatured protein then the substrate specificity would possibly target regions of the

protein not normally accessible in native conformation such as hydrophobic residues within the core.

To address the role of substrate specificity in defining the activity of ISP, the proteolytic digestion of various protein and peptide substrates will be investigated. The pH dependant activity of ISP will also be determined to compare with the known pH profile of the ESPs and correlate activity with cellular pH levels.

6.2 Characterisation of ISP proteolytic activity towards protein substrates

One hypothetical physiological role for ISP is that they may be involved in the processing of denatured or misfolded protein within the bacterial cell. If this were the case we would expect ISP to show minimal activity towards protein in native conformation and have a higher proteolytic activity towards protein that is denatured or misfolded.

To test this hypothesis the processed ISP was assayed against three protein substrates: Malate dehydrogenase (MDH), lactate dehydrogenase (LDH) and bovine serum albumin (BSA). The substrates were chosen as ubiquitous proteins with distinct functions not linked with ISP. The proteolytic digestion of the substrates was assayed before and after heating to 95°C to assess whether a natively folded or denatured protein would act as the better substrate. To allow comparison with the ESPs, Savinase from *Bacillus lentus* was used as a representative of the ESPs. Savinase is a scavenging protease with a broad specificity (Siezen *et al.*, 1997) and is therefore expected to degrade most proteins.

Native BSA was a poor substrate for active ISP with very little digestion observed (Figure 6.1 A). In contrast proteolytic degradation of denatured BSA was markedly higher (Figure 6.1 B). Proteolytic degradation of denatured BSA by Savinase was also much higher than that observed for native BSA. However, the overall level of proteolysis of BSA, both native and denatured conformations, by Savinase was much higher compared to ISP.

The denatured form of MDH acted as a good substrate for active ISP, with almost complete digestion of MDH observed after 60 minutes (Figure 6.2 B). In contrast, native MDH was a poor substrate with little or no digestion observed (Figure 6.2 A). Both forms of MDH were good substrates for Savinase (Figure 6.2 C & D), with denatured being digested slightly more efficiently and rapidly.

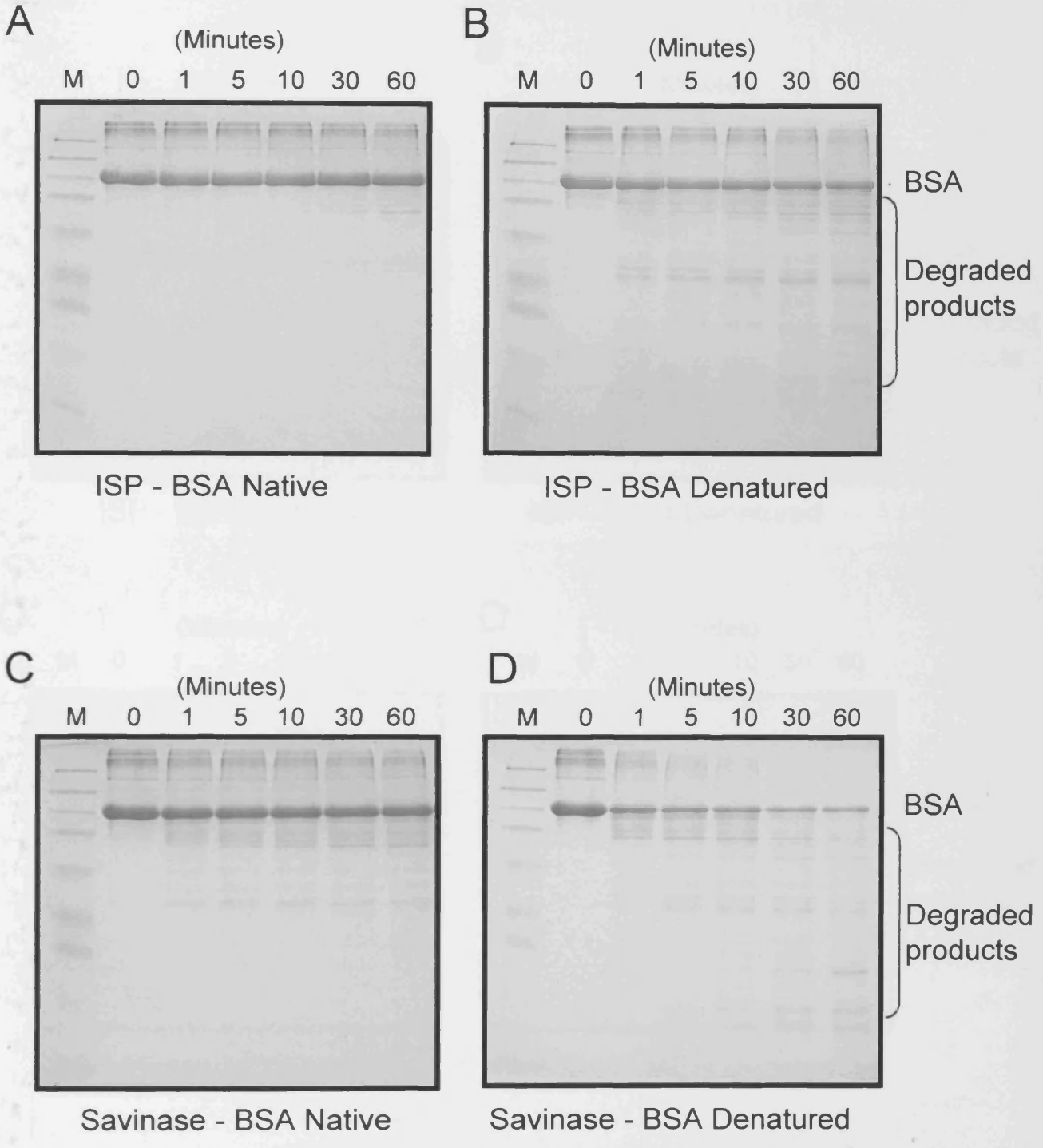


Figure 6.1. Proteolysis of bovine serum albumin (BSA) by ISP (A & B) and Savinase (C & D). The substrate BSA was assayed in its native form (A & C) and after denaturation at 95°C for 5 minutes (B & D). Fractions were taken at time intervals as indicated and digestion patterns analysed by SDS-PAGE.

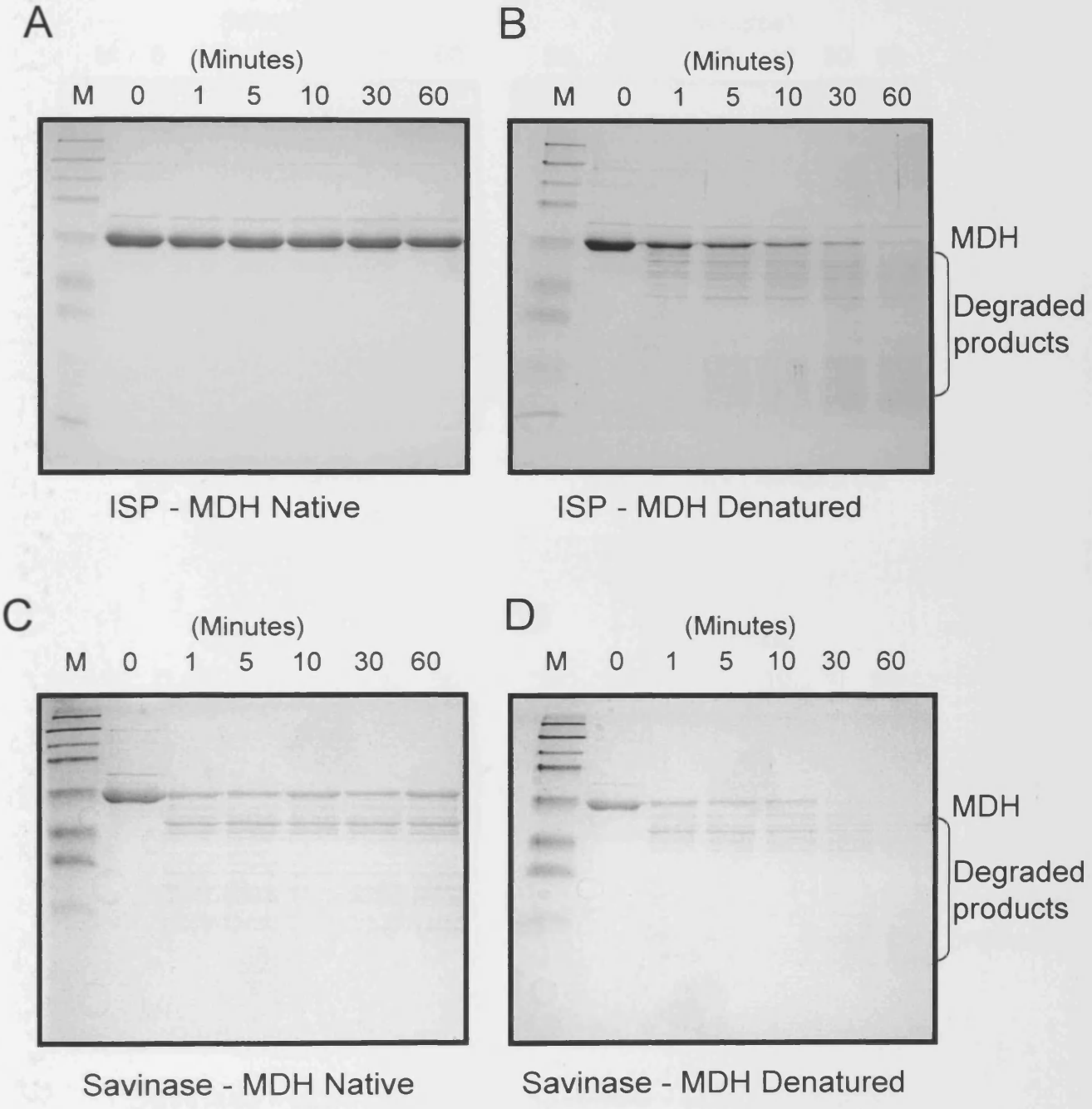


Figure 6.2. Proteolysis of malate dehydrogenase (MDH) by ISP (A & B) and Savinase (C & D). The substrate MDH was assayed in its native form (A & C) and after denaturation at 95°C for 5 minutes (B & D). Fractions were taken at time intervals as indicated and digestion patterns analysed by SDS-PAGE.

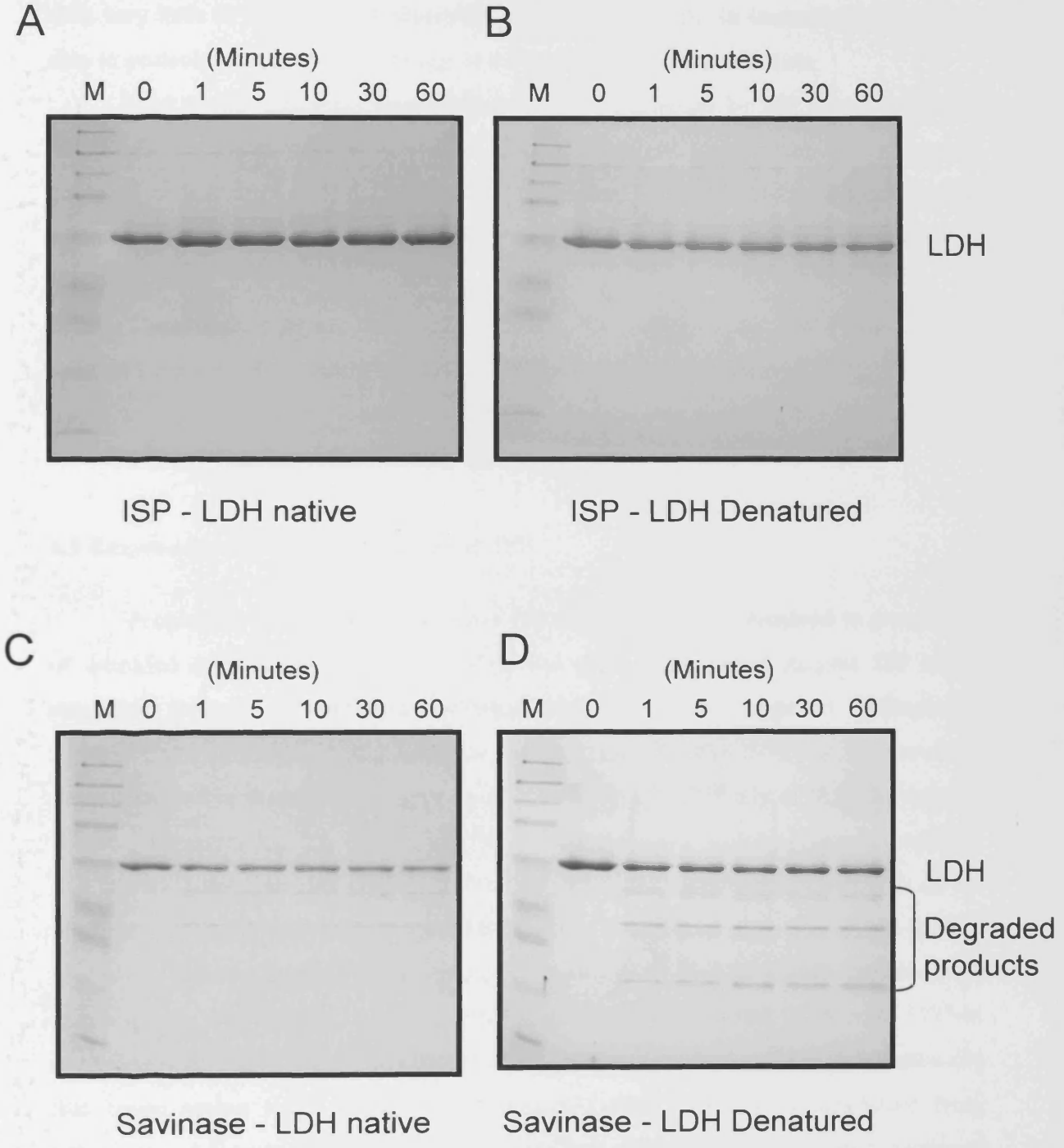


Figure 6.3. Proteolysis of lactate dehydrogenase (LDH) by ISP (A & B) and Savinase (C & D). The substrate LDH was assayed in its native form (A & C) and after denaturation at 95°C for 5 minutes (B & D). Fractions were taken at time intervals as indicated and digestion patterns analysed by SDS-PAGE.

Both native and denatured forms of LDH were a poor substrate for active ISP with very little or no digestion observed (Figure 6.3 A & B). In contrast Savinase was able to proteolyse denatured LDH and to a lesser degree native substrate.

Due to the numerous degradation products generated by ISP when denatured MDH and BSA were used as substrates it was decided not to determine the processing positions as planned. If fewer bands were observed, MALDI-TOF mass spectroscopy was to be used to help establish where ISP hydrolysed substrate to gain information on ISP specificity.

These data support the hypothesis that ISP has a higher specificity towards unfolded compared to native protein. The assays were performed using the same concentration of ISP and Savinase. Therefore the data also suggests that ISP has lower overall activity than Savinase for these particular substrates.

6.3 Enzyme kinetic characterisation of ISP

Proteolysis experiments suggested that the ISP could be involved in processing of unfolded or misfolded proteins within the cell. This would suggest ISP might recognise features of unfolded/misfolded proteins such as exposed hydrophobic residues and may therefore play a role in substrate specificity of the ISPs. To provide a more quantitative and detailed analysis of the substrate specificity of ISP, the enzyme kinetic parameters K_M, k_{cat} and k_{cat}/K_M were measured using various synthetic peptides. The chromogenic peptides (Figure 6.4) consisted of four amino acids linked to an N-terminal succinyl group, to aid in solubility, and a C-terminal para-nitroanilide (pNA). Hydrolysis of the peptide bond linking the amino acids and release of the pNA chromophore (designated P1-P1') was measured at 405 nm using an UV/vis spectrophotometer. The active sites of most proteases contain individual subsites (S) that house amino acids from the substrate (P) (based on the nomenclature from Schechter and Berger 1967). The P1 site is defined as the site cleaved at its C-terminus and binds within the S1 subsite (Figure 1.6). The residues N-terminal of P1 are numbered P2, P3, and P4 and bind the corresponding S subsites. The P' sites are C-terminal to the scissile bond and are accommodated by the S' subsites. The number of residues that bind the enzyme subsites varies from one protease to another. The important subsites within the subtilisins are the S1 and S4 sites, which preferentially bind large hydrophobic residues (Gron *et al.*, 1992). The S2 site accommodates most

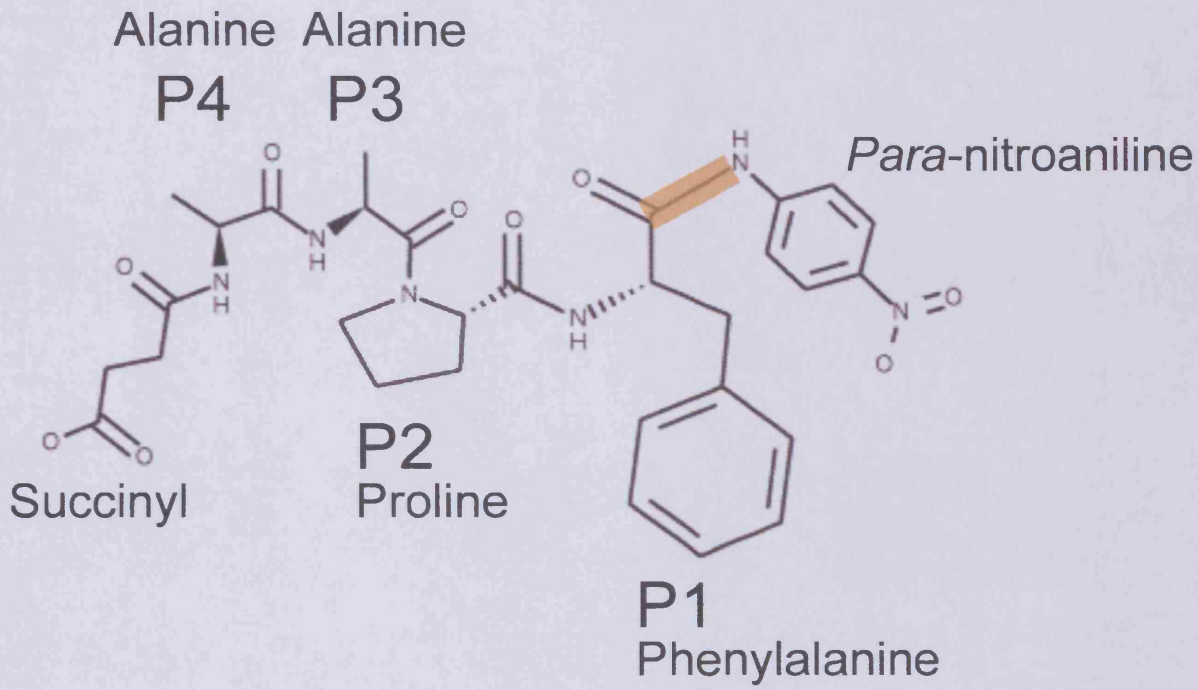


Figure 6.4. Structural representation of Succinyl-AAPF-pNA. An example of the peptides used to determine substrate specificity for ISP. The residues constituting P1-P4 and the chromophore (*para*-nitroaniline) are labelled. The scissile bond is highlighted orange.

residues as the side chain is directed towards the solvent and the shallow S3 site favours small amino acids (Ballinger *et al.*, 1995).

The substrates used in this assay were FAAF, AAPF, AAPNle, AAPNva, AAPL, AAPM, AAVA, AAPK, AAPOrn, AAPE and YVAD. The amino acids are given in standard single letter abbreviation except for the unnatural amino acids Nle (norleucine), Nva (norvaline) and Orn (ornithine).

The enzyme kinetic results were determined by fitting the initial rate of hydrolysis, visualised by the release of the chromogenic pNA moiety, and substrate concentration to Michaelis Menten parameters using GraphPad Prism software (Figure 6.5). The enzyme kinetic parameters for ISP are shown in Table 6.1. The order of the peptides in terms of the ISP specificity constant (k_{cat}/K_M) is shown below: FAAF>>AAPF>AAPNle>AAPNva>AAPL>AAPM>AAVA>AAPK. ISP was unable to hydrolyse the peptide substrates YVAD, AAPOrn and AAPE, which have charged hydrophilic residues at P1. The enzyme kinetic analysis using the chromogenic peptides revealed that Suc-FAAF-pNA was the preferred substrate for ISP. The k_{cat}/K_M for ISP towards the Suc-FAAF-pNA substrate was approximately sixty-times greater than that of the next favoured substrate, Suc-AAPF-pNA (121 $s^{-1}M^{-1}$ and 1.8 $s^{-1}M^{-1}$, respectively). This large difference is due to a hundred-fold increase in the K_M for the AAPF substrate. However the turnover number k_{cat} for FAAF and AAPF are relatively similar (8 s^{-1} and 12 s^{-1} respectively). Savinase also preferred Suc-FAAF-pNA (Tindbaek 2004), although there was only a four-fold difference in catalytic efficiency over Suc-AAPF-pNA. Moreover, the K_M for Suc-FAAF-pNA was only five-fold lower than Suc-AAPF-pNA for Savinase (0.233 and 1.156 mM respectively) in comparison to the one hundred-fold difference observed for ISP. However, the turnover number for Savinase with these two substrates was much higher (11-15 fold) compared to the ISP. These data suggests that although the K_M values may be similar for ISP and Savinase, the ESP shows a much higher peptide turnover and therefore catalytic efficiency than ISP.

Using the peptides AAPF, AAPNle, AAPNva, AAPL, AAPM, AAPK, AAP Orn and AAPE where P2, P3 and P4 residues were constant (AAP) the P1 position was studied. The k_{cat} values for ISP against these peptides were consistently low (0.2-0.7 s^{-1}), resulting in low k_{cat}/K_M values (0.8 - 1.4 mMs^{-1}). The exception was Suc-AAPF-

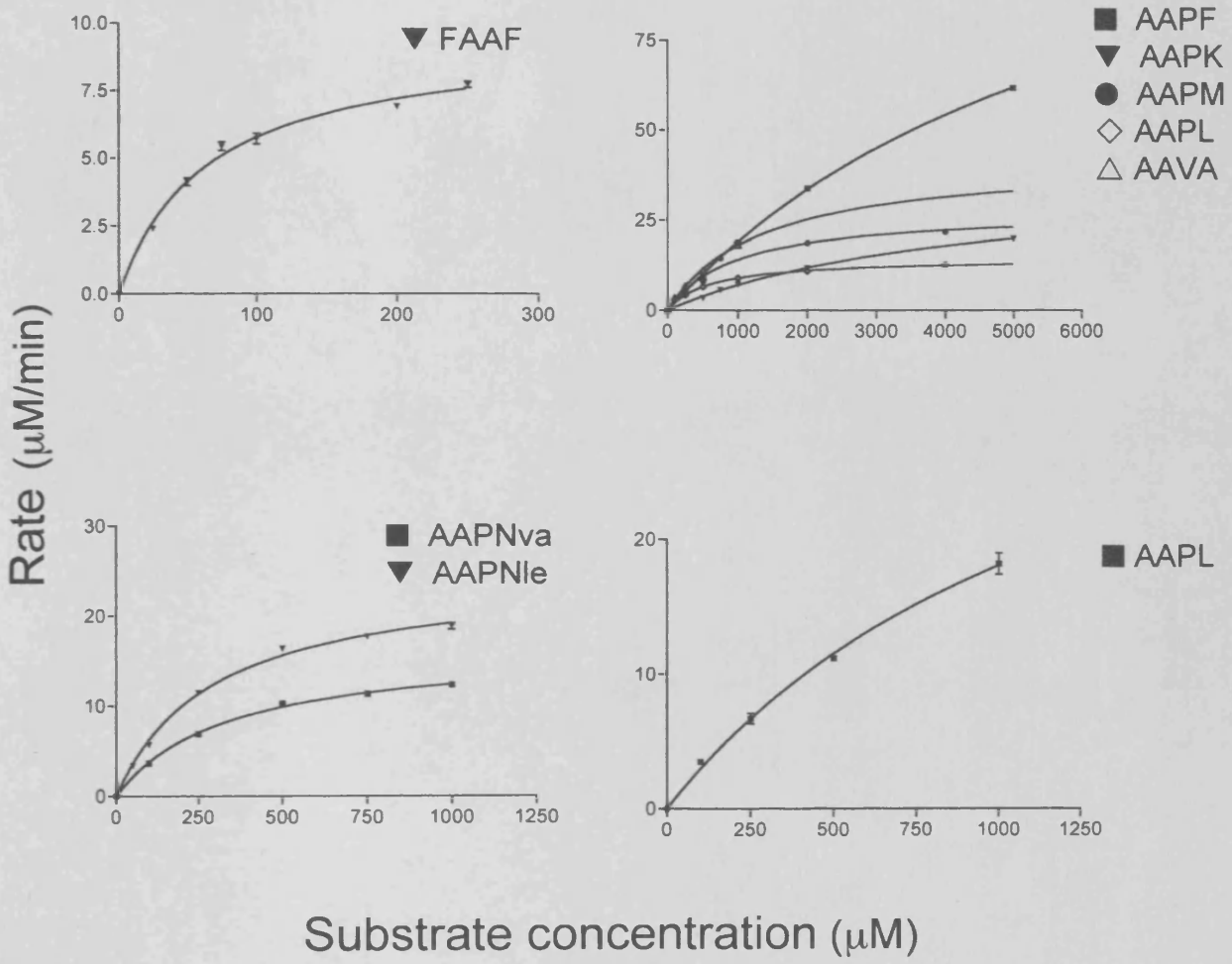


Figure 6.5. Enzyme kinetic analysis of processed ISP. The rate of ISP hydrolysis against substrate concentration were fitted to the Michaelis Menton equation using GraphPad Prism software (Version 2-http://www.graphpad.com).

Table 6.1. Enzyme kinetic analysis for ISP and ESP Savinase.

Substrate	ISP			Savinase [a]		
P^4 P^3 P^2 $P^1 \downarrow P^{1'}$	K_m (mM)	k_{cat} (s^{-1})	k_{cat} / K_m ($s^{-1}mM^{-1}$)	K_m (mM)	k_{cat} (s^{-1})	k_{cat} / K_m ($s^{-1}mM^{-1}$)
Suc-Phe-Ala-Ala-Phe-pNA	0.0661	8	121	0.233	90	418
Suc-Ala-Ala-Pro-Phe-pNA	6.475	12	1.8	1.156	185	160
Suc-Ala-Ala-Pro-Nle-pNA	0.3013	0.4	1.4			
Suc-Ala-Ala-Pro-Nva-pNA	0.344	0.3	0.8			
Suc-Ala-Ala-Pro-Leu-pNA	1.315	0.7	0.5			
Suc-Ala-Ala-Pro-Met-pNA	0.960	0.5	0.5			
Suc-Ala-Ala-Val-Ala-pNA	0.596	0.2	0.4	0.039	12	294
Suc-Ala-Ala-Pro-Lys-pNA	4.177	0.6	0.2			
Suc-Ala-Ala-Pro-Orn-pNA	ND	ND	ND			
Suc-Ala-Ala-Pro-Glu-pNA	ND	ND	ND			
Suc-Tyr-Val-Ala-Asp-pNA	ND	ND	ND	0.618	10	16

ND; Not detectable. [a] The kinetic values for Savinase were taken from Tindbaek *et al.,* (2004).

pNA peptide, one of the preferred substrates, where although a large K_M was observed the k_{cat} was higher than that shown using Suc-FAAF-pNA as the substrate. This suggests that even though the affinity of ISP for Suc-AAPF-pNA was much lower, once bound the catalytic turnover is higher. A difference in the remaining peptides is observed using the K_M values, which increases depending on the size and charge of the residue. ISP was able to hydrolyse all of the peptides containing hydrophobic, non-polar residues at P1 (norleucine, norvaline, leucine and methionine). This is in contrast to the peptides containing acidic (glutamic acid) or basic (ornithine) residues at this position, where activity was undetectable. However ISP was able to hydrolyse the peptide containing basic lysine, which is similar to ornithine, but with a high K_M and low k_{cat} suggesting that this was an inefficient and poor substrate for ISP.

ISP was also able to hydrolyse the Suc-AAVA-pNA peptide and showed a relatively low K_M, however the k_{cat} value was also very low resulting in a negligible specificity constant (k_{cat}/K_M). In comparison, the specificity constant shown by Savinase against Suc-AAVA-pNA is greater than 700 fold increase compared to that shown by ISP (294 and 0.4 $s^{-1}M^{-1}$ respectively) (Tindbaek 2004). This is due to a fifteen-fold decrease in K_M and a sixty-fold increase in k_{cat}. Interestingly ISP was unable to hydrolyse the highly polar peptide substrate, Suc-YVAD-pNA, whereas Savinase showed reasonable activity with a K_M comparable to the preferred substrates.

6.4 pH dependent activity of ISP

The protonation state of histidine residues is highly dependent on the pH of its environment. The requirement of a deprotonated His86 gives rise to the basic pH dependence of subtilisins. The majority of subtilisins have basic pH dependent activity, due to the catalytic site histidine, and are normally active in the pH range 7-10 (Van der Laan *et al.*, 1991, Betzel *et al.*, 1992). The substrate used in this assay, Suc-AAPF-pNA, was successfully used by Tindbaek *et al.*, (2004) in studying the pH dependence of Savinase. It was noted that alkaline hydrolysis of the peptide bond linking the pNA may hydrolyse above pH 11. Therefore measurements taken above this pH may overestimate activity. With this taken into account Suc-AAPF-pNA substrate was used to study ISP pH dependent activity.

The pH range for ISP (Figure 6.6) showed a sigmoidal curve with optimum activity within pH 8-10. Savinase showed a wider pH range from 7-11. This may be

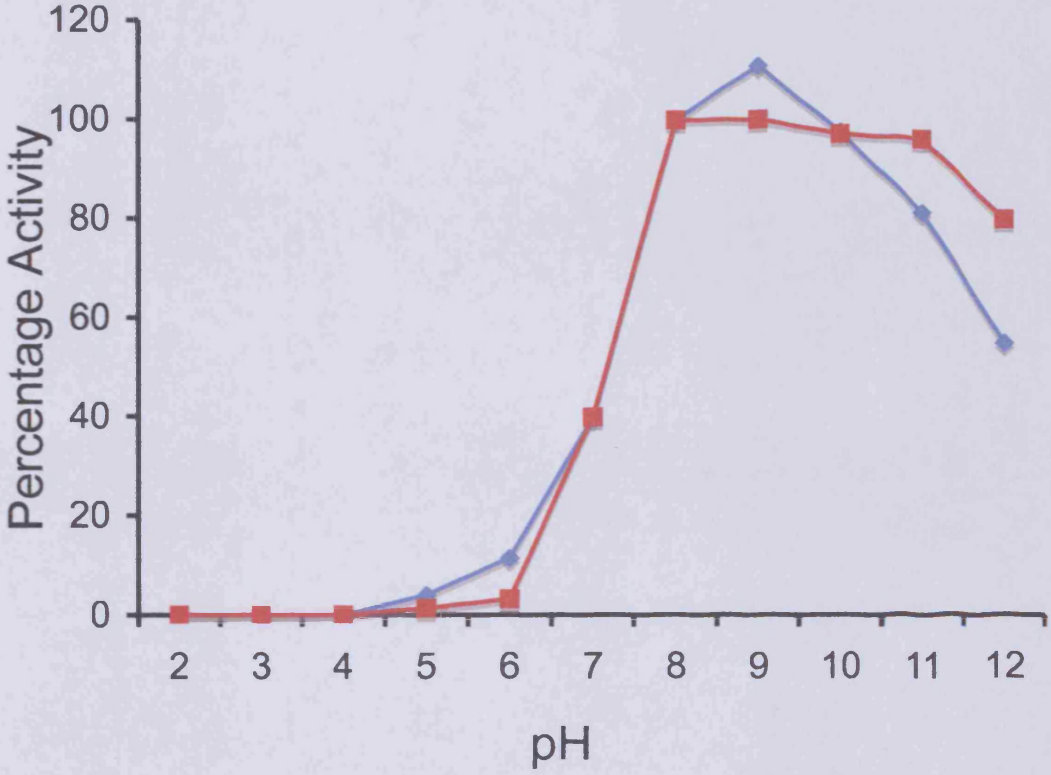


Figure 6.6. pH dependant activity of ISP and Savinase. ISP (◆) and Savinase (■) activity were determined using the chromogenic peptide Suc-AAPF-pNA at the pH values shown in the graph.

due to evolutionary pressure from the harsh extracellular environment into which many *Bacillus* species live and secrete the scavenging ESPs.

6.5 The effect of calcium and EDTA on ISP activity

The subtilisin family is highly dependent on calcium binding for stability (Bryan 2000). The harsh extracellular environment into which the majority of subtilisins are secreted requires high stability in order to withstand denaturation and proteolysis. Sequence homology suggests that the main high affinity calcium-binding site conserved in subtilases is present in the ISPs (Figure 4.1) and studies have shown that the divalent metal chelator EDTA causes loss of activity (Kurotsu *et al* 1982, Koide *et al* 1986). However, the most recent report suggested that ISP is active in the presence or absence of calcium (Sheehan and Switzer 1990).

To address this, the activity of processed ISP was measured in the presence of calcium or the metal ion-chelating agent, EDTA, using the peptide substrate Suc-FAAF-pNA. ISP was incubated in buffer supplemented with calcium (1 mM or 10 mM) or EDTA (0.01 mM, 0.1 mM or 1 mM) and rate of pNA release determined by a change in absorbance at 405 nm using a UV/spectrophotometer. Only small changes were observed in ISP activity (Figure 6.7) in the presence of excess calcium (1 or 10 mM). A similar reduction in protease activity in excess calcium has been observed with the ESPs, thought to be caused by a shift in the calcium binding and dissociation equilibrium resulting in protein unfolding (Bryan 2000). Low concentrations of EDTA (0.01 mM) had no effect on the activity of ISP. However, increasing the EDTA concentration beyond 0.1 mM dramatically decreased ISP activity to below 5% of that observed in the absence of EDTA. As metal ions are not thought to be directly involved in catalysis, the most likely explanation was that calcium removal caused a conformational change in ISP, to a partially unfolded state. This is confirmed using CD spectroscopy (Figure 6.8) in which ISP is shown to have characteristic CD spectra in 1 mM $CaCl_2$ with troughs at 208 and 222 nm. Deconvolution of this spectra indicated ISP contained 18% α-helices and 23% β-sheet, which is consistent with the spectra observed for ISP^{S250A} (Figure 4.11). A slight shift in the CD spectra is observed at 208 nm using 10 mM $CaCl_2$, which may explain the decrease in activity in the same calcium concentration. However a large change in CD spectra is observed in EDTA with a large decrease in the percentage of α-helices and corresponding increase in β-sheet. A

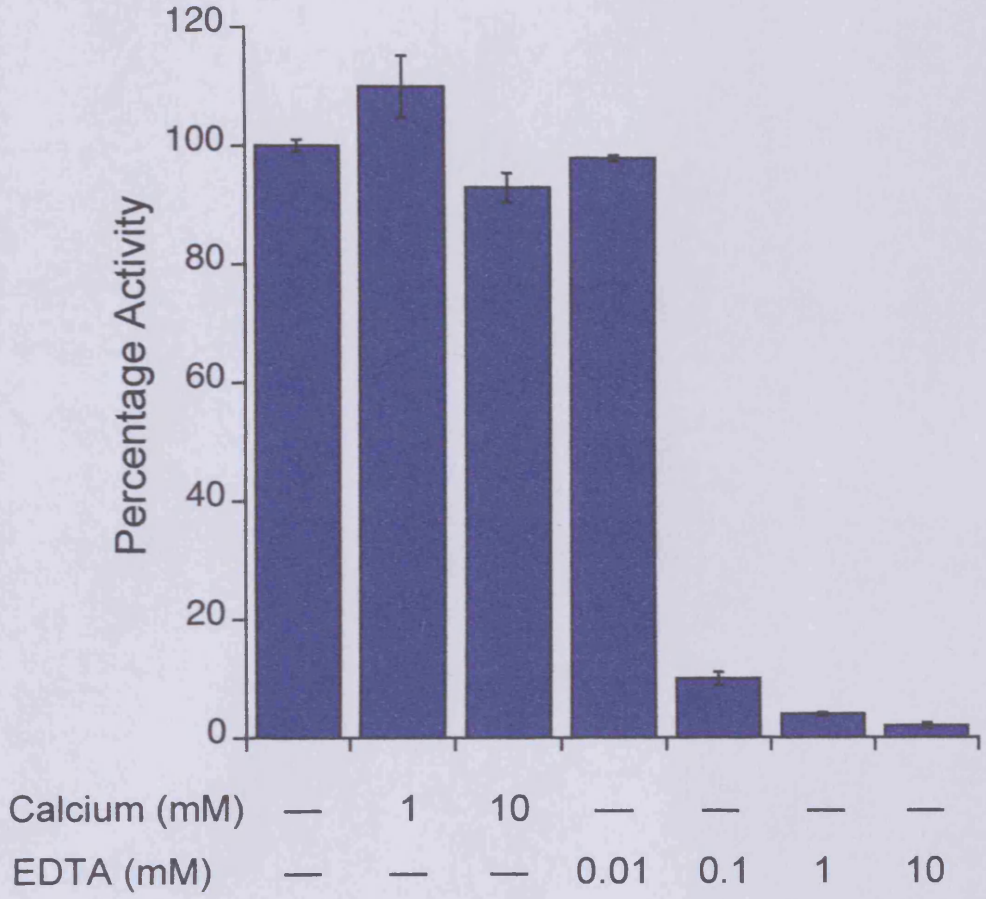
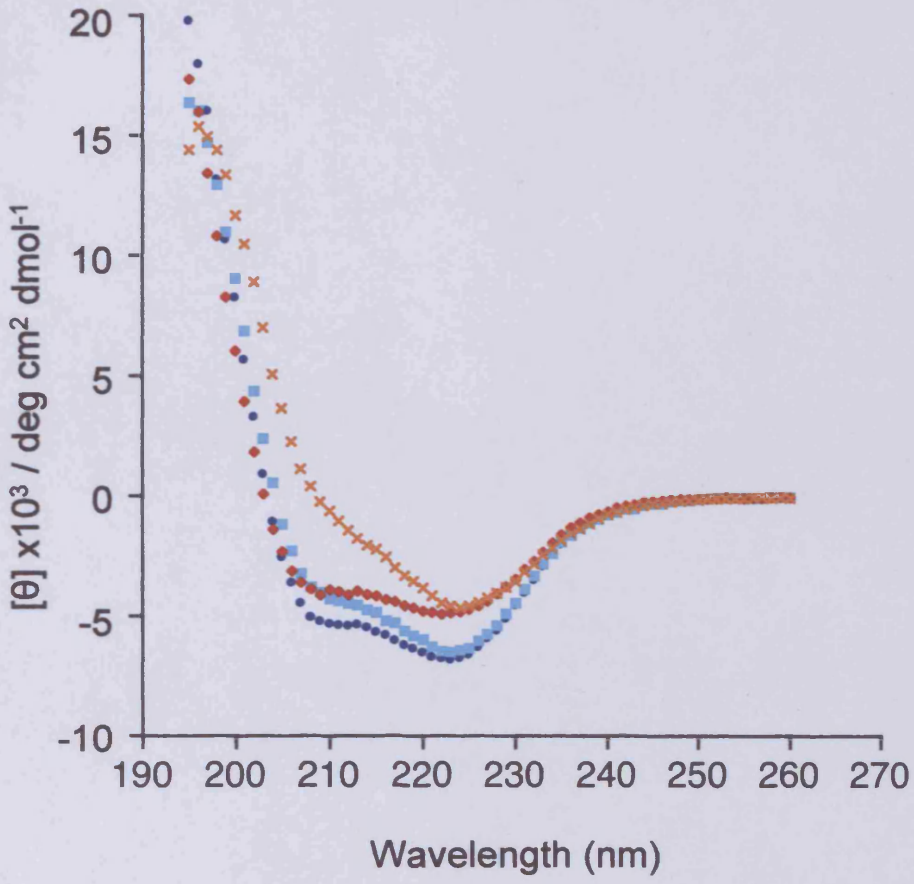


Figure 6.7. Analysis of ISP activity in the presence of specific concentrations of either calcium or EDTA. ISP was incubated for 10 minutes in buffer containing 50 mM Tris-HCl pH 8.0, 0.2 M ammonium sulphate supplemented with either calcium (1 or 10 mM) or EDTA (0.01, 0.1, 1 or 10 mM). Activity was determined by the hydrolysis of the chromogenic substrate Suc-FAAF-pNA.

A



B

		α-helix	β-strand	β-turns	others
Calcium	1 mM	18	23	14	45
	10 mM	18	25	10	47
EDTA	0.1 mM	16	23	14	47
	1 mM	5	34	12	49

Figure 6.8 Circular dichroism analysis of ISP in buffer containing calcium or EDTA. The CD spectra for ISP (**A**) (treated with 1 mM PMSF) and incubated in buffer (containing 20 mM sodium phosphate pH 8.0 and 0.1 M ammonium sulphate) supplemented with either $CaCl_2$ (1 mM ● or 10mM■) or EDTA (0.1 mM ◆ or 1 mM ×). Deconvolution of the spectra is shown (**B**) determined using CDSSTR software.

change in fluorescence spectroscopy was also observed in 10 mM $CaCl_2$, which indicated a slight change in structure surrounding one or more of the tryptophan residues and therefore a change in tertiary structure (Figure 6.9). However a larger decrease in fluorescence intensity was observed with the addition of 0.1 mM EDTA. This decrease in fluorescence intensity was also shown in 1 mM EDTA however this was accompanied by a blue shift in λ_{max}, suggesting that the change in structure had caused one or more of the tryptophan residues to become buried within structure. The activity and biophysical analysis of ISP in the presence of EDTA indicate that calcium binding is crucial for structure and therefore activity of ISP.

If ISP is undergoing a conformational change with the addition of EDTA, such as partial unfolding, it is possible that the unfolded ISP may become a target for any remaining active ISP. To assess this, ISP was incubated in buffer with and without EDTA and aliquots were taken at various time points. ISP self-degradation was analysed by SDS-PAGE (Figure 6.10 A & B). Active ISP incubated with EDTA showed considerable degradation compared to the sample in the absence of EDTA, which showed no protein degradation. A second control using the inactive ISP^{S250A} analysed under the same conditions was used to eliminate the presence of any contaminating factors that may cause the degradation. No degradation of ISP^{S250A} with or without EDTA was observed (Figure 6.10 C & D).

Therefore these data shows not only that calcium plays a key role in the stability of ISP but also that ISP preferentially degraded unfolded/partially folded ISP thus reinforcing the view that ISP is involved in protein processing of unfolded protein.

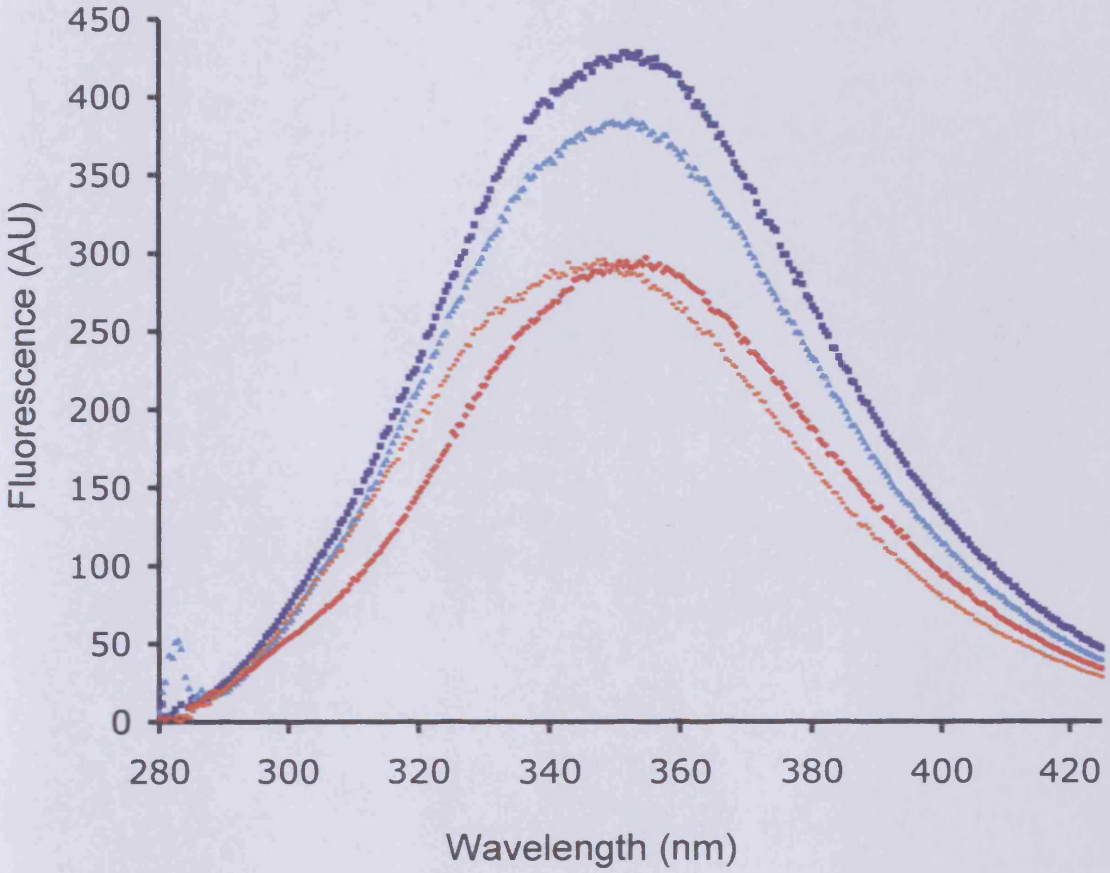


Figure 6.9 Fluorescence spectroscopic analysis of ISP in buffer containing calcium or EDTA. ISP was incubated in for one hour in buffer (containing 20 mM sodium phosphate pH 8.0 and 0.1 M ammonium sulphate) supplemented with either $CaCl_2$ (1 mM ● or 10mM■) or EDTA (0.1 mM ◆ or 1 mM ×). ISP was treated with 1 mM PMSF.

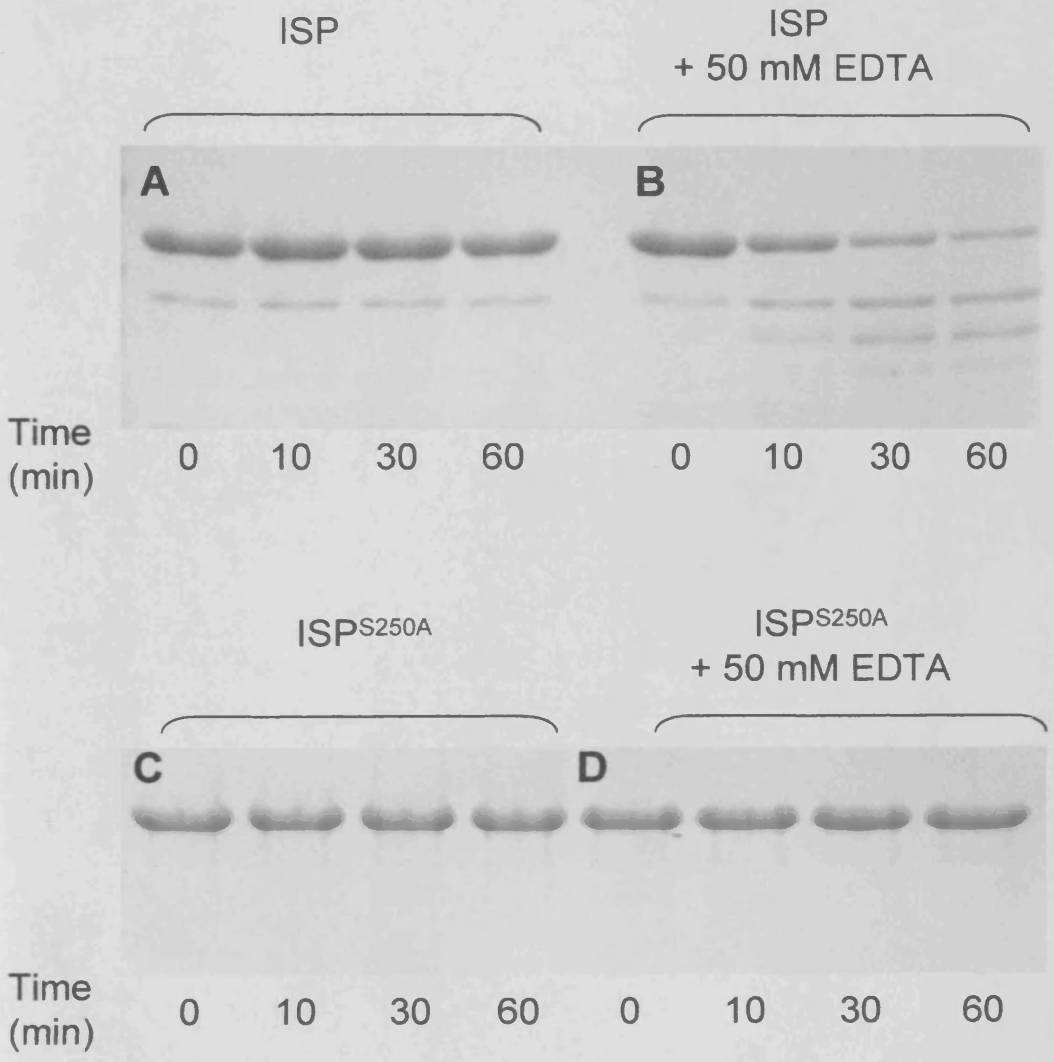


Figure 6.10 Effects of EDTA on ISP self degradation. Active (processed) ISP was incubated with **(A)** no EDTA and **(B)** 50 mM EDTA. As a control ISP^{S250A} was incubated under the same conditions (**C** no EDTA; **D** 50 mM EDTA). Fractions were taken at time intervals of 0, 10, 30 and 60 minutes and analysed using SDS-PAGE.

6.6 Discussion

The ESPs are a highly studied group of subtilases that are secreted from the Bacilli to digest extracellular proteins as both a defence mechanism against competing proteases and to allow ingestion of the amino acid for nutrition (Siezen 1997). To suit this role ESPs have low substrate specificity and high catalytic efficiency. They have also evolved a high kinetic stability and resistance to proteolysis thought to increase the lifespan of the protease. A structural alignment of the novel ISP structure and Savinase from *Bacillus lentus* reveals considerable homology (Figure 6.11 A). However the active site residues of ISP do appear to be shifted compared to those of Savinase (Figure 6.11 B). It is unclear whether this arrangement of the catalytic triad is limited to the serine to alanine mutant or is due to the presence of the N-terminal sequence (not shown in Figure 6.11) that binds back over the ISP active site and inhibits the protease. Further analysis is required which would be possible with determining the structure of a N-terminal truncated ISP.

6.6.1 Proteolytic activity of ISP

The ISP have relatively high sequence homology with the ESPs (40-50%). However, the cellular location of the ISPs, within the cell, suggests that their activity must be either highly regulated or their substrate specificity much lower to avoid non-specific proteolysis damaging the cell. ISP clearly prefers denatured forms of BSA and MDH over the folded conformations (Figure 6.1 and 6.2). This is in contrast to Savinase; a low substrate specificity scavenging protease that was capable of hydrolysing both substrates in either state. These data also suggested that the inherent activity of ISP was lower than that observed with Savinase, suggesting that there is a difference in turnover of substrates. Interestingly, ISP was unable to degrade the LDH substrate in native or unfolded conformations whereas Savinase showed degradation but only of the unfolded protein. This reinforces the inference that ISP has a higher specificity than Savinase and potentially bacilli ESPs in general.

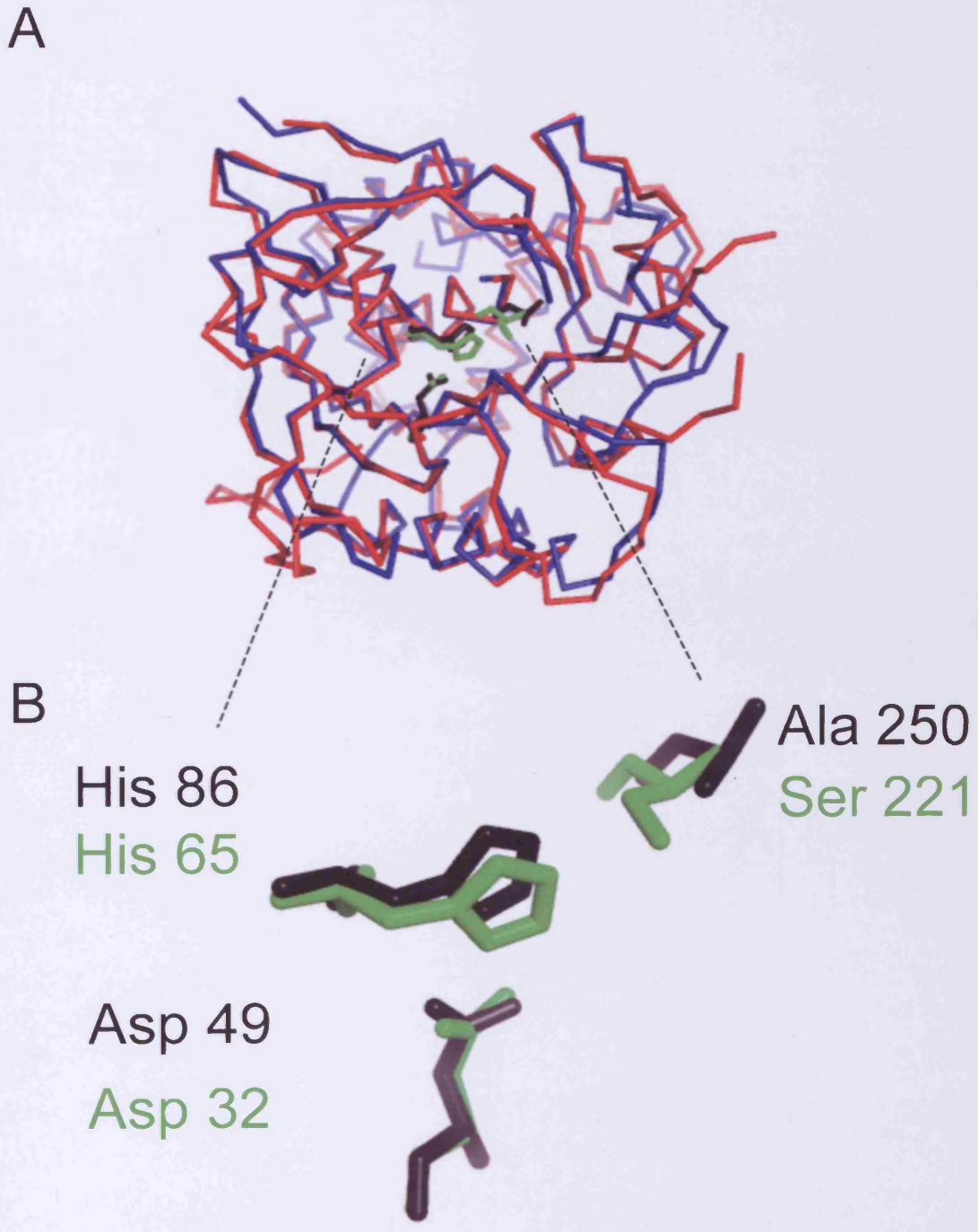


Figure 6.11. Structure alignment of ISP and BPN'. (A) ISP (red) and Savinase (blue) were aligned using MacPymol. (B) The superimposed active site residues ISP (black) and Savinase (green)

The enzyme kinetic parameters determined for ISP using synthetic peptide substrates confirm not only higher substrate specificity than that of Savinase but also a lower overall turnover of substrates. The overwhelming preference for the Suc-FAAF-pNA substrate suggested that ISP has a preference for large hydrophobic residues at P1 and P4. Although ESPs have also been shown to prefer large hydrophobic residues at P1 and P4 they are capable of binding and hydrolysing substrates containing polar hydrophilic residues (Tindbaek *et al.*, 2004) demonstrating their broader substrate specificity.

Using this set of peptides the P1 site was the most analysed with eight peptides where P2, P3 and P4 are constant (AAP) and P1 is changed. The enzyme kinetic parameters indicated that ISP preferred hydrophobic residues at position P1 rather than the basic or acid residues. This reinforces the hypothesis that ISP may be involved processing of misfolded/unfolded proteins by the recognition of residues normally hidden within the core of a protein.

The difference shown between the two preferred substrates, Suc-FAAF-pNA and Suc-AAPF-pNA, was primarily based on changes in K_M. The proline at P2 or the phenylalanine at P4 could be the cause of this large difference. However, since the P2 residue normally faces into the solvent and has little effect on substrate preference, it is more likely that the large hydrophobic phenylalanine of Suc-FAAF-pNA at P4 is better suited to binding at S4 position than the alanine. This in turn suggests that the S4 site of ISP preferentially binds large hydrophobic residues. It is not possible to directly compare the substrate specificity presented here with that shown by Lee *et al.*, (2004) who suggested that ISP have a preference for XXAA. With the data presented here (Table 6.1) it is not inconceivable that any residue may be tolerated at P3 and P4 (XX). Since the P3 residue is generally orientated away from the active site the S3 site is not normally a large factor in substrate binding. However the structure shows that the S4 site is a deep relatively hydrophobic pocket (Figure 6.15, described later) and may not tolerate any residue at this site. It is also plausible that S1 and S2 sites would prefer hydrophobic residues, as shown here ISP preferred substrate contains an phenylalanine and alanine at these sites. However, the enzyme kinetic data suggests that these sites are not constrained to alanines and may tolerate many more residues, especially larger hydrophobic residues. The structure also confirms that the S1 site is a deep hydrophobic pocket (Figure .13) and may accept the majority of hydrophobic residues.

They also suggested that ISP was involved in specific processing of protein such as ClpC and EF-tu. EF-tu is involved in aminoacyl-tRNA ribosome binding (Jonák *et al.*, 1986) and ClpC is involved in protein processing of misfolded proteins during stress (Krüger *et al.*, 2000). Therefore ISP may be involved in regulation of specific proteins such as ClpC and EF-tu via proteolysis. In addition, the specificity of ISP shown in this study suggests that it overwhelmingly prefers large hydrophobic residues at P1 and P4 and sections of proteins normally not accessible. This supports the hypothesis that ISP may be involved in the misfolded or unfolded protein response whereby ISP recognises features of such proteins (e.g. exposed hydrophobic residues) and targets them for degradation.

ISP activity was pH dependent, with highest activity in the pH range 8-10 and peak activity at pH 9. Due to the requirement of a deprotonated histidine as part of the catalytic mechanism this was expected, however this range is narrower than that of the Savinase whose activity is stable from pH 8-11. The pH within Bacillus species is 8.1 (Magill 1994) suggesting that the pH is not regulating activity of the protease within the cell. However the pH is known to change during the different growth stages of the cell including a decrease in pH throughout sporulation which may in turn decrease ISP activity (Magill 1996).

6.6.2 Structure of active ISP

The new structure of ISP^{S250A}, with the N-terminal sequence bound, allows analysis of the substrate binding pockets (Figure 6.12). This shows the S1 binding pocket accommodating Ile7 and the P1' residue Pro8 which shifts the scissile bond away from the active site residue, in the case of ISP^{S250A} alanine 250. The S2 pocket surrounds Leu6, whereas the S3 site is a small shallow groove, similarly to the ESPs (Gron 1992), where the P3 residue, Ile5, is orientated into the solvent. The S4 site is shown to form a narrow cavity housing Phe4. Analysis of the subsites of ISP^{S250A} indicates the S1 site (Figure 6.13) forms a relatively hydrophobic pocket comprised of Ala180, Ala 181, Leu154 and Phe194. This deep pocket is similar to those found in most ESPs that have been shown to prefer large hydrophobic residues (Estell 1986, Gron 1992). This supports enzyme kinetic data presented in this section where the preferred substrates both contained a phenylalanine at P1 (Table 6.1). Analysis of the

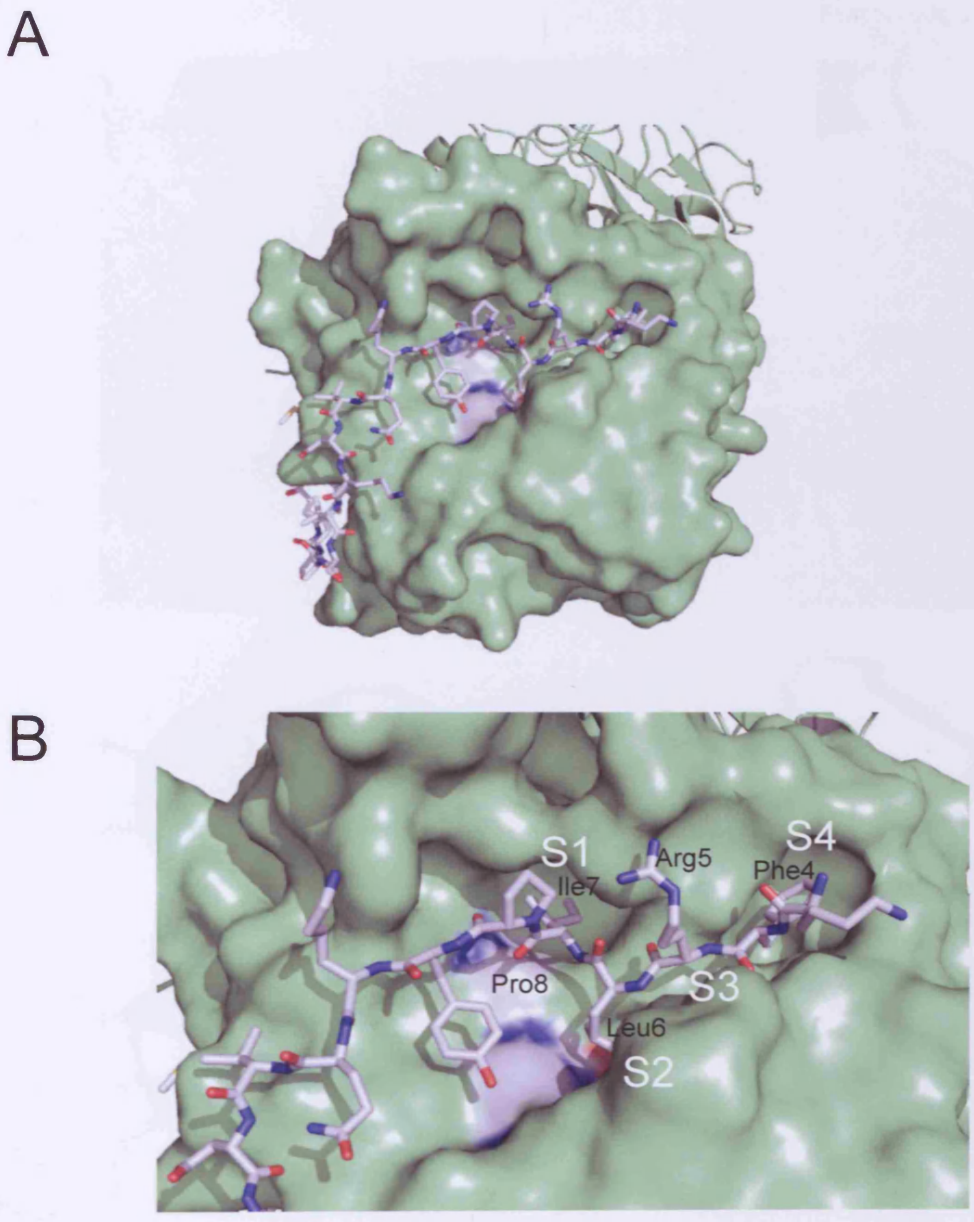


Figure 6.12 Substrate binding of ISP^{S250A}. (A) Location and binding of the ISP N-terminal extension (B) Enlarged view of substrate binding in ISP^{S250A} showing the Ile7 in the S1 binding site, Leu6 in the S2 site, Arg5 in the S3 site (although the side chain is directed into the solvent) and Phe4 in the S4 site. N-terminal extension shown as stick (coloured CPK) and active site shown as surface, (coloured CPK).

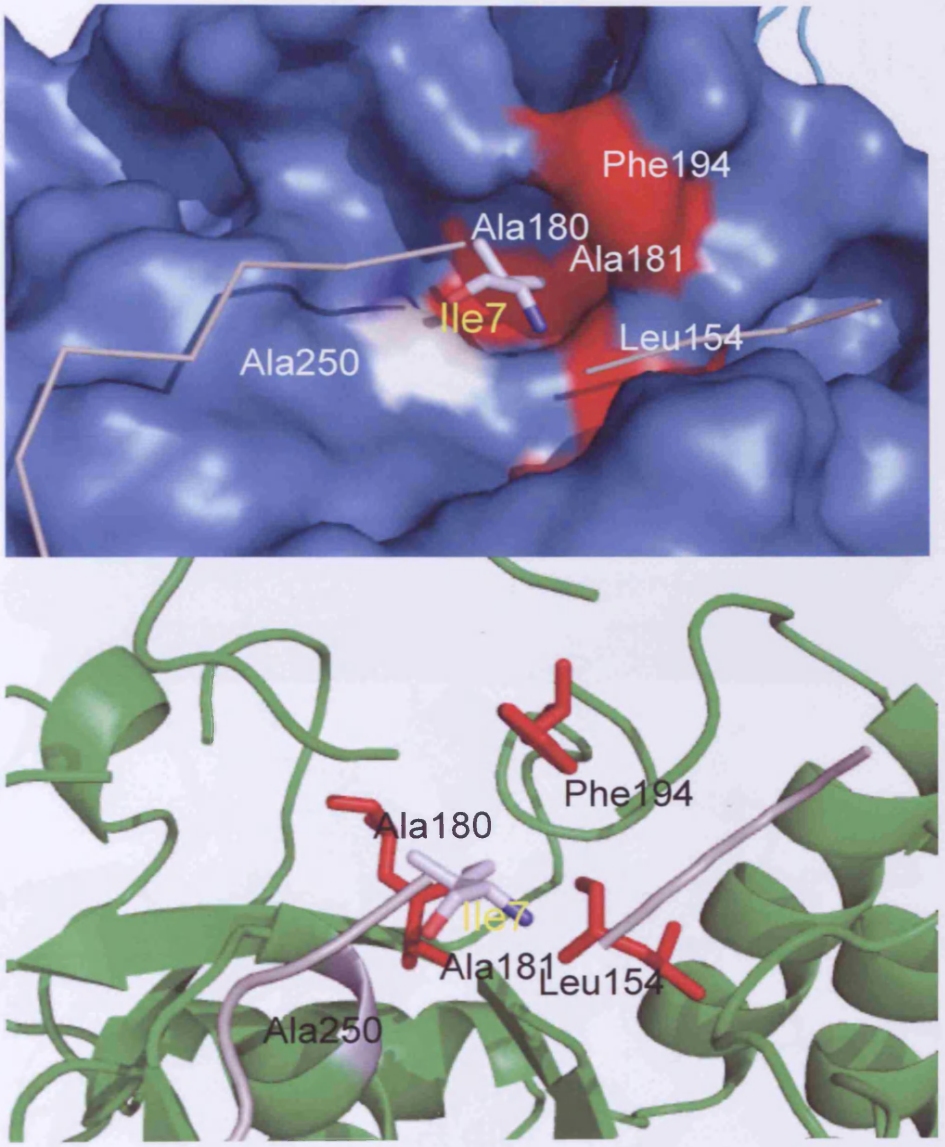


Figure 6.13 ISP^{S250A} S1 substrate binding site. The amino acids constituting the S1 substrate binding site (red) with the N-terminal Ile7 bound (CPK , labeled yellow). The remaining N-terminal of ISP overlaid (grey). Active site A250 is shown as surface (white).

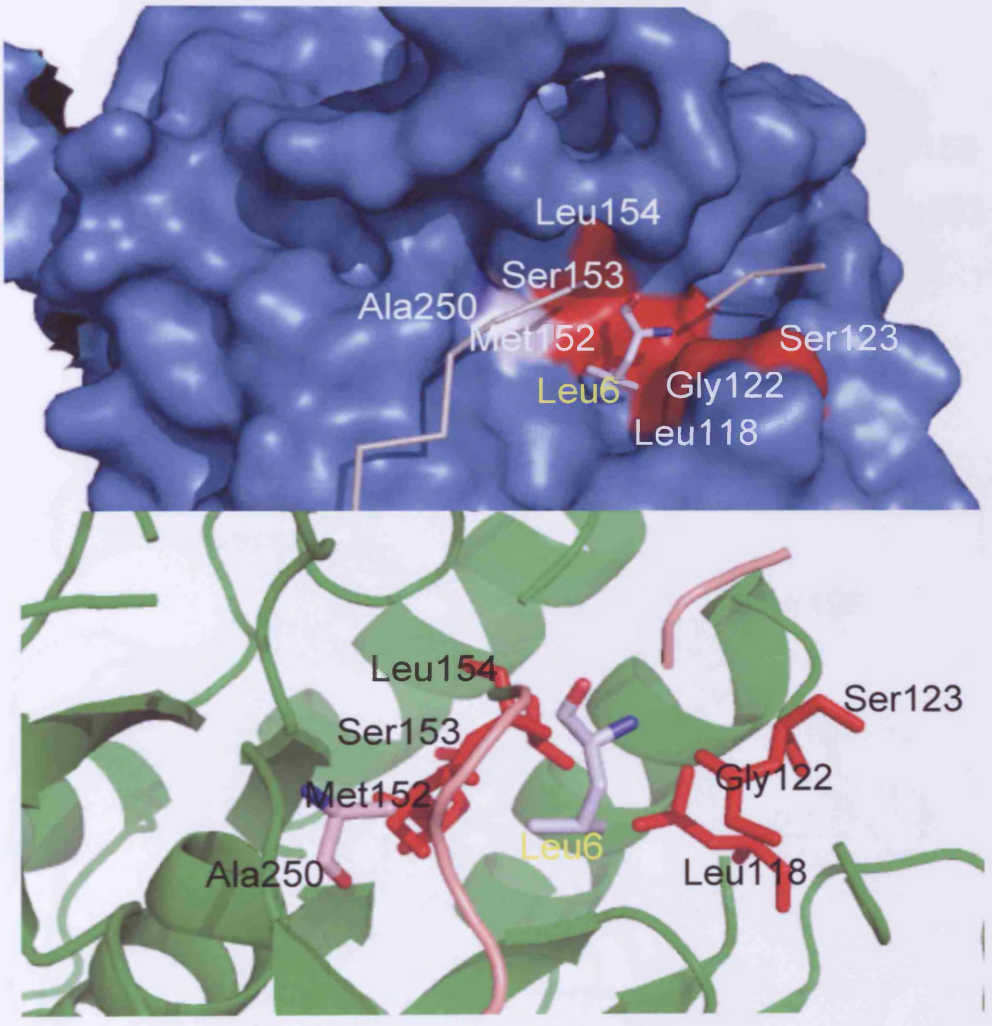


Figure 6.14 ISP^{S250A} S2 substrate binding site. The amino acids constituting the S2 substrate binding site (red) with the N-terminal Leu6 bound (CPK , labeled yellow). The remaining N-terminal of ISP overlaid (grey). Active site A250 is shown as surface (white).

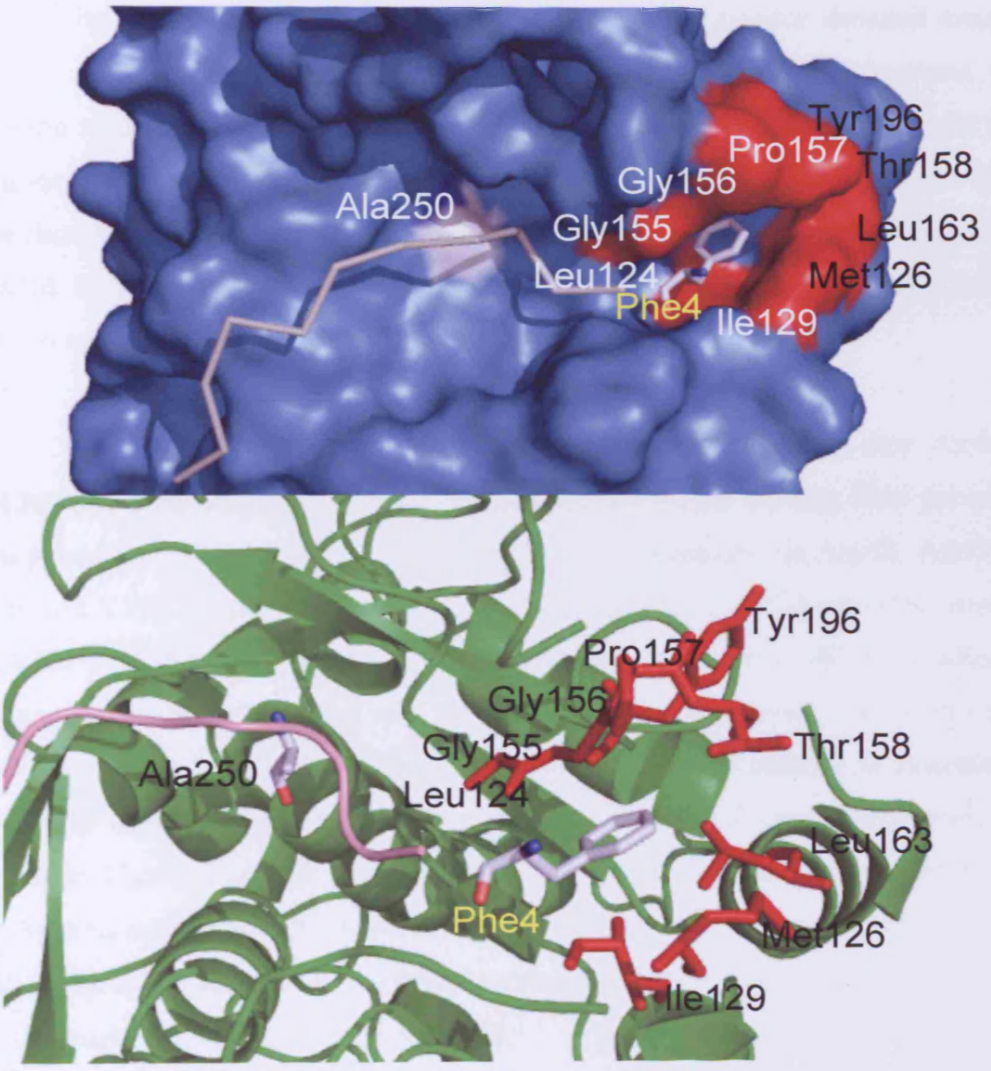


Figure 6.15 ISP^{S250A} S4 substrate binding site. The amino acids constituting the S4 substrate binding site (red) with the N-terminal Leu6 bound (CPK, labeled yellow). The remaining N-terminal of ISP overlaid (grey). Active site A250 is shown as surface (white).

structure shows that the S2 site (Figure 6.14) is lined with Leu118, Gly122, Ser123, Met152, Ser153 and Leu154, and houses Leu6 of the N-terminal extension. All of the peptides contained a hydrophobic residue at P2, however it would be interesting to determine the contribution of this site to substrate binding with further analysis using peptides with alternative residues at P2, should these become available. Interestingly, the S3 site is similar to that of the ESPs with the P3 residue directed towards the solvent. One of the most important observations from this novel structure, together with the role of the S1 site, is the orientation of the S4 site. The S1 and S4 sites are traditionally vital in substrate binding within the subtilisins (Siezen *et al.*, 1997) and are now shown to be the case with ISP. The S4 site (Figure 6.15) of ISP^{S250A} is shown to be flanked by Leu124, Met126, Ile129, Gly155, Gly156, Pro157, Ser160, Leu163 and Tyr196 creating a highly hydrophobic cavity.

The majority of subtilisins have one or more calcium binding sites. As described in Chapter 4, the structure of ISP^{S250A} revealed one metal binding sites per protomer. This is comparable to site A within the ESP and coordinates via Asp58, Ala97, Thr99, S101 and V103 residues. Here it is shown that the removal of the divalent metal, most probably calcium, by EDTA severely decreased the activity of ISP confirming its importance to the function of ISP. The decrease in ISP activity, shown in Figure 6.7 following the addition of EDTA suggests that metal ion binding is essential for an active ISP conformation. In addition, the change in the CD and fluorescence spectra, shown in Figures 6.8 and 6.9, with EDTA suggest that there is a structure change, which in turn hastened self-degradation of ISP. This also provides further evidence for ISP recognising features of unfolded proteins. It must be noted that there was a contaminating protein observed below ISP in Figure 6.8A that may be responsible for the degradation of ISP. However, no degradation of ISP is seen prior to the addition of EDTA indicating that it is more likely that the structural change due to the removal of calcium allows the remaining active ISP to degrade any unfolded ISP.

With the combination of proteolytic analysis of ISP showing a preference for unfolded/misfolded protein and the enzyme kinetic data revealing a preference for the Suc-FAAF-pNA substrate, these data support the hypothesis that ISP may be involved in processing of regions of proteins not normally accessible in native conformation. This is reinforced with data showing that removal of calcium allows the self-

degradation of ISP. It must be noted that the role of ISP may be involved in the specific processing of a protein. However the data shown here suggest that if this were the case the recognition sequence would contain hydrophobic residues, especially at P1 and P4.

Chapter 7
Folding Properties of ISP

7 Folding properties of ISP

7.1 Introduction

Anfinsen suggested in 1973 that the primary amino acid sequence contains sufficient information for a protein to fold into its native three-dimensional confirmation. However, the rules governing how changes in this primary amino acid sequence give rise to changes in protein structure, function and folding are still not fully understood. More recent theories concerning protein folding, such as the "new view', suggest there is a stochastic search in which proteins are funnelled through preferentially stable conformations, still dictated by the primary sequence, before reaching their native state (Dobson, 2004). Proteins with high sequence homology have been used to show how amino acid sequence dictates structure and function of a protein. However the folding mechanisms of these homologous proteins is normally similar. Furthermore, the majority of folding studies have been performed on small monomeric proteins. Such proteins are perceived to have simpler folding pathways and mechanisms that are easier to investigate compared to larger multimeric proteins.

The ISPs and ESPs share relatively high sequence homology, however there are distinct differences in their primary sequence that suggests there will be significant changes in their folding properties. The ESPs contain a prodomain that acts as an intramolecular chaperone or foldase that catalyses the folding process, allowing the mature subtilisin to reach its native conformation on a biologically relevant timescale (Bryan 2000, Fu *et al.,* 2000, Fisher *et al.,* 2007). Autocatalytic removal of the ESP prodomain creates a high kinetic barrier to refolding (Section 1.7). As a consequence the mature protease is unable to refold upon denaturation. The requirement of a highly specific intramolecular chaperone coupled with kinetic stability is relatively rare in proteins as the stability of most proteins is dictated by thermodynamics. Thus this has marked the ESPs out for special attention in terms of protein folding. However, it is still not known how the prodomain aids folding at the molecular level. In contrast, ISPs lack a classical prodomain. It has been shown by Subbian *et al.*, (2004) that the ISP from *Bacillus subtilis* can refold in the absence of the prodomain. However, the protease presented in this paper is monomeric and has approximately 52% sequence identity similarities with the protease from *Bacillus clausii*, studied in this project. Therefore within this chapter we will study whether the ISP from *Bacillus clausii* N-

terminal extension can act like a prodomain or if folding of ISP is independent of a prodomain. Furthermore, it needs to be established if ISPs from *Bacillus clausii* is thermodynamically stable and can be reversibly denatured, as opposed to the mature ESPs, which are kinetically stable and irreversibly unfolded.

To investigate the folding properties of ISP, protein unfolding and refolding will be performed using the denaturants urea and guanidinium chloride (GdnHCl) and monitored using protease activity, CD and fluorescence spectroscopy.

7.2 Refolding of processed ISP

The majority of subtilisins require a prodomain to facilitate folding to a mature and active 3D structure. Expression of ESPs lacking the prodomain results in the formation of a partially folded intermediate (Shinde and Inouye 2000). The ISPs lack of a prodomain is novel within the subtilisins, however as shown in Chapter 5, *Processing and activation of ISP from Bacillus clausii*, the N-terminal extension shares a role with the ESP prodomain in that they both inhibit the nascent protease, but via very different mechanisms. It therefore stands to reason that the ISP N-terminal extension may also have a role in protein folding.

Therefore to determine if processed ISP is capable of refolding to an active conformation, 8 M of the denaturant urea was used to unfold ISP. The sample was then diluted to give a final urea concentration below 0.1 M. Enzymatic activity of this sample was determined against the chromogenic substrate Suc-FAAF-pNA (Figure 7.1). For comparison the activity of processed ISP in the absence and presence of urea was also determined, representing folded and unfolded protein. The activity of ISP in native conformation prior to unfolding was deemed 100%. The unfolded ISP in the presence of 8 M urea had no activity towards the chromogenic substrate. However upon removal by dilution of the denaturant, the recovered activity of ISP was 88% of that observed for the native, folded ISP (Figure 7.1). This indicates that ISP can refold in the absence of the prodomain, in contrast to that observed for the ESPs.

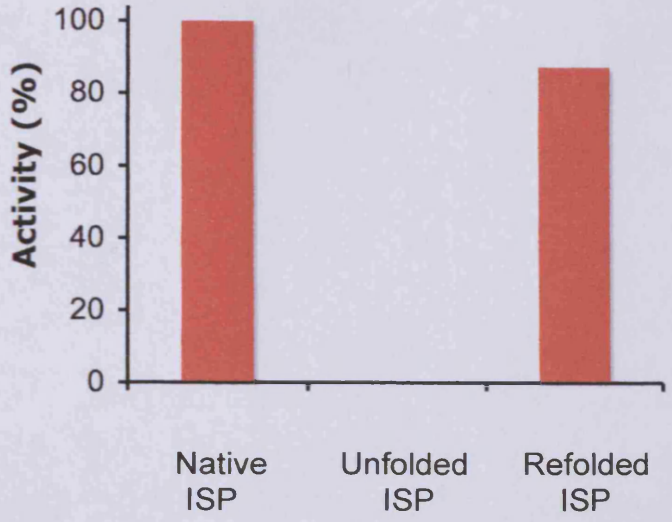


Figure 7.1 Refolding of processed ISP. The activity of processed ISP was determined in native (without denaturant), unfolded (addition of 8 M urea) and refolded (following dilution to reduce the urea concentration to below 0.1 M) conformations. Enzymatic activity was determined using the chromogenic substrate Suc-FAAF-pNA

7.3 Biophysical analysis of ISP^{S250A} refolding

As shown in the previous section ISP was capable of refolding without the N-terminal extension. To complement this study, the biophysical techniques CD, to probe changes in secondary structure and fluorescence spectroscopy, as an indicator of changes in the solvent exposure around tryptophan residues were used to study refolding of ISP independent of activity. However due to the hypothesised role for the ISPs in the processing of unfolded/misfolded proteins, supported with data in Chapter 6, *Characterisation of the enzymatic activity of ISP*, the deactivated ISP^{S250A} will be used to avoid proteolysis complicating analysis.

CD spectra revealed that ISP^{S250A} is essentially denatured in 6 M GdnHCl and 8 M urea (Figure 7.2 A&B) with a decrease in ellipticity, especially at 222 nm normally associated with α-helical structure. Upon removal of GdnHCl, the CD spectra was similar to that observed for the native, folded protein with the return of CD signal at 222 nm and 208 nm suggesting fully folded protein. The CD analysis of ISP^{S250A} refolding after urea denaturation (Figure 7.2B) also suggests that ISP^{S250A} refolds, with

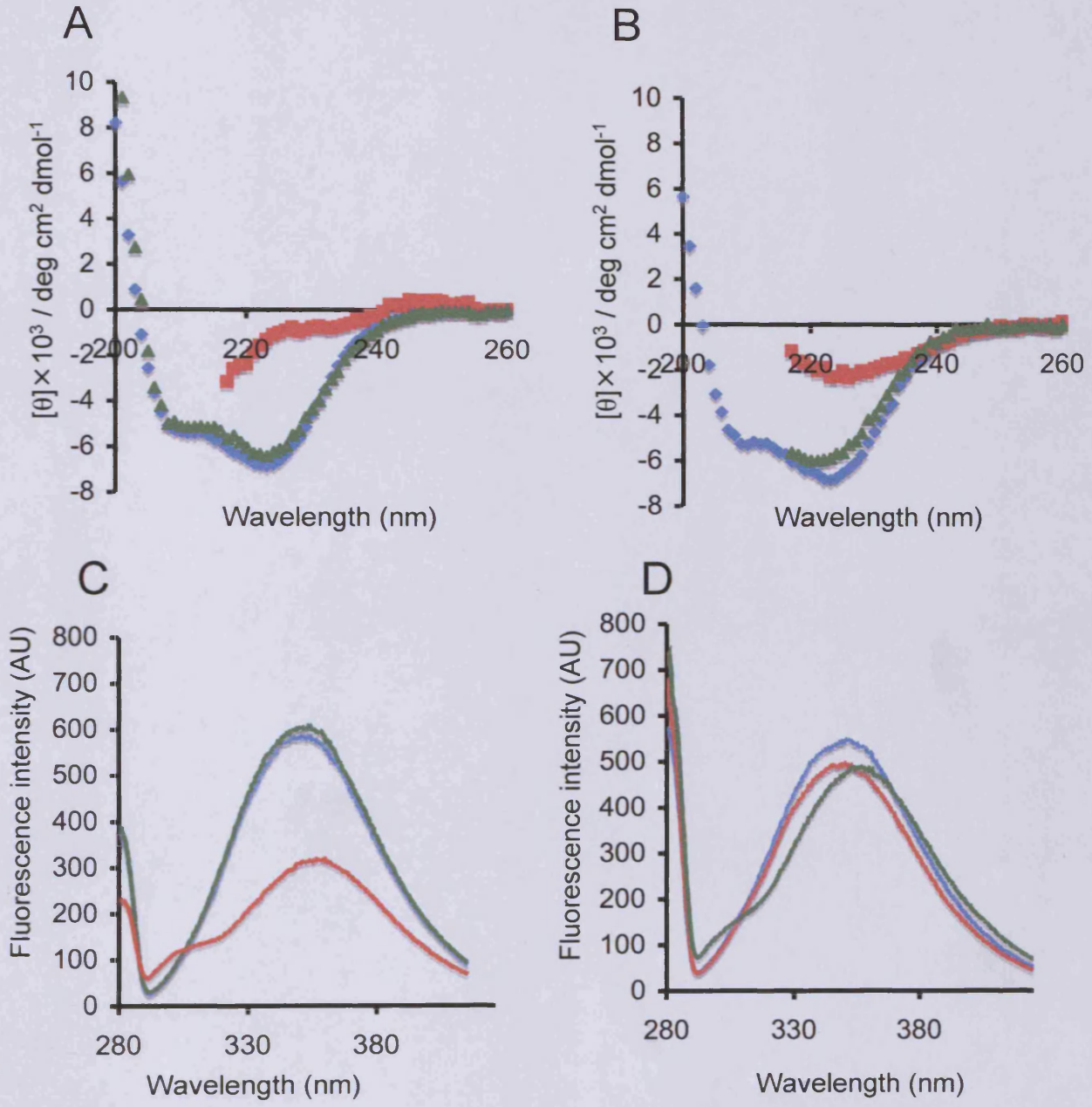


Figure 7.2 Unfolding and refolding of ISP^{S250A}. Unfolding and refolding analysed by CD spectroscopy with ISP^{S250A} denatured in 6 M GdnHCl (A) or 8 M urea (B). Unfolding and refolding of ISP^{S250A} analysed by fluorescence in 6 M GdnHCl (C) or 8 M urea (D). In each panel the folded, native samples are coloured blue, the chemically denatured samples are coloured red and the refolded samples after dilution of the denaturant are coloured green.

the generation of a similar folded spectrum with a similar trough at 222 nm. However the CD signal at 222 nm only returns to approximately 80% coinciding with that seen with the activity refolding in Section 7.1.

Fluorescence spectra confirmed denaturation of ISP^{S250A} in the presence of 6 M GdnHCl (Figure 7.2 C) with a drop in fluorescence intensity and a slight red shift in λ_{max} for emission. The drop in fluorescence and red shift is due to the quenching effect of the solvent on exposure of the tryptophan since the energy of the emitted photon is reduced. On removal of denaturant the fluorescence intensity and λ_{max} returned to that observed with native, folded protein indicating full refolding of ISP^{S250A}. The fluorescence spectra revealed that unfolding in urea (Figure 7.2D) did not reduce the fluorescence intensity significantly but does cause a red shift in the fluorescence maxima, which was restored upon removal of urea.

These results confirms that ISP can refold in the absence of the classical prodomain found in other subtilisins.

7.4 Thermal denaturation of ISP^{S250A}

Thermal unfolding of ISP^{S250A} was analysed using CD and fluorescence. The major changes to the CD spectra were observed at 222 nm. With fluorescence spectroscopy the major change was seen at 330 nm. Therefore these wavelengths were used to monitor the folded/unfolded transition caused by increasing the temperature of the system. As shown in Figure 7.3, the CD ellipticity becomes less negative with increased temperature, suggesting a reduction in secondary structure. The curve suggested that there were two thermal transitions midpoints (T_m) of 56 and 69°C. Fluorescence spectra revealed a decrease in intensity (Figure 7.3) as temperature increased, with a T_m calculated to be 55°C.

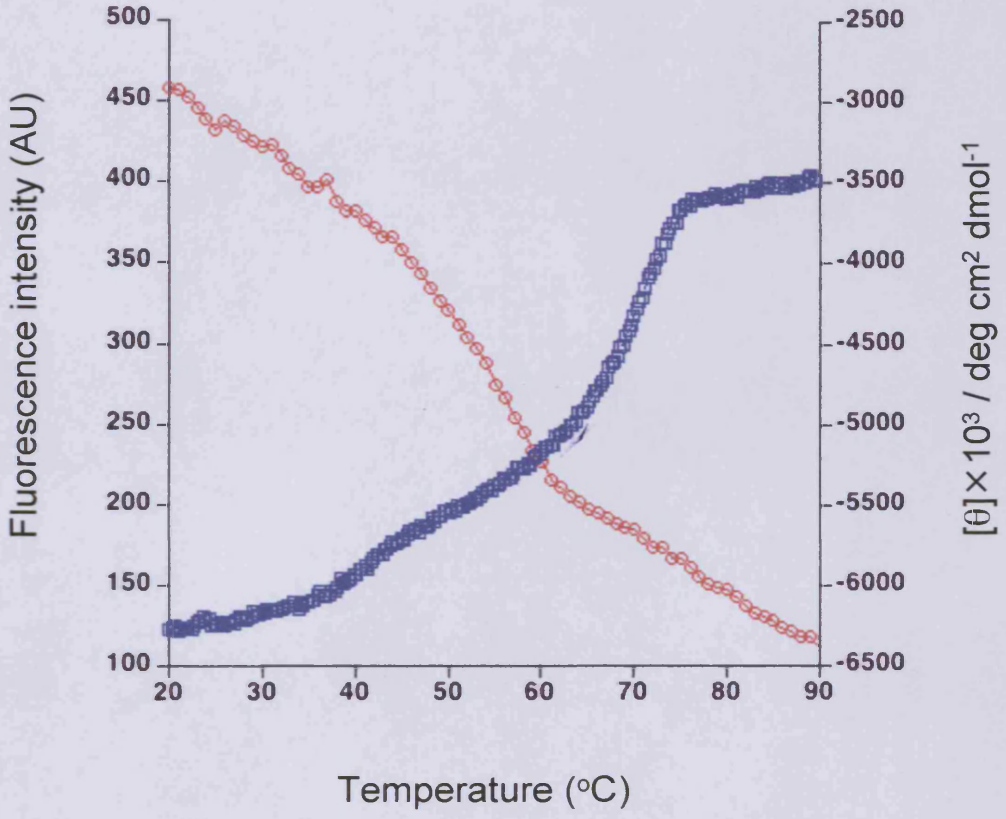


Figure 7.3 Thermal denaturation of ISP^{S250A} as analysed by CD (blue) and fluorescence (red) spectroscopy. Using CD the thermal denaturation profile of ISP^{S250A} monitored at 222 nm. With fluorescence the sample was excited at 280 nm and emission monitored at 330 nm. The T_m of ISP^{S250A} using fluorescence was 55 °C. The T_m of ISP^{S250A} using CD was 56 and 69 °C.

7.5 Chemically induced denaturation of ISP^{S250A}

Urea and guanidinium chloride (GdnHCl) are, at present, the most frequently used protein denaturants. It is unclear how these chemical cause protein denaturation but it is known to be involved in the disruption of the non-covalent interactions associated with protein structure. The equilibrium unfolding and refolding of ISP^{S250A} will be analysed using urea and GdnHCl and variation in structure monitored by changes in CD at 222 nm and fluorescence emission at 330 nm after excitation at 280 nm. This will therefore allow an estimation of the conformation stability in terms of free energy of folding.

As shown in Figure 7.4, the unfolding and refolding curves for ISP^{S250A} using GdnHCl and monitored using CD at 222 nm are very similar. Once again it has been demonstrated that the ISP from *Bacillus clausii* can refold once denatured. As it has been established that ISP^{S250A} is a dimer, analysis of the curves and extrapolating thermodynamic parameters has proved difficult. The curves, where possible, were fitted to an equation for a two-state dimer to monomer $N_2 \leftrightarrow 2U$ (Section 2.3.14). The curve fitting provides the parameter free energy change (ΔG^{H_2O}) that is a direct measure of the stability of a protein in the absence of denaturant. Also calculated is the denaturant susceptibility parameter *m*, which is associated with the gradient of the slope and gives an indication of the change in stability with denaturant. This value also strongly correlates with the solvent accessibility of apolar residues upon unfolding. A large *m* value normally indicates a large or steep change in stability and solvent exposure of apolar residues i.e. a large buried hydrophobic core of the protein that is rapidly exposed. In contrast a small *m* value suggests a lesser change in stability and therefore reduced exposed buried apolar surface area. Using this model the ΔG^{H_2O} for unfolded ($\Delta G_{N\text{-}U}{}^{H_2O}$) and refolded ($\Delta G_{U\text{-}N}{}^{H_2O}$) ISP^{S250A} using GdnHCl (Figure 7.4) were calculated to be 10.7 and -8.5 $kcalmol^{-1}$ (Table 7.1), respectively. These values suggest that ISP^{S250A} is relatively stable. For example, ubiquitin and CI-2 have a $\Delta G_{N\text{-}U}{}^{H_2O}$ of 7.1 and 7.0 $kcalmol^{-1}$, respectively (Jackson 1998). The *m* values are also similar (2.2 and 2.0 $kcalmol^{-1}M^{-1}$, respectively). The *m* values suggest that there is a substantial change in stability and solvent exposure on unfolding. The midpoints for the main transitions, C_m, were very similar (1.1 and 1.0 M GdnHCl respectively). These low

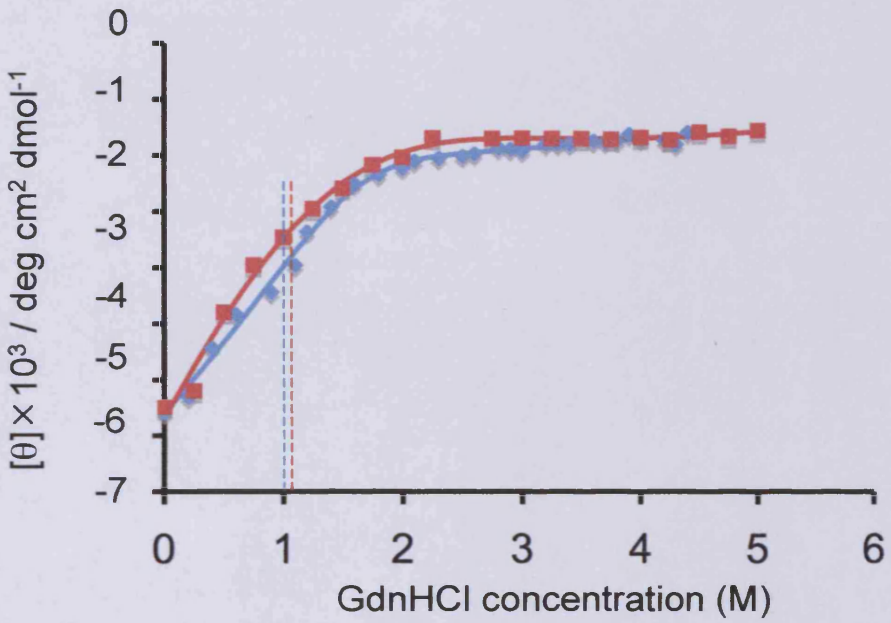


Figure 7.4 Unfolding and refolding of ISP^{S250A} with GdnHCl as analysed by CD spectroscopy. The CD spectroscopy profiles for the unfolding (blue) and refolding (red), measured by CD at 222 nm. The curves were fitted to a two state $N_2 \Leftrightarrow 2U$ curve using Kaleidagraph 2 software. The C_m values showing the transition points are shown as dotted line calculated using a first derivative analysed with Origin8 software, OriginLab.

Table 7.1 Comparison of different thermodynamic parameters for unfolding and refolding of ISP^{S250A}.

		GdnHCl	+/-	Urea	+/-
CD	$\Delta G_U^{H_2O}$ - unfolding	10.7 kcalmol^{-1}	1.5	10.0 kcalmol^{-1}	4.5
	m - unfolding	2.2 kcalmol^{-1}M^{-1}	0.5	0.36 kcalmol^{-1}M^{-1}	7.0
	$\Delta G_U^{H_2O}$ - refolding	-8.5 kcalmol^{-1}	0.8	-13.3 kcalmol^{-1}	10.4
	m - refolding	2.0 kcalmol^{-1}M^{-1}	0.5	1.2 kcalmol^{-1}M^{-1}	0.6
Fluorescence	$\Delta G_U^{H_2O}$ - unfolding	NA		12.9 kcalmol^{-1}	0.9
	m - unfolding	NA		1.3 kcalmol^{-1}M^{-1}	0.2
	$\Delta G_u^{H_2O}$ - refolding	-10.3 kcalmol^{-1}	0.6	-12.9 kcalmol^{-1}	2.3
	m - refolding	4.2 kcalmol^{-1}M^{-1}	0.8	1.7 kcalmol^{-1}M^{-1}	0.6

Curves were fitted to two-state $N_2 \leftrightarrow 2U$ equation (Walters *et al.*, 2009). (NA - Non applicable due to potential intermediate transitions.

concentrations of GdnHCl suggest that ISP^{S250A} is highly susceptible to the denaturant GdnHCl.

The unfolding and refolding of curves ISP^{S250A} with GdnHCl monitored using fluorescence were dissimilar (Figure 7.5). There was a pronounced suggestion of a secondary transition, especially during the unfolding of ISP^{S250A}. Two C_m values were determined 0.6 and 2 M. Due to time restraints and the complicated nature of the fitting algorithms the curves were not fitted to the three-state equations ($N_2 \leftrightarrow 2I \leftrightarrow 2U$ or $N_2 \leftrightarrow I_2 \leftrightarrow 2U$) that included either a monomeric (2I) or dimeric (I_2) intermediate, as was planned. The refolding of ISP^{S250A} from GdnHCl monitored by fluorescence was fitted to a two state equation and revealed a $\Delta G_u^{H_2O}$ of -10.3 kcalmol^{-1} (Table 7.1), which is comparable to that shown using CD spectroscopy.

The unfolding of ISP^{S250A} using urea (Figure 7.6), monitored using CD, also suggested that three state unfolding may occur. However, due to the lack of data points around the first transition it was not possible to determine if the transition is real. Fitting the data to a three-state model would not allow the parameters to be resolved. Consequently, the data fitted better to a two-state equation. The $\Delta G_u^{H_2O}$ was calculated to be 10.0 kcalmol^{-1}, which is comparable to the values shown for GdnHCl. The refolding of ISP^{S250A} using urea showed a transition at approximately 2 M similar to unfolding however the transition at 0.6 M was absent. Furthermore, the refolded

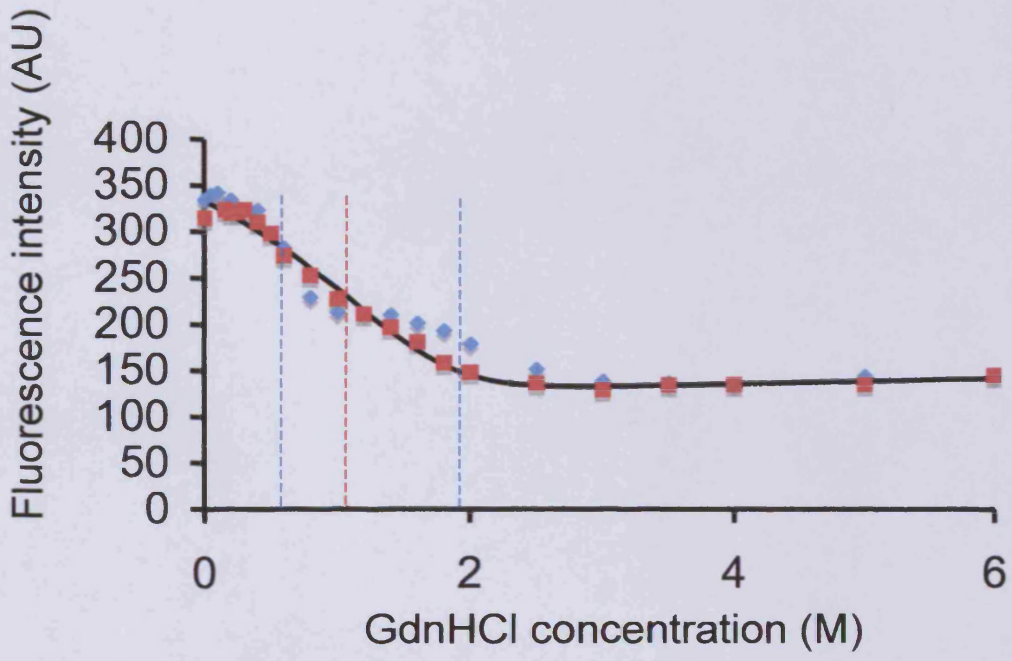


Figure 7.5 Unfolding and refolding of ISP^{S250A} with GdnHCl as analysed by Fluorescence spectroscopy. Fluorescence intensity was measured at 330 nm with excitation at 280 nm. The refolding curve (red) was fitted to a two state $N_2 \Leftrightarrow 2U$ curve (black) using Kaleidagraph 2 software. The unfolding curve (blue) suggests there may be an intermediate in the unfolding of ISP^{S250A}. The C_m values showing the transition points are shown as dotted lines and were calculated using a first derivative analysed with Origin8 software, OriginLab.

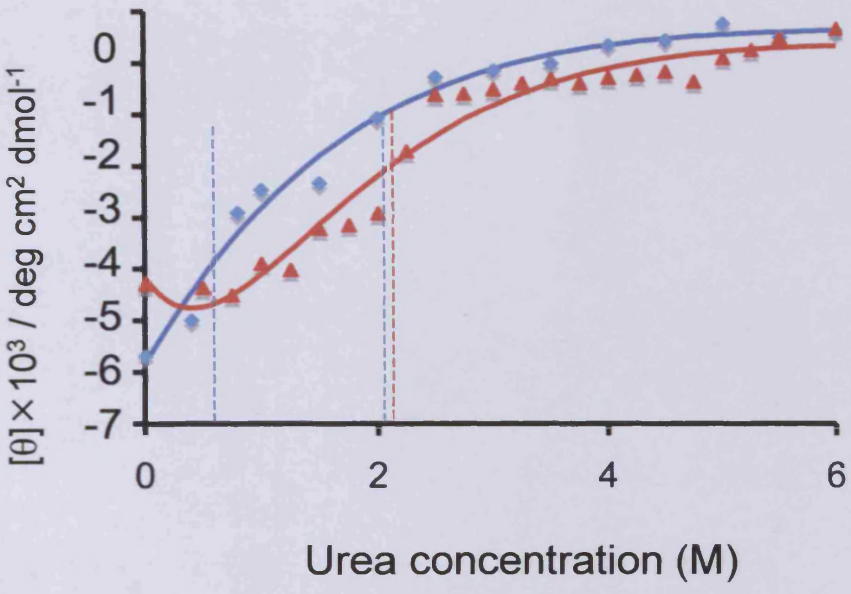


Figure 7.6 Unfolding and refolding of ISP^{S250A} with urea as analysed by CD spectroscopy. The CD spectroscopy profiles for the unfolding (blue) and refolding (red), measured by CD at 222 nm. The curves were fitted to a two state $N_2 \Leftrightarrow 2U$ curve using Kaleidagraph software. The C_m values showing the transition points are shown as dotted lines and were calculated using a first derivative analysed with Origin8 software, OriginLab.

ellipticity did not reach the same signal level observed for the folded ISP^{S250A} prior to unfolding. This suggests that there was either an experimental problem with refolding such as the length of sample incubation or that the refolding of ISP^{S250A} follows a different refolding path. It may also be the case that ISP^{S250A} forms a partially unfolded form that could aggregate. The curve was fitted to a two-state equation with a $\Delta G_u^{H_2O}$ calculated to be -13.3 $kcalmol^{-1}$ however the error value was high suggesting this value may be overestimated, compared to data using GdnHCl.

When fluorescence was used to monitor ISP^{S250A} unfolding and refolding with urea (Figure 7.7) there was no indication of an intermediate transition. Therefore the two-state model was used. The $\Delta G_u^{H_2O}$ and m values for unfolding and refolding of ISP^{S250A} were similar (12.9 $kcalmol^{-1}$, -12.9 $kcalmol^{-1}$ and 1.3 $kcalmol^{-1}M^{-1}$ and 1.7 $kcalmol^{-1}$ M^{-1}, respectively). Similarly to the data using CD to monitor unfolding and refolding (Figure 7.6), refolded ISP^{S250A} did not revert to the same fluorescence magnitude observed in 0 M urea. This may also suggest that ISP^{S250A} refolds to a slightly different conformation with the removal of urea.

7.6 Unfolding of ISP^{S250A} monitored by SEC

The dimerisation of ISP has been established by SEC and confirmed with the recent crystal structure. As shown in the previous section, it is unclear whether ISP folding/unfolding pathways include an intermediate state. Therefore SEC was used to monitor changes in hydrodynamic volume of ISP^{S250A} in various concentrations of urea. As shown in Figure 7.8, with the addition of 2 M urea ISP^{S250A} elutes earlier (14.4 ml) compared to that in the absence of urea (14.8 ml). That suggests that ISP^{S250A} has a higher hydrodynamic volume in 2 M urea. In addition, the profile peak is comparable to ISP in native conformation suggesting the presence of only one ISP^{S250A} form. In 4 M urea, the peak at 14.4 ml was still observed. However, the major peak elution was broad with peak elution at 11.9 ml suggesting a large change in hydrodynamic volume. A further decrease in elution volume is shown in 8 M urea indicating that unfolding was not complete in 4 M urea.

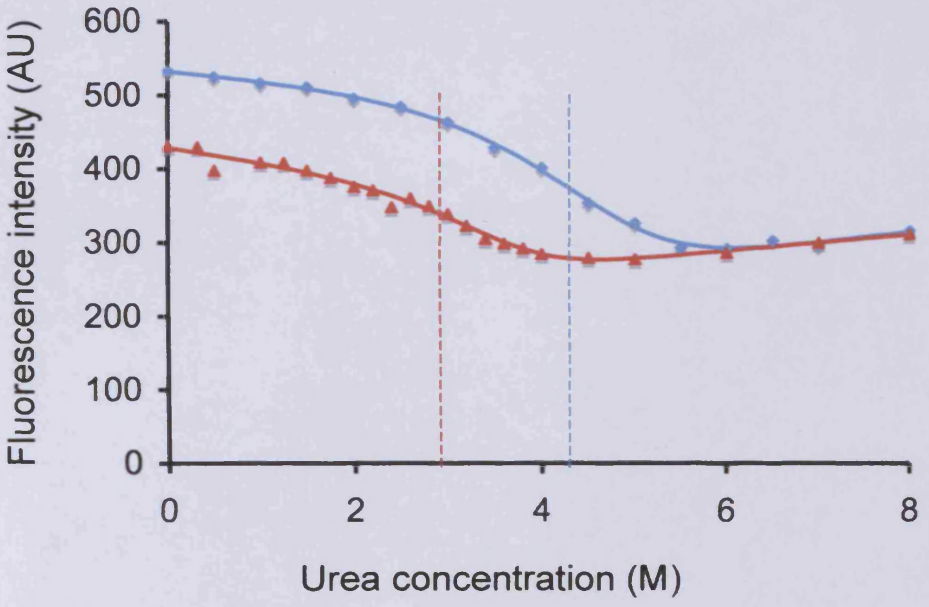


Figure 7.7 Unfolding and refolding of ISP^{S250A} with urea as analysed by fluorescence spectroscopy. The fluorescence spectroscopy profiles for the unfolding (blue) and refolding (red) of of ISP^{S250A}, measured at 330 nm with excitation at 280 nm. The curves were fitted to a two state $N_2 \Leftrightarrow 2U$ curve using Kaleidagraph software. The C_m values showing the transition points are shown as dotted lines and were calculated using a first derivative analysed with Origin8 software, OriginLab.

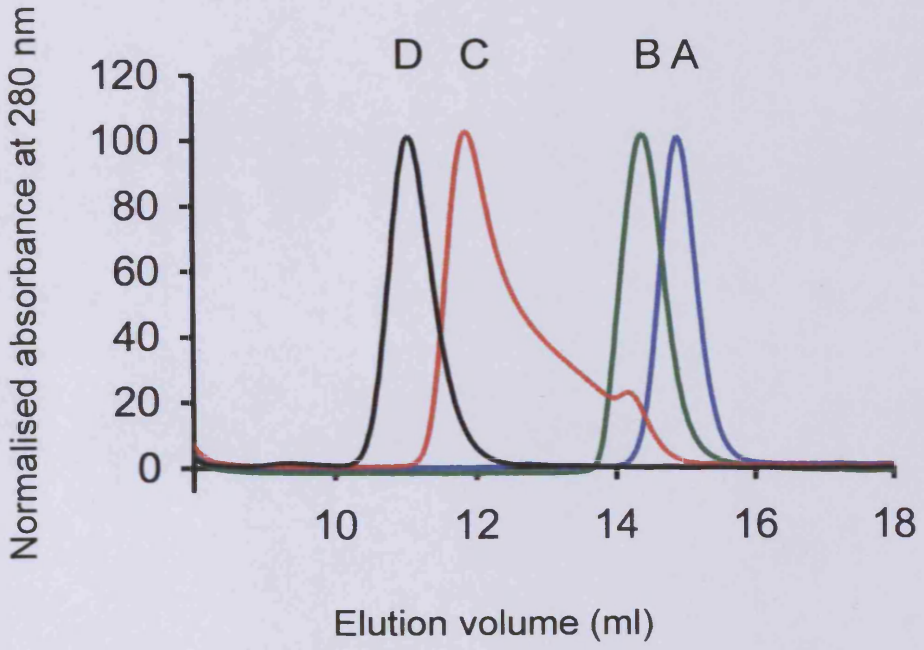


Figure 7.8. Unfolding of ISP^{S250A} using urea as analysed by size exclusion chromatography. Elution profiles for 50 mM ISP^{S250A} in (A) 0 M urea, (B) 2 M urea, (C) 4 M urea and (D) 8 M urea. Conditions: 50 mM sodium phosphate (pH 7.5), 0.5 M ammonium sulphate, 1 mM calcium.

7.7 Discussion

Understanding how the linear sequence of amino acids folds into a defined 3D structure remains to this day one of the most important questions in biology. It is generally assumed that proteins with similar sequences and structures fold by equivalent pathways and mechanism. However, this does not appear to be the case for the ISPs and ESPs. While the core structure of the ESPs and ISPs studied here are very similar (Figure 4.6, Section 4.4) the ISP appears to be thermodynamically stable and does not require the classical prodomain for folding, unlike the ESPs that require a prodomain for folding (Eder *et al.*, 1993). In the absence of the prodomain the ESPs are trapped as a partially folded intermediate. In contrast, in the presence of the prodomain the ESPs fold to an active conformation that is kinetically stable upon cleavage of the prodomain. A large free energy barrier to unfolding that is created is an advantage to the ESPs, which function in the harsh extracellular environment, by enhancing longevity via a decreased susceptibility to proteolysis and denaturation. The high-energy barrier decouples the folding and unfolding pathways resulting in a large activation energy for folding in the absence of the prodomain (Shinde *et al.*, 1995, Fu *et al.*, 2000). Both the full length and processed forms of ISPs on the other hand, display thermodynamic stability commonly observed for most other proteins that can be reversibly unfolded. Refolding of full length ISP^{S250A} in GdnHCl revealed almost complete restoration of CD and fluorescence signal (Figure 7.2), confirming that ISP^{S250A} can be reversibly unfolded. However, when using urea as the denaturant, the CD signal returned to approximately 80%. This matches the percentage of activity restored following refolding of processed ISP from urea (Figure 7.2). It was thought that this decrease in processed ISP activity was due to proteolysis, which may be true, however, since the ISP^{S250A} is inactive, it may be the case that not all of the ISP^{S250A} population fully refolds from urea, with a proportion remaining unfolded or forming aggregates. This is not uncommon when using urea, and is normally due to the sample not reaching equilibrium. However, the samples were left overnight before being analysed and it is unlikely that longer incubation may allow full refolding of ISP^{S250A}. Although the action of the two denaturants is not fully known, GdnHCl induced denaturation is highly reversible (Rumfeldt *et al.*, 2005) which may be a result of the greater solubilising ability towards polar rather than non polar amino acids resulting in stabilising the protein structure during refolding (Mayr *et al.*, 1993). This may account for the lack of

full refolding of ISP^{S250A}. It is thought that both urea and GdnHCl initially dispel water from around the protein and hydrogen bonding to peptide backbone NH. Following the breakage of inter-residue hydrogen bonds, the denaturants penetrate the core of the protein and preferentially hydrogen bond to the protein (backbone and polar side chains) thus breaking more inter-residue hydrogen bonds and the loss of structure leading to protein denaturation (Hua *et al.,* 2008, Lim *et al.*, 2009).

The thermal transition mid point (T_m) of ISP^{S250A} were similar using either CD or fluorescence (56°C and 55°C) (Figures 7.3). This is similar to the value measured by Shinde *et al.*, 2004 (58.7 °C) suggesting that the ISPs share similar stabilities. However there was a second transition observed using CD that gave a T_m of 69 °C. This may be the result of dimerisation of ISP from *Bacillus clausii*. The T_m values are only slightly lower than those seen with the ESP's, which are normally above 60°C (Bryan *et al.,* 1992, Tindbaek *et al.,* 2004). The ESP's are robust, compact proteases that are highly suited to the harsh environment into which they are secreted. The ISPs on the other hand have a more complex quaternary structure but still maintain high structural stability. This is confirmed by the values for $\Delta G_u^{H_2O}$ for ISP^{S250A}, which ranged from 8.5 to 13.3 kcalmol^{-1}. This suggests that ISP^{S250A} is stable, for example $\Delta G_u^{H_2O}$ value for cytochrome is 6.9 kcalmol^{-1} and ubiquitin is 7.0 kcalmol^{-1} (Jackson 1998). The *m* values were different when using either GdnHCl or urea (averaged to 2.8 and 1.1 kcalmol^{-1}, respectively). This is a common phenomenon where GdnHCl is normally twice as effective a denaturant than urea (Myers *et al.*, 1995). This is also observed case with ISP^{S250A}; the *m* values for ISP^{S250A} suggest that there is considerable change in stability upon unfolding with GdnHCl and is similar to most proteins in its susceptibility to the denaturants GdnHCl and urea. For example, α-chymotrypsin has *m* values using GdnHCl and urea of 4.1 and 2.1 kcalmol^{-1}, respectively (Myers *et al.*, 1995).

The ESPs also bind calcium at a high affinity site that contributes 5 kcalmol^{-1} to unfolding of the protease adding to the already high kinetic stability (Bryan *et al.,* 1992). This has been shown to be a large factor in the inability of ESPs to refold (Siezen *et al.,* 1997). As revealed in Chapter 4, ISP^{S250A} has reversible metal ion binding, therefore removing this kinetic barrier seen with the ESPs. This may also explain the high stability of ISP^{S250A}.

Most protein folding studies utilize small (<150 amino acids) monomeric proteins to investigate the pathway and mechanisms of folding. ISPs are a lot more

complicated both in terms of amino acid content (>300 amino acids) and quaternary structure (dimeric). Thus trying to accurately determine thermodynamic properties and folding pathways is difficult. However, we have attempted to address this challenge and endeavoured to devise a simple model by which to full length ISP^{S250A} fold. However even this provided a significant challenge, as outlined below, because the available data could not be routinely fitted to a singe defined model.

The ability of an ISP to refold has been shown by Subbian *et al.,* (2004) with an ISP from *Bacillus subtilis*. They observed similar ΔG values presented in this project (8.86 kcal/mol versus 8.5 to 13.3 kcal/mol). In addition the C_m values seen using GdnHCl were also very similar (0.8 and 2.5 versus 0.6 and 2 M (Figure 7.5)). Their view is similar to that formed within this project that ISPs and ESPs fold via alternative mechanisms that involved separate intermediate states. However, the model used in their study relied on a homology model that assumed a monomeric ISP structure rather than the dimer configuration used to fit data here and does not suggest why there is a difference in the folding intermediate between ESPs and ISPs.

It has also been shown here that an intermediate state during the unfolding and refolding of ISP may occur. Although the nature of this intermediate has not been identified it does not prevent the ISP from folding to an active conformation. The SEC profiles that shows an initial increase in hydrodynamic volume, with a decreased elution volume, in 2 M urea. This suggests that ISP^{S250A} may be either shifting to partially unfolded dimers or unfolded monomers. Since one may expect partially unfolded monomers to have a lower hydrodynamic volume and elute later than the dimer, this may suggest ISP is forming partially unfolded dimers. In 4M urea the peak observed in 2 M urea is still present but the major peak elutes much earlier suggesting a large increase in hydrodynamic volume. This may suggest that the breakdown of the dimer occurs during this stage before fully unfolding as the urea concentration is increased to 8 M. This hypothesis requires more data to support this view such a repetition of data points at key transitions and denaturation at different ISP^{S250A} concentrations.

Whilst some of the denaturation curves had regions that may have suggested additional transitions were present, more data points are required to address this. The majority of the unfolding/refolding curves fit to the two-state ($N_2 \leftrightarrow 2U$) equations with comparable thermodynamic parameters. Interestingly, ISP^{S250A} did not refold to the same extent when using urea in any of the techniques used to monitor refolding (Figure

7.1, 7.2, 7.6 and 7.7). The fact that all the methods displayed a similar extent of refolding suggests that this must be a result of the denaturant action on ISP^{S250A}. It may be the case that aggregates of protein are formed or that ISP is refolding via a different pathway when urea is used.

It will be important to establish which model is accurate for ISP folding as this may provide an insight into the evolution of thermodynamic rather than kinetic folding. The folding intermediate whether it is monomeric or dimeric may allow the ISP to circumvent the requirement for a prodomain essential within the ESPs. We have presented data suggesting that a dimeric intermediate occurs during the folding of ISP. This may suggest that the dimerisation acts as a stabilising factor, similar to the ESP's prodomain, directing correct folding to native conformation.

Within this section it has been shown that ISP can refold without a classical prodomain via thermodynamic rather than kinetic stability shown with the ESPs. There also appears to be an intermediate on the folding pathway. However, it was not determined if this is a monomeric or dimeric intermediate.

Chapter 8
General Discussion

8. General Discussion

The ISPs are a little studied family within the subtilisin family despite having approximately 40-50% sequence homology with the highly studied ESPs and constituting over 80% of the intracellular protease activity. However, the ISPs contain novel sequence elements for example, the N-terminal extension not present in the ESPs (Figure 5.1). Therefore one of the main objectives of this project has been to determine if these novel sequence elements of ISP translate into significant differences in both tertiary and quaternary structure. This in turn would lead us to prescribe functional roles to these novel sequences that will help elucidate their physiological relevance. This involves key features of the ISP such as 3D structure, quaternary structure, post-translational modification, enzymatic activity and folding mechanism. Therefore, to allow such a detailed analysis of the ISPs, the gene encoding a representative ISP from *Bacillus clausii* was cloned into an expression vector for the protein to be produced in a convenient host system (*E.coli* BL21Gold). One of the first important experimental findings was the ability to produce a large amount of soluble recombinant ISP protease in *E.coli*. This shows that the ISP protease can be produced in a folded and stable state without any detriment to the non-native bacterial host. This suggests that the proteolytic activity of ISP is not detrimental to the *E.coli* host proteins, unlike intracellular expressed ESPs (Subbian *et al.*, 2004).

8.1 Structure-function relationships of ISP

During this project we have collaborated with Professor Keith Wilson from York University who solved the first crystal structure of an ISP. The structure of a deactivated variant containing a mutation of the active site serine to an alanine (ISP^{S250A}), to avoid any problems due to unwanted proteolysis, was determined to 2.7 and 2.5 Å resolution. These structures allowed the analysis of the structure-function relationship, especially with regards to the sequence elements novel to ISP amongst the subtilases. As predicted from the relatively high sequence identity, ISP^{S250A} has a typical subtilisin core similar to the ESPs. The overall subtilisin structure is highly conserved, and the differences being restricted to a number of surface loops and the extensions of ISP^{S250A} at the N- and C-termini.

8.1.1 Quaternary structure of ISP

Whilst the general fold of ISP^{S250A} is similar to that of the bacterial ESPs the dimeric quaternary structure is a unique feature amongst the subtilisins. SEC and AUC analysis suggested that full length ISP^{S250A} formed dimers in line with other studies on the processed version (Strongin *et al.*, 1978, Koide *et al.*, 1986, Tsuchiya *et al.*, 1997). The structure allowed the dimer interface to be defined, which comprises residues predominantly from the C-terminal region of ISP^{S250A}. The C-terminal of ISP from *Bacillus clausii* shows very little sequence homology to that found in the ESPs and is not highly conserved within the ISPs (Figure 1.5). Therefore the role of this region was not even speculated.

The C-terminal extension is in ISP^{S250A}, residues 301-320, form an extensive crossover to the adjoining subunit making the major contribution to the interface. In addition helices comprising residues 277-287 from each subunit associate at the interface to stabilise the dimer. ISP dimerisation was initially thought to play a role in the regulation of ISP activity, however the two active sites lie far apart in the dimer on the opposite side of the protein from the dimer interface. Therefore the role of dimerisation in ISP maturation is unclear. However one may speculate that if dimerisation is not involved in the regulation of ISP activity, it may have a function similar to the ESPs prodomain and forms a scaffold allowing ISP to fold to its native mature conformation.

8.1.2 Role of the N-terminal extension

Apart from the C-terminus, another region in the primary structure of ISPs that differs significantly from the ESPs is at the N-terminus. The most notable difference is the absence of a traditional prodomain associated with the ESPs, which is instead replaced by a short N-terminal extension. The LIPY/F sequence motif in the N-terminal extension is conserved, hinting at a common role for this region amongst the ISPs. The ISPs are commonly isolated with the N-terminal extension post-translationally removed but it was not known what is responsible for this cleavage and whether this processing is relevant to function or just an artefact of protein purification (Strongin *et al.*, 1978, Sheehan and Switzer 1990, Tsuchiya *et al.*, 1997). Our analysis revealed that ISP from *Bacillus clausii* was produced as a full-length protein that was processed to a smaller

product, with the first 18 amino acids missing from the N-terminus and the His Tag section missing from the C-terminus. The full-length ISP was found to be inactive but activity was gained on processing. Given that *E.coli* in which ISP was expressed does not contain an ISP homolog, it is unlikely that the processing was an artefact produced by *E.coli* proteases. However, this was not enough evidence to assume processing was not an artefact of purification. Therefore the activity of ISP against ISP^{S250A} as a substrate was determined. This revealed identical processing, with the same 18 residues removed from the N-termini. Western blotting using a His Tag antibody verified that the C-terminal was present in the full ISP and during activation but was removed later. This suggests that the main processing event occurs at the N-terminal resulting in activation of ISP and that the N-terminal extension regulates ISP activity. The potential regulatory role of the N-terminal extension was confirmed as it acted as a potent inhibitor of proteolytic activity in its own right when added to processed ISP *in vitro* as a synthetic peptide. The peptide equivalent to the first 17 amino acids of ISP (excluding the N-terminal methionine) substantially reduced proteolytic activity with a calculated K_i of 5.0×10^{-7} M.

Analysis of the full length ISP^{S250A} structure provides an insight into the mode of inhibition by the N-terminal extension. The N-terminal extension is shown to bind back across the active site where main chain of ISP^{S250A} is shifted away from the active site by a kink induced by Pro8. This moves the scissile peptide bond out of position for hydrolysis to occur. The overall position of the catalytic triad in ISP^{S250A} is similar to that seen in other subtilisins, but with a significant shift in the position of the mutated alanine. The main chain of the catalytic Ala250 in the ISP^{S250A} mutant is displaced by 1.6 Å from that expected for formation of a typical catalytic triad, unlike the situation for equivalent mutants of the ESPs where the alanine main chain remains in essentially the same position as in the wild-type enzyme. This shift may contribute to the inhibition of ISP with the catalytic triad clicking into an active conformation on maturation.

This first structural analysis of an ISP therefore sheds light on the mechanism of inhibition of ISPs by the N-terminal extension, which differs from other modes of inhibition of subtilisins, and to our knowledge other serine proteases.

The inhibitory role of the N-terminal extension may be compared with two other modes of inhibition seen in the subtilisin family: inhibition of the mature ESP by its cognate prodomain before (Comellas-Bigler *et al.*, 2004, Tanaka *et al.*, 2007) and after (Gallagher *et al.*, 1995, Jain *et al.*, 1998) autoprocessing and inhibition by natural

peptide inhibitors (Bode and Huber 1992, Laskowski and Qasim, 2000), such as CI2 (Radisky and Koshland 2002). All three have a common element in their mode of inhibition, in that a peptide chain binds at the active site, preventing access of potential substrates to the active site. However, the nature of the inhibition varies significantly (Figure 8.1).

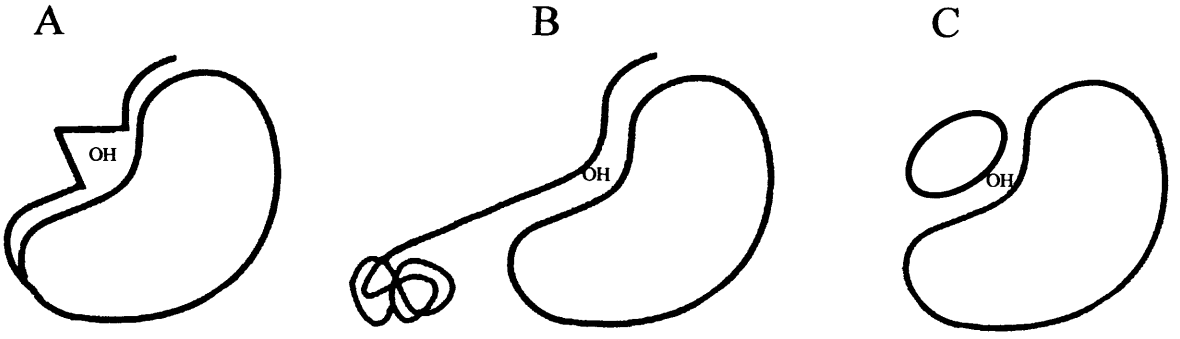


Figure 8.1 Different modes of inhibition of proteinacous inhibitors of subtilases. (A) ISP inhibition by N-terminal extension. (B) ESP inhibition by prodomain. (C) ESP inhibition by CI2. (OH represents hydroxyl group of the active site serine residue).

In comparison with the proESPs, the N-terminal extension of ISP^{S250A} occupies a space over the active site largely equivalent to that taken up by residues -1 to -4 from the autocatalytic cleavage point between the prodomain and mature ESP. The vital difference with ISP^{S250A} is the deviation of the main-chain away from the active site induced by Pro8, with a contribution from Tyr9 to promoting formation of the proline bulge, so shifting the target peptide bond out of position for hydrolysis to occur. In contrast, the proESP retains the scissile bond in a position to allow autocatalytic processing that leads to the subsequent detachment of the prodomain, as highlighted in the recently determined structures of the unautoprocessed pro-Tk subtilisin from the archaea *Thermococcus kodakarensis* (Tanaka *et al.*, 2007). The difference between the ISPs and ESPs reflects their different biological role; the N-terminal extension prevents ISP becoming active within the cell until required. In contrast the ESP prodomain is part of the folding mechanism to ensure the enzyme only forms a functional conformation on secretion from the cell. Unregulated protease activity with broad

substrate specificity within the cell, such as that exhibited by bacterial ESPs, can be lethal (Subbian *et al.*, 2004).

There are some similarities in the mode of binding with the N-terminal extension of ISP^{S250A} and the mechanism of subtilisin inhibition by natural proteinacous inhibitors such as CI2. In particular the main chain residues 3-7 (KFRLI) of ISP^{S250A} and 55-59 (TIVTK) of CI2 are in similar locations when bound. However, the ISP^{S250A} N-terminal sequence has a quite different orientation to that of CI2 bound to BPN' with the Pro8 causing a bulge in the backbone taking the scissile bond away from the active site. Furthermore, the interactions between CI2 and BPN' are restricted to the substrate binding pocket whereas the N-terminal extension of ISP^{S250A} interacts with regions of the protein beyond the substrate binding pocket. It has been suggested that intracellular inhibitors of ISPs exist to regulate protease activity (Shiga *et al.*, 1993) but the likely primary mode of inhibition is through the N-terminal extension.

Cleavage of the N-terminal extension at similar positions has been observed for ISPs from other species (Strongin *et al.*, 1978, Strongin *et al.*, 1979a, Tsuchiya *et al.*, 1997) and together with conservation of the LIPY/F motif suggests a common role for the N-terminal extension amongst the ISP class of subtilisins. However, the N-terminal cleavage point between Leu18 and Ser19 is on a separate face of the enzyme and is distant from the active site. As the whole of the N-terminal extension makes extensive contacts with the rest of the protein, therefore it is difficult to envisage autocatalytic removal of the N-terminal extension as occurs for prodomain processing in ESPs. It is clear that ISP itself is responsible for the precise processing of the N-terminal extension but this must be an inter- rather than intra-molecular event, in a manner akin to the trypsinogen activation, which requires the active protease trypsin to process and active the zymogen. This may explain the relatively slow activation of ISPs *in vitro* and *in vivo* (Ruppen *et al.*, 1988, Sheehan and Switzer 1990), as one would expect autoactivation to be a relatively quick, unimolecular process. Furthermore, given that Pro8 introduces a kink at the position where the N-terminal extension interacts with the active site preventing hydrolysis of the target peptide bond, it is unlikely that there are any intermediate autocatalytic processing events of the N-terminal extension occurring prior to its full removal.

8.1.3 Folding characteristics of ISP

The ESPs rely on their prodomain to act as an intramolecular chaperone, so as to catalyse folding of the mature subtilisin to reach its native functional conformation. The N-terminal extension of ISP is too short to facilitate folding in a manner akin to a classical prodomain. Our observations that processed ISP could refold to an active conformation substantiate this. This important result also demonstrated folding is reversible therefore ISP is thermodynamically stable as opposed to the mature forms of ESPs that are predominantly kinetically stable. Chemical denaturation studies with full length ISP^{S250A} confirmed the reversibility of folding, with a ΔG folding calculated to be -9.0 $kcalmol^{-1}$, which represents one of the first determination of the free energy of folding of the dimeric ISP and thus stability of a subtilisin.

Chemical and thermal denaturation studies suggested that an intermediate transition maybe present on the folding pathway. However it was not possible to determine if this was a monomeric or dimeric intermediate state. The SEC data suggested that a dimeric intermediate might be formed. The folding curves were fitted to a two-state $2U \leftrightarrow N_2$ transition, however there were suggestions that an intermediate state may occur especially with GdnHCl monitored via fluorescence. These two models are presented in a schematic in Figure 8.2.

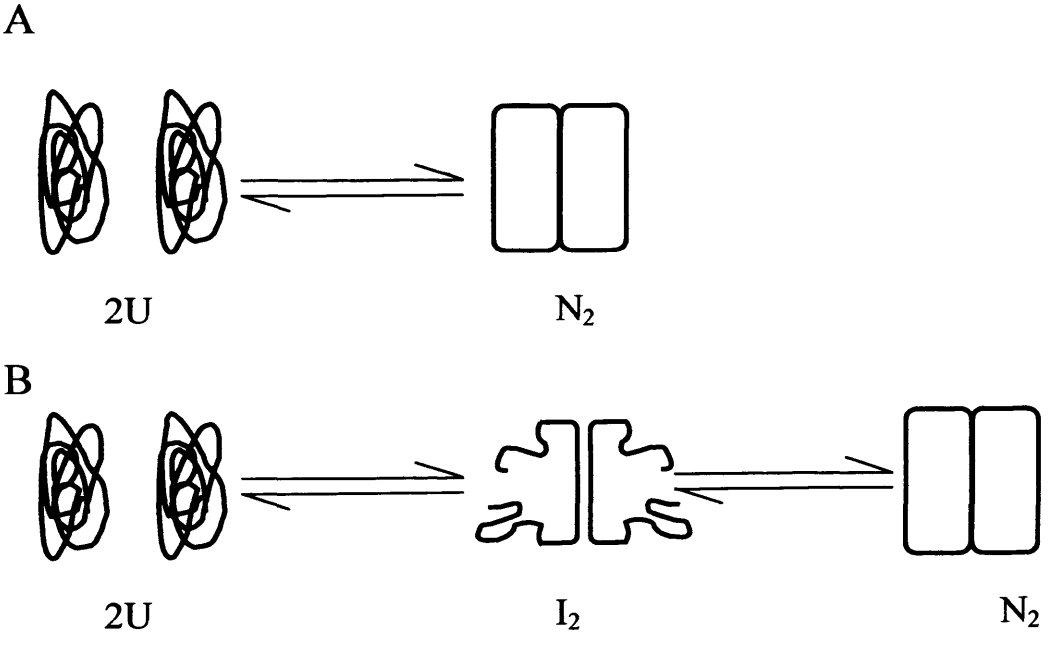


Figure 8.2 Potential folding pathways of ISP^{S250A}. (A) Two-state $2U \leftrightarrow N_2$ transition. (B) Three state $2U \leftrightarrow I_2 \leftrightarrow N_2$ transition.

It will be important to establish which model is accurate as this may give an insight into thermodynamic rather than kinetic stability. One may speculate that dimerisation of ISP may circumvent the requirement of a prodomain that is essential in the ESPs.

8.1.4 Metal ion binding of ISP

Calcium is known to play an important structural role in subtilisins, but the role of calcium in ISPs has not been clearly defined. Analysis of sequence alignments suggested that the high-affinity calcium-binding site present in the majority of bacterial ESPs is retained in ISPs, the differences being limited to conservative substitutions that still retain the potential to bind a metal ion.

The presence of a divalent metal ion chelator has a major influence on the structure and therefore activity of ISP. In the presence of EDTA, there were significant changes to the CD spectra of both ISP^{S250A} and the processed ISP suggesting a substantial conformational change. Proteolytic activity of ISP was also adversely affected, with a near total loss of ISP activity at EDTA concentrations greater than 0.01 mM. Although it is not definitive that calcium is the ligand bound at site A, the decrease in activity and breakdown in structure shown here indicate that metal ion binding is crucial for ISP structure and therefore function. The structure of ISP^{S250A} indicates that the coordination geometry at this site is very similar to the ESPs with the main chain carbonyl oxygens of Ala97, Ser101 and Val103 and the side chain oxygens of Asp58 and Thr99 coordinating the ligand. However, in the unprocessed form the sixth coordinating site contains a water molecule. One may speculate that this processing occurs near this site and may allow structural changes permitting a sixth coordinating ligand with residue Glu20 or Ser19 occupying this site.

8.1.5 Physiological role for ISPs

The physiological function of the ISPs is still unknown despite constituting approximately 80% of the *Bacillus* intracellular proteolytic activity (Burnett *et al* 1986, Orrego *et al* 1973). The bacterial ESPs are scavenging proteases with a low substrate specificity. Other members of the subtilase family include the proprotein convertases, which are involved in specific proteolysis of proproteins and have much higher

substrate specificities, which demonstrates that high substrate specificity can be built into the subtilisin scaffold (Van de Ven *et al.*, 1993). Initially the ISPs were thought to be involved in sporulation but it was later shown that bacteria lacking the ISP gene still sporulated normally (O'Hara 1990). To date the only role suggested for the ISPs involves the regulation of proteins such as ClpC and EF-Tu (Lee *et al.*, 2004). The proteins have diverse function within protein processing and tRNA ribosomal binding, respectively. These were identified using degradomic screening with the ISP from *B.subtilis* to degrade cellular extracts. ISP was shown to degrade the proteins with a preference for XXAA recognition sites. Therefore, ISP may have a role in protein regulation via proteolysis. In a broader context, ISP has been shown in this project to preferentially degrade unfolded protein substrates rather that protein in native conformation. This was in contrast to Savinase, used as a representative of the ESPs, which showed proteolytic degradation of the substrates in native or unfolded conformation. Enzyme kinetic analysis using chromogenic tetrapeptides revealed two preferred substrates for ISP, Suc-FAAF-pNA and Suc-AAPF-pNA. The k_{cat} values were similar (8 and 12 s^{-1}) however there was a large increase in K_M (0.066 and 6.5 mM) suggesting that ISP has a preference for phenylalanine at P1 and P4. It is not clear if proline at P2 adversely affects peptide binding but it is apparent that a large hydrophobic residue is preferred at P4. The remaining peptides assayed suggested that ISP tolerated hydrophobic residues at P1 and P2. However, ISP hydrolysed peptides containing a charged residue at P1 poorly, if at all. The ESPs demonstrate higher substrate turnover for all of the peptides tested. ISP activity was only comparable to the ESPs with the Suc-FAAF-pNA as a substrate however the k_{cat}/K_M parameters for ISP were still four times lower (418 and 121 $s^{-1}mM^{-1}$) (Tindbaek *et al.*, 2004). This confirmed the hypothesis that ISPs must have higher substrate specificity than the ESPs but also indicated that the ISP overall substrate turnover is lower. ISP preference for hydrophobic residues at P1 and P4 reinforces the evidence suggesting that the ISPs may play a role in processing of misfolded or unfolded proteins within the cell. It may be speculated that the presence of an intracellular protease akin to the ESPs would not be tolerated within the cell. Therefore, it is apparent that the core subtilisin fold has undergone divergent evolution to allow the precise and regulatable protease activity within the cell. The question still remains as to the nature of the common ancestor and whether the ISPs represent the more ancient ancestor than the ESPs, with elements such as a prodomain and signal sequence being incorporated later. The degradation of

misfolded or denatured proteins is normally tightly regulated within both eukaryotic and prokaryotic cells. As a first option, instead of degradation, cells utilise chaperone or stabilising factors to refold protein. Unlike eukaryotic cells bacteria are less able to organise misfolded protein into separate organelles for sorting. Misfolded or aggregated protein is formed into inclusion bodies for refolding or degradation via proteolysis. However, under stressful conditions, such as changes in temperature or pH, or during stationary phase, misfolded or aggregated protein is degraded by proteases for nutrition. Some of the main proteases involved in degradation of abnormal proteins are the Lon and Clp proteases. As discussed earlier, ISP may have a role in regulation of ClpC by proteolysis. The Lon protease is directly involved in degrading approximately 50% of abnormal protein in *E.coli* as well as its many other roles of protein processing within the cell cycle and differentiation (Melderen and Aertsen 2009). The Bacillus species generally encode two Lon proteases but they are not responsible for degradation of misfolded proteins. Therefore it is feasible that ISPs may have a role in protein processing within Bacillus either as a proteolytic regulatory protease or in the degradation of misfolded protein.

8.2 Future work

8.2.1 Structure of ISP

The structure of ISP^{S250A} has been determined, however there are still questions to be addressed. Processed ISP has been shown to refold to 88% activity, which suggests that the N-terminal extension does not play a role in protein folding. Furthermore, rudimentary structural analysis (e.g. CD and fluorescence) suggests that processing does not result in any major conformational changes. To address these ideas, engineered truncated versions of ISP with the N-terminal region removed could be produced and subjected to folding and crystallography studies. The latter would be particularly interesting, as we could map the fine structural changes occurring on processing that further define the regulation mechanism of the N-terminal extension. A variant with the native serine at the catalytic position 250 found in wild-type ISP would be required for activity studies and a deactivated variant, with a serine to alanine mutation at position 250, required for structure determination by X-ray crystallography and structural assays. The processed active ISP is stable but is still a potential protease

and, as established in Section 6.2, any unfolded protein may be degraded and adversely affect crystallography.

The exact metal ion binding within the ISP structure is unclear especially if there is any metal binding occurring at the dimer interface. The structure of ISP indicates that the high affinity calcium-binding site found in the ESPs is conserved and that the ion is highly likely to be calcium. Removal of the metal ion from ISP had dramatic effects on the structure of ISP including dimerisation. Therefore it is important to understand the structure of metal ion binding and its importance in protein stability.

The dimerisation of ISPs appears to be key in folding and thermodynamic stability as oppose to kinetic folding found in the ESPs. It would therefore be crucial to determine the K_d of the dimer interaction and understand conditions where the interaction breaks down, such as changes in pH or ionic buffering. Using the structure as a guide, one could systematically analyse the contribution of residues at the dimer interface through protein engineering.

There is only approximately 50% sequence identity between the ISPs therefore it would be interesting to determine what exact changes are responsible for the different structural and functional properties as opposed to those occurring due to genetic drift. For example, the LIPY region within the N-terminal extension is highly conserved but not all of the N-terminal extension. This could be achieved through the study of an ISP from a different Bacilli species such as *B. subtilis* or *B. anthracis.*

8.2.2 Function of ISPs

It has been shown using chromogenic peptides that ISPs have a preference for large hydrophobic residues at P1 and P4. This translates at a protein level to a preference for unfolded substrates. It would be important for understanding the physiological function of the ISPs to use *in vivo* assays such as gene deletion or deactivation. It would also be interesting to analyse the enzymatic activity of ISP against different peptide substrates, with changes at P2, and protein substrates especially native Bacillus proteins to identify potential protein targets. Ideally one could take a proteomic approach to determine which proteins are degraded at different stages of the cellular life cycle.

The ESPs are used extensively in the biotechnology industry, however their reliance of a prodomain means that structural changes to increase functionality have to ensure the prodomain mediated folding is not inhibited. It would therefore be beneficial if the ESPs could fold independently of a prodomain as the ISPs can. If this is not possible perhaps ISPs may be a scaffold to evolve increased stability and functionality. Furthermore, there is no known human homolog of ISP. However, they have been found in species that are known pathogens such as *Bacillus anthracis* and *Clostridium difficile* therefore the ISPs may be a target for drug design such as inhibitors. However we would need to establish how essential the ISPs is to the bacterial cell.

8.2.3 Folding of ISP

The folding studies presented in this work demonstrated the reversible thermodynamic folding of ISP without the requirement of a prodomain. It would be important to identify the exact mechanism for prodomain independent folding of ISP. This would involve analysing the role of ISP dimer formation to determine if this is a stabilising factor forming, for example, a scaffold for folding. The folding studies also indicated a potential monomeric or dimeric intermediate transition on the folding pathway. Repeating the folding assays with more data points especially around the transition may clearly demonstrate if the transition is genuine or noise in the data. The nature of any transition is important in understanding the role of dimerisation in ISP folding and its significance in stabilisation of ISP in the absence of a classical prodomain. This may be accomplished by changing the protein concentrations to determine where concentration dependent differences are observed in the thermodynamic parameters.

The data presented in this project represents the first detailed molecular analysis of an ISP. It has hopefully put forward some answers to questions such as the role of the conserved N-terminal extension, the dimeric quaternary structure of ISPs and a physiological role. It also suggests that the ISPs may be used as a realistic tool for analysis of, for example, the folding properties of homologous proteins in terms of thermodynamic versus kinetic stability and novel inhibition mechanisms.

References

Almog, O., Gallagher, T., Tordova, M., Hoskins, J., Bryan, P., Gilliland, G. L. 1998. Crystal structure of calcium-independent subtilisin BPN' with restored thermal stability folded without the prodomain. *Proteins.* Apr 1;**31**(1):21-32.

Amara, U., Rittirsch, D., Flierl, M., Bruckner, U., Klos, A., Gebhard, F., Lambris, J.D., and Huber-Lang, M. 2008. Interaction between the coagulation and complement system. *Adv Exp Med Biol* **632:** 71-79.

Anfinsen, C. 1973. Principles that govern the folding of protein chains. *Science* **181:** 223-230.

Band, L., Henner, D.J., and Ruppen, M. 1987. Construction and properties of an intracellular serine protease mutant of Bacillus subtilis. *J. Bacteriol.* **169:** 444-446.

Bankar, S.B., Bule, M.V., Singhal, R.S., and Ananthanarayan, L. 2009. Glucose oxidase-an overview. *Biotechnol Adv* **27:** 489-501.

Barrett, A., and Rawlings, N. 1995. Families and clans of serine peptidases. *Archives of Biochemistry and Biophysics* **318:** 247-250.

Barrett, A.J., and McDonald, J.K. 1986. Nomenclature: protease, proteinase and peptidase. *Biochem J* **237:** 935.

Barrett, A.J., and Rawlings, N.D. 2007. 'Species' of peptidases. *Biol Chem* **388:** 1151-1157.

Beynon, R., and Bond, J. 1989. Proteolytic enzymes - A practical approach, pp. 59. IRL Press, Oxford.

Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S., Wilson, K.S. 1992. Crystal structure of the alkaline proteinase Savinase from Bacillus lentus at 1.4 A resolution. *J Mol Biol.* **20**;223(2):427-45

Blow, D.M. 1976. Structure and mechanism of chymotrypsin. *Acc. Chem. Res.* **9:** 145-152.

Blow, D.M., Birktoft, J.J., and Hartley, B.S. 1969. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* **221:** 337-340.

Bode, W., and Huber, R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* **204:** 433-451.

Bode, W., Papamokos, E., and Musil, D. 1987. The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech Hirudo medicinalis. Structural analysis, subtilisin structure and interface geometry. *Eur. J. Biochem.* **166:** 673-692.

Borgese .N., Brambillasca. S., Colombo. S. 2007. How tails guide tail-anchored proteins to their destinations. *Curr Opin Cell Biol.* Aug;**19**(4):368-75.

Bryan, P. 1987. Protein Engineering. *Biotech. Adv.* **5:** 221-234.

Bryan, P. 2000. Protein engineering of subtilisin. *Biochemica et biophysica acta* **1543:** 203-222.

Bryan, P., Alexander, P., Strausberg, S., Schwarz, F., Lan, W., Gilliland, G., Gallagher, D.T. 1992. Energetics of folding subtilisin BPN'. *Biochemistry.* Jun 2;**31**(21):4937-45.

Bullock, T.L., Breddam, K., and Remington, S.J. 1996. Peptide aldehyde complexes with wheat serine carboxypeptidase II: implications for the catalytic mechanism and substrate specificity. **255**(5):714-25.

Burnett, T.J., Shankweiler, G.W., and Hageman, J.H. 1986. Activation of intracellular serine proteinase in *Bacillus subtilis* cells during sporulation. *J Bacteriol* **165:** 139-145.

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

Cheng, Y., and Aronson, A. 1977. Alterations in spore coat processing and protein trunover in a Bacillus cereus mutant with a defective postexponential intracellular protease. *Proc Natl Acad Sci U S A* **74:** 1254-1258.

Chilson, O., and Chilson, A. 2003. Perturbation of folding and reassociation of lactate dehydrogenase by proline and trimethylamine oxide. *Eur. J. Biochem* **270:** 4823-4834.

Coates, L., Tuan, H.F, Tomanicek, S., Kovalevsky, A., Mustyakimov, M., Erskine, P., and Cooper, J., 2008. The Catalytic Mechanism of an Aspartic Proteinase Explored with Neutron and X-ray Diffraction. *J. Am. Chem. Soc* **130**, 7235–7237 9 7235

Comellas-Bigler, M., Maskos, K., Huber, R., Oyama, H., Oda, K., and Bode, W. 2004. 1.2 A crystal structure of the serine carboxyl proteinase pro-kumamolisin; structure of an intact pro-subtilase. *Structure* **12:** 1313-1323.

Dalboge, H., Bayne, S., and Pedersen, J. 1990. *In vivo* processing of N-terminal methionine in *E. coli*. *FEBS Lett* **266:** 1-3.

Davail, S., Feller, G., Narinx, E., Gerday, C. 1994. Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile Bacillus TA41. *J Biol Chem.* **269**(26):17448-53.

Davidson, C.J., Tuddenham, E.G., and McVey, J.H. 2003. 450 million years of hemostasis. *J Thromb Haemost* **1:** 1487-1494.

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

Davies, D.R. 1990. The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* **19:** 189-215.

Davis, B., Khumtaveeporn, K., Bott, R., and JB, J. 1999a. Altering the specificity of subtilisin *Bacillus lentus* through the introduction of positive charge at single amino acid sites. *Bioorg Med Chem* **11:** 2303-2311.

Davis, B.G., Khumtaveeporn, K., Bott, R.R., and Jones, J.B. 1999b. Altering the specificity of subtilisin Bacillus lentus through the introduction of positive charge at single amino acid sites. *Bioorg Med Chem* **7:** 2303-2311.

Dee, D.R., Filonowicz, S., Horimoto, Y., and Yada, R.Y. 2009. Recombinant prosegment peptide acts as a folding catalyst and inhibitor of native pepsin. *Biochim Biophys Acta.* **1794**(12):1795-801

Derewenda, Z.S., Derewenda, U., and Kobos, P.M. 1994. (His)C epsilon-H...O=C < hydrogen bond in the active sites of serine hydrolases. *J Mol Biol* **241:** 83-93.

Dobson, C. 2004. Principals of protein folding, misfolding and aggregation. *Seminars in Cell and Developmental Biology* **15:** 3-16.

Dodson, G., and Wlodawer, A. 1998. Catalytic triads and their relatives. *Trends Biochem Sci* **23:** 347-352.

Donepudi, M., Grütter MG. Structure and zymogen activation of caspases. 2002. *Biophys Chem.* Dec 10;101-102:145-53

Dubendorff, J.W., and Studier, F.W. 1991. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* **219:** 45-59.

Eder, J., Rheinnecker, M., and Fersht, A. 1993a. Folding of subtilisin BPN': characterisation of a folding intermediate. *Biochemistry* **32:** 18-26.

Eder, J., Rheinnecker, M., and Fersht , A. 1993b. Folding of subtilisin BPN': Role of the pro-sequence *J Mol Biol* **233:** 293-304.

Eder, J., and Fersht, AR. 1995. Pro-sequence-assisted protein folding. *Mol Microbiol.* **16**(4):609-14.

Fisher, K.E., Ruan, B., Alexander, P.A., Wang, L., and Bryan, P.N. 2007. Mechanism of the kinetically-controlled folding reaction of subtilisin. *Biochemistry* **46:** 640-651.

Fu, X., Inouye, M., and Shinde, U. 2000. Folding pathway mediated by an intramolecular chaperone. The inhibitory and chaperone functions of the subtilisin propeptide are not obligatorily linked. *J Biol Chem* **275:** 16871-16878.

Fujinaga, M., Cherney, M.M., Oyama, H., Oda, K., and James, M.N. 2004. The molecular structure and catalytic mechanism of a novel carboxyl peptidase from Scytalidium lignicolum. *Proc Natl Acad Sci U S A* **101:** 3364-3369.

Gallagher, D.T. 1992. Energetics of folding subtilisin BPN'. *Biochemistry* **31:** 4937-4945.

Gallagher, T., Gilliland, G., Wang, L., and Bryan, P. 1995. The prosegment-subtilisin BPN' complex: crystal structure of a specific 'foldase'. *Structure* **3:** 907-914.

Georgieva, D.N., Stoeva, S., Voelter, W., Genov, N., and Betzel, C. 2001a. Differences in the specificities of the highly alkalophilic proteinases Savinase and Esperase imposed by changes in the rigidity and geometry of the substrate binding sites. *Arch Biochem Biophys* **387:** 197-201.

Georgieva, D.N., Stoeva, S., Voelter, W., Genov, N., and Betzel, C. 2001b. Substrate specificity of the highly alkalophilic bacterial proteinase esperase: relation to the x-ray structure. *Curr Microbiol* **42:** 368-371.

Gleason, W., Fu, Z., Birktoft, J., and Banaszak, L. 1994. Refined crystal structure of mitochondrial malate dehydrogenase from porcine heart and the consensus structure for dicarboxylic acid oxidoreductases. *Biochemistry* **33:** 2078-2088.

Gottesman, S., Clark, W.P., and Maurizi, M.R. 1990. The ATP-dependent Clp protease of Escherichia coli. Sequence of clpA and identification of a Clp-specific substrate. *J Biol Chem* **265:** 7886-7893.

Gottesman, S., Roche, E., Zhou, Y., and Sauer, R.T. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12:** 1338-1347.

Grøn, H., and Breddam, K. 1992. Interdependency of the binding subsites in subtilisin. *Biochemistry*. Sep 22;**31**(37):8967-71.

Hageman, J.H., and Carlton, B.C. 1973. Effects of mutational loss of specific intracellular proteases on the sporulation of Bacillus subtilis. *J Bacteriol* **114**: 612-617.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166:** 557-580.

Hedstrom, L. 2002. Serine protease mechanism and specificity. *Chem Rev* **102:** 4501-4524.

Hengen, P. 1995. Purification of His-Tag fusion proteins from *Escherichia coli. Trends Biochem Sci* **20:** 285-286.

Herbaud, M., L., Guiseppi, A., Denizot, F., Haiech, J., and Kilhoffer, M., C. 1998. Calcium signalling in *Bacillus subtilis. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research.* December **10**;212-226

Hernick, M., and Fierke, C.A. 2005. Zinc hydrolases: the mechanisms of zinc-dependent deacetylases. *Archives of Biochemistry and Biophysics* **433** 71–84

Horimoto, Y., Dee, D.R., and Yada, R.Y. 2009. Multifunctional aspartic peptidase prosegments. *Nat Biotechnol* **25:** 318-324.

Hu, Z., Zhu, X., Jordan, F., and Inouye, M. 1994. A Covalently Trapped Folding Intermediate of Subtilisin E: Spontaneous Dimerization of a Prosubtilisin E Ser49Cys Mutant in Vivo and Its Autoprocessing *in Vitro*. *Biochemistry* **33:** 562-569.

Hua, L., Zhou, R., Thirumalai, D., and Berne, B.J. 2008. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci U S A* **105:** 16928-16933.

Huang, H.W., Chen, W.C., Wu, C.Y., Yu, H.C., Lin, W.Y., Chen, S.T., and Wang, K.T. 1997. Kinetic studies of the inhibitory effects of propeptides subtilisin BPN' and Carlsberg to bacterial serine proteases. *Protein Eng* **10:** 1227-1233.

Huntington, J.A. 2008. How Na+ activates thrombin--a review of the functional and structural data. **389:** 1025-1035. Production of active subtilisin E in *Escherichia coli. J Biol Chem* **252:** 7859-7864.

Ikemura, H., Takagi, H., and Inouye, M. 1987. Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli. J Biol Chem.* **262**(16):7859-64.

Inouye, M., Fu, X., and Shinde, U. 2001. Substrate-induced activation of a trapped IMC-mediated protein folding intermediate. *Nat Struct Biol* **8:** 321-325.

Jackson, S.E. 1998. How do small single-domain proteins fold? *Fold Des.* **3**(4):R81-91.

Johnson, WC., 1999. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins.* May 15;**35**(3):307-12.

Jonák, J., Pokorná, K., Meloun ,B., Karas, K. 1986. Structural homology between elongation factors EF-Tu from Bacillus stearothermophilus and Escherichia coli in the binding site for aminoacyl-tRNA. *Eur J Biochem.* Jan 15;**154**(2):355-62.

Kataoka, Y., Takada, K., Oyama, H., Tsunemi, M., James, M.N., and Oda, K. 2005. Catalytic residues and substrate specificity of scytalidoglutamic peptidase, the first member of the eqolisin in family (G1) of peptidases. *FEBS Lett* **579:** 2991-2994.

Kerjan, P., Keryer, E., and Szulmajster, J. 1979. Characterization of a thermosensitive sporulation mutant of Bacillus subtilis affected in the structural gene of an intracellular protease. *Eur J Biochem* **98:** 353-362.

Khan, A.R., and James, M.N. 1998. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci* **7:** 815-836.

Kneusel, R., Crowe, J., Wulbeck, M., and Ribbe, J. 1998. Procedures for the Analysis and Purification of His-Tagged Proteins. *Molecular Diagnosis of Infectious Diseases.* (ed. U. Reischl), pp. 293. Humana Press.

Koide, Y., Nakamura, A., Uozumi, T., and Beppu, T. 1986. Cloning and sequenceing of the major intracellular serine protease gene of *Bacillus subtilis. J Bacteriol* **167:** 110.

Krüger, E., Witt,E., Ohlmeier, S., Hanschke, R., and Hecker, M. 2000. The Clp Proteases of Bacillus subtilis Are Directly Involved in Degradation of Misfolded Proteins. *J Bacteriology.* **182**(11): 3259-3265

Kumar, C., Malik, R., and Tiwari, M. 1998. Novel enzyme-based detergents: An Indian perspective. *Current science* **75:** 1312-1318.

Kurotsu, T., Marahiel, M., Muller, K., and Horst, K. 1982. Characterisation of an intracellular serine protease from sporulating cells of *Bacillus brevis. J Bacteriol* **151:** 1466-1472.

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

Laskowski, M., Jr., and Kato, I. 1980. Protein inhibitors of proteinases. *Annu Rev Biochem* **49:** 593-626.

Laskowski, M., and Qasim, M.A. 2000. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim Biophys Acta* **1477:** 324-337.

Lee, A.Y., Goo Park, S., Kho, C.W., Young Park, S., Cho, S., Lee, S.C., Lee, D.H., Myung, P.K., and Park, B.C. 2004. Identification of the degradome of Isp-1, a major intracellular serine protease of Bacillus subtilis, by two-dimensional gel electrophoresis and matrix- assisted laser desorption/ionization-time of flight analysis. *Proteomics* **4:** 3437-3445.

Lew, R.A., Boulos, E., Stewart, K.M., Perlmutter, P., Harte, M.F., Bond, S., Reeve, S.B., Norman, M.U., Lew, M.J., Aguilar, M.I., et al. 2001. Substrate analogs incorporating beta-amino acids: potential application for peptidase inhibition. *FASEB J* **15:** 1664-1666.

Li, Y., Hu, Z., Jordan, F., and Inouye, M. 1995. Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants. *J Biol Chem* **270:** 25127-25132.

Li, Y., and Inouye, M. 1994. Autoprocessing of prothiolsubtilisin E in which active site serine 221 is altered to cysteine *J Biol Chem* **269:** 4169-4174.

Liao, Y.D., Jeng, J.C., Wang, C.F., Wang, S.C., and Chang, S.T. 2004. Removal of N-terminal methionine from recombinant proteins by engineered *E. coli* methionine aminopeptidase. *Protein Sci.* **13**(7):1802-10

Lim, W.K., Rosgen, J., and Englander S.W. 2009. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *PNAS.* **106** (8) 2595-2600

Magill, N., Cowan, A., Koppel, D., and Setlow, P. 1994. The internal pH of the forespore compartment of *Bacilus magaterium* decreases by about 1 pH unit during sporulation. *J Bacteriol* **176:** 2252.

Magill, N.G., Cowan, A.E., Leyva-Vazquez, M.A., Brown, M., Koppel, D.E., and Setlow, P. 1996. Analysis of the relationship between the decrease in pH and accumulation of 3-phosphoglyceric acid in developing forespores of Bacillus species. *J Bacteriol* **178:** 2204-2210.

Matagne, A., Joris, B., and Frere, J.M. 1991. Anomalous behaviour of a protein during SDS/PAGE corrected by chemical modification of carboxylic groups. *Biochem J* **280 (Pt 2):** 553-556.

Matthews, B.W. 1988. Structural basis of the action of thermolysin and related zinc peptidases,. *Acc. Chem. Res.* **21:** 333-340.

Matthews, D.A., Alden, R.A., Birktoft, J.J., Freer, T., and Kraut, J. 1977. Re-examination of the charge relay system in subtilisin comparison with other serine proteases. *J Biol Chem* **252:** 8875-8883.

Mayr, L.M., and Schmid, F.X. 1993. Stabilization of a protein by guanidinium chloride. *Biochemistry* **32:** 7994-7998.

McPhalen, C.A., and James, M.N.G. 1988. Structural Comparison of Two Serine Proteinase-Protein Inhibitor Complexes: Eglin-C-Subtilisin Carlsberg and CI-2-Subtilisin Novo? *Biochemistry.* **27:** 6582-6598

Millet, J. 1971. Characterization of a cytoplasmic endopeptidase in *Bacillus megaterium* undergoing sporulation. *C R Acad Sci Hebd Seances Acad Sci D.* **272**(13):1806-9

Misselwitz, R., Hausdorf, G., Welfle, K., Höhne, WE., Welfle H. 1995. Conformation and stability of recombinant HIV-1 capsid protein p24 (rp24). *Biochim Biophys Acta.* Jul 3;**1250**(1):9-18.

Morihara, K. 1987. Using proteases in peptide synthesis. *Trends in Biotechnology* **5:** 164-170.

Mortensen, U. H., and Breddam, K. 1994. A conserved glutamic acid bridge in serine carboxypeptidases, belonging to the cx/p hydrolase fold, acts as a pH-dependent protein-stabilizing element. *Protein Science* **3**: 838-842.

Mulder, F., Schipper, D., Bott, R., and Boelens, R. 1999. Altered flexibility in substrate-binding site of related native and engineered high alkaline Bacillus subtilisins. J *Mol Biol* **292:** 111-123.

Myers, J.K.,Pace, C.N., and Scholtz, J. M. 1995. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Science.* **4**: 2138-2148

Nakayama, T., Munoz, L.E., Sadaie, Y., and Doi, R.H. 1978. Spore coat protein synthesis in cell-free systems from sporulating cells of *Bacillus subtilis. J. Bacteriol.* **135:** 952-960.

Neurath, H. 1984. Evolution of proteolytic enzymes. *Science* **224:** 350-358.

O'Hara, M.B., and Hageman, J.H. 1990. Energy and calcium ion dependence of proteolysis during sporulation of *Bacillus subtilis* cells. *J Bacteriol* **172:** 4161-4170.

Ong, B.E., Shaw, E., and Schoellmann, G. 1964. An Active Center Histidine Peptide of α-Chymotrypsin. *Journal of the American Chemical Society*: 1271–1272.

Oosterbaan, R.A., van, A., and Cohen, J.A. 1962. An acetyl-peptide from acetylchymotrypsin. *Biochim Biophys Acta* **63:** 204-206.

Orrego, C., Kerjan, P., Manca de Nadra, M.C., Szulmajster, J. 1973. Ribonucleic acid polymerase in a thermosensitive sporulation mutant (ts-4) of *Bacillus subtilis. J Bacteriol.* **116**(2): 636-647

Pantoliano, M. W., Whitlow, M., Wood, J. F., Rollence, M. L., Finzel, B. C., Gilliland, G. L., Poulos, T. L., and Bryan P. N. 1988. The engineering of binding affinity at metal ion binding sites for the stabilization of proteins: subtilisin as a test case. *Biochemistry* **27:** 8311-17.

Paulick MG, Bertozzi CR. 2008 The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry.* **47**(27):6991-7000

Pejler, G., Abrink, M., Ringvall, M., and Wernersson, S. 2007. Mast cell proteases. *Adv Immunol* **95:** 167-255.

Perry, J., Stanley, J., and Lory, S. 2002. Chapter 31 Industrial microbiology Part VII Applied microbiology. *Microbial life*, pp. 779-. Sinaver associated publishing, Sunderland, Massecussetts.

Pham, C.T. 2008. Neutrophil serine proteases fine-tune the inflammatory response. *Int J Biochem Cell Biol* **40:** 1317-1333.

Pierre Kerjan, E.K., Jekisiel Szulmajster,. 1979. Characterization of a thermosensitive sporulation mutant of *Bacillus subtilis* affected in the structural gene of an intracellular protease. *European Journal of Biochemistry* **98:** 353-362.

Pshezhetsky, A.V., and Hinek, A. 2009. Serine carboxypeptidases in regulation of vasoconstriction and elastogenesis. *Trends Cardiovasc Med* **19:** 11-17.

Radisky, E. S., and Koshland, D. E., Jr. 2002. A clogged gutter mechanism for protease inhibitors. *Proc Natl Acad Sci U S A* **99:** 10316-10321.

Radisky, E.S., Kwan, G., Karen Lu, C.J., and Koshland, D.E., Jr. 2004. Binding, proteolytic, and crystallographic analyses of mutations at the protease-inhibitor interface of the subtilisin BPN'/chymotrypsin inhibitor 2 complex. *Biochemistry* **43:** 13648-13656.

Radisky, E.S., Lu, C.J., Kwan, G., Koshland, D.E., Jr. 2005. Role of the intramolecular hydrogen bond network in the inhibitory power of chymotrypsin inhibitor 2. *Biochemistry*. May 10;**44**(18):6823-30.

Rastall, R. 2007. Novel Enzyme Technology for Food Applications. Woodhead publishing limited. Abington, Cambridge, UK. ISBN 1 84569 132 6

Rawlings, N., and Barrett, A. 1994. Families of serine peptidases. *Methods in Enzymology* **244:** 19-61.

Read, T. D., Peterson, S. N., Tourasse, N., Baillie, L. W., Paulsen, I. T., Nelson, K. E., Tettelin, H., Fouts, D. E., Eisen, J. A., Gill, S. R., Holtzapple, E. K., Okstad, O. A., Helgason, E., Rilstone, J., Wu, M., Kolonay, J. F., Beanan, M., J., Dodson, R. J., Brinkac, L. M., Gwinn, M., DeBoy, R. T., Madpu, R., Daugherty, S. C., Durkin, A. S., Haft, D. H., Nelson, W. C., Peterson, J. D., Pop, M., Khouri, H. M., Radune, D., Benton, J. L., Mahamoud, Y., Jiang, L., Hance, I. R., Weidman, J. F., Berry, K. J., Plaut, R. D., Wolf, A. M., Watkins, K. L., Nierman, W. C., Hazen, A., Cline, R., Redmond, C., Thwaite, J.E., White, O., Salzberg, S.L., Thomason, B., Friedlander, A. M., Koehler, T.M., Hanna, P.C., Kolstø, A.B., Fraser, C.M. 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature.* **423**(6935):81-6.

Ricardo, R., and Phelan, K. 2008. Trypsinizing and subculturing mammalian cells. *J Vis Exp.* **12**;(16). pii: 755.

Richter, C., Tanaka, T., and Yada, R.Y. 1998. Mechanism of activation of the gastric aspartic proteinases: pepsinogen, progastricsin and prochymosin. *Biochem J* **335 (Pt 3):** 481-490.

Rockwell, N.C., and Fuller, R.S. 2002. Specific modulation of Kex2/furin family proteases by potassium. *J Biol Chem.* **277**(20):17531-7

Rockwell, N.C., and Thorner, J.W. 2004 The kindest cuts of all: crystal structures of Kex2 and furin reveal secrets of precursor processing. *Trends Biochem Sci.* **29**(2):80-7

Royer, C.A. 2006a. Probing Protein Folding and Conformational Transitions with Fluorescence. *Chemical Reviews* **106:** 1769-1784.

Ruan, B., London, V., Fisher, K.E., Gallagher, D.T., and Bryan, P.N. 2008. Engineering Substrate Preference in Subtilisin: Structural and Kinetic Analysis of a Specificity Mutant. *Biochemistry.* Jun 24;**47**(25):6628-36.

Rumfeldt, J.A.O., Stathopulos, P.B., Chakrabarrty, A., Lepock, J.R., Meiering, E.M. 2005. Mechanism and Thermodynamics of Guanidinium Chloride-induced Denaturation of ALS-associated Mutant Cu,Zn Superoxide Dismutases. *J Mol Biol.* **355** (1). 106-123.

Ruppen, M.E., Van Alstine, G.L., and Band, L. 1988. Control of intracellular serine protease expression in *Bacillus subtilis. J Bacteriol* **170:** 136-140.

Sambrook, J.F., EF Maniatis, T. 1989. Bacterial media, antibiotics and bacterial strains. In *Molecular cloning - A laboratory manual*, 2nd ed, pp. A1. Cold Spring Harbor Laboratory Press.

Sastry, K.J., Srivastava, O.P., Millet, J., FitzJames, P.C., and Aronson, A.I. 1983. Characterization of *Bacillus subtilis* mutants with a temperature-sensitive intracellular protease. *J. Bacteriol.* **153:** 511-519.

Schechter, I., and Berger, A. 1967. *Biochem. Biophys. Res. Commum* **27:** 157.

Schoellmann, G., and Shaw, E. 1963. Direct evidence for the presence of histidine in the active center of chymotrypsin. *Biochemistry* **2:** 252-255.

Schmidt, R., Decatur, A, L., Rather, P., Moran, C.P., and Losickl, R. 1994. *Bacillus subtilis* lon protease prevents inappropriate transcription of genes under the control of the sporulation transcription factor. *J Bacteriol.* **176** (21) 6528-6537

Sebaihia, M., Wren, B.W, Mullany, P., Fairweather, N.F, Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeño-Tárraga, A.M., Wang, H., Holden, M.T., Wright, A., Churcher, C., Quail, M.A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B., Parkhill, J. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet.* **38**(7):779-86.

Seidah, N.G., and Chretien, M. 1999. Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res* **848:** 45-62.

Sheehan, S., and Switzer, R. 1990a. Intracellular serine protease 1 of *Bacillus subtilus* is formed *in vivo* as an unprocessed, active protease in stationary phase. *J Bacteriol* **172:** 473-476.

Sheehan, S.M., and Switzer, R.L. 1990b. Intracellular serine protease 1 of *Bacillus subtilis* is formed in vivo as an unprocessed, active protease in stationary cells. *J Bacteriol* **172:** 473-476.

Sheehan, S.M., and Switzer, R.L. 1991. Intracellular serine protease-4, a new intracellular serine protease activity from *Bacillus subtilis. Arch Microbiol* **156:** 186-191.

Shiga, Y., Yamagata, H., and Udaka, S. 1993. Characterization of the gene encoding an intracellular proteinase inhibitor of *Bacillus subtilis* and its role in regulation of the major intracellular proteinase. *J. Bacteriol.* **175:** 7130-7137.

Shinde, U., Fu, X., and Inouye, M. 1999a. A pathway for conformational diversity in proteins mediated by intramolecular chaperones. *J Biol Chem* **274:** 15615-15621.

Shinde, U., Fu, X., and Inouye, M. 1999b. A pathway for conformational diversity in proteins mediated by intramolecular chaperones *J Biol Chem* **274:** 15615-15621.

Shinde, U., Fu, X., and Inouye, M. 2000. Intramolecular chaperones: polypeptide extensions that modulate protein folding. *Cell & Developmental Biology* **11:** 35-44.

Shinde, U., and Inouye, M. 1995. Folding pathway mediated by an intramolecular chaperone: characterisation of the structural changes in pro-subtilisin E coincident with autoprocessing. *J Mol Biol* **252:** 25-30.

Siezen, R., and Leunissen, J. 1997. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein science* **6:** 501-523.

Siezen, R., Vos, W., de Leunissen, J., and Dijkstra, B. 1991. Homology modelling and protein engineering strategy for subtilases, the family of subtilisin-like serine proteases. *Protein engineering* **4:** 719-737.

Sleat, D. E., Wiseman, J. A., El-Banna, M., Kim, K., H., Mao, Q., Price, S., Macauley, S., L., Sidman, R., L., Shen, M., M., Zhao, Q., Passini, M., A., Davidson, B., L., Stewart, G., R., Lobel, P. 2004. A mouse model of classical late-infantile neuronal ceroid lipofuscinosis based on targeted disruption of the CLN2 gene results in a loss of tripeptidyl-peptidase I activity and progressive neurodegeneration. *J Neurosci.* Oct 13;**24**(41):9117-26.

Squire, P., Moser, P., and O'Konski, C. 1968. The hydrodynamic properties of bovine serum albumin monor and dimer. *Biochemistry* **7:** 4261-4272.

Storer, A.C., and Menard, R. 1994. Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol* **244:** 486-500.

Strausberg, S.L., Alexander, P.A., Gallagher, D.T., Gilliland, G.L., Barnett, B.L., and Bryan, P.N. 1995. Directed evolution of a subtilisin with calcium-independent stability. *Biotechnology (N Y)* **13:** 669-673.

Strongin, A., Izotova, L., Abramov, Z., Gorodetsky, D., Ermakova, L., Baratova, L., Belyanova, L., and Stepanov, V. 1978. Intracellular serine protease of *Bacillus subtilis:* sequence homology with extracellular subtilisins. *J Bacteriol* **133:** 1401-1411.

Stroud, R.M., Kossiakoff, A.A., Chambers, J.L. 1977. Mechanisms of zymogen activation. *Annu Rev Biophys Bioeng.* **6**:177-93

Stryer. 1995. *Biochemistry*, 4th ed. W H Freeman and company, New York.

Studier, F.W., and Moffatt, B.A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189:** 113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185:** 60-89.

Stumpe, M.C., and Grubmuller, H. 2008. Polar or apolar--the role of polarity for urea-induced protein denaturation. *PLoS Comput Biol* **4:** e1000221.

Subbian, E., Yabuta, Y., and Shinde, U. 2004. Positive selection dictates the choice between kinetic and thermodynamic protein folding and stability in subtilases. *Biochemistry* **43:** 14348-14360.

Subbian, E., Yabuta, Y., and Shinde, U. 2005. Folding pathway mediated by an intramolecular chaperone: intrinsically unstructured propeptide modulates stochastic activation of subtilisin. *J Mol Biol* **347:** 367-383.

Takeuchi, Y., Tanaka, S., Matsumura, H., Koga, Y., Takano, K., Kanaya, S. 2009. Requirement of a unique Ca(2+)-binding loop for folding of Tk-subtilisin from a hyperthermophilic archaeon. *Biochemistry*. Nov 10;**48**(44):10637-43.

Tanaka, S., Saito, K., Chon, H., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. 2007. Crystal structure of unautoprocessed precursor of subtilisin from a hyperthermophilic archaeon: evidence for Ca2+-induced folding. *J Biol Chem* **282:** 8246-8255.

Terabe, M., Kojima, S., Taguchi, S., Momose, H., Miura K. 1994. Three novel subtilisin-trypsin inhibitors from Streptomyces: primary structures and inhibitory properties. *J Biochem.* **116**(5):1156-63.

Tindbaek, N., Svendsen, A., Oestergaard, P., and Draborg, H. 2004. Engineering a substrate-specific cold adapted subtilisin. *Protein engineering, design and selection* **2:** 149-156.

Tsuchiya, K., Ikeda, I., Tsuchiya, T., and Kimura, T. 1997. Cloning and expression of an intracellular alkaline protease gene from alkalophilic *Thermactinomyces* sp. HS682. *Biocsci. Biotech. Biochem* **61:** 298-303.

Vallet, V., Chraibi, A., Gaeggeler, H.P., Horisberger, J.D., and Rossier, B.C. 1997. An epithelial serine protease activates the amiloride-sensitive sodium channel. *Nature* **389:** 607-610.

Van Der Laan, J.C., Gerritse, G., Mulleners, L.J., Van Der Hoek, R.A., Quax, W.J. 1991. Cloning, characterization, and multiple chromosomal integration of a Bacillus alkaline protease gene. *Appl Environ Microbiol.* **57**(4):901-9

Van de Ven, W.J., Roebroek, A.J., and Van Duijnhoven, H.L. 1993. Structure and function of eukaryotic proprotein processing enzymes of the subtilisin family of serine proteases. *Crit Rev Oncog* **4:** 115-136.

Van Melderen, L., Aertsen, A. 2009. Regulation and quality control by Lon-dependent proteolysis. *Research in Microbiology.* **160**(9):645-651

Voordouw, G., Milo, C., and Roche, R.S. 1976. Role of bound calcium ions in thermostable, proteolytic enzymes. Separation of intrinsic and calcium ion contributions to the kinetic thermal stability. *Biochemistry* **15:** 3716-3724.

Walters, J., Milam, S.L and Clark, A.C. 2009. Chapter 1 Practical Approaches to Protein Folding and Assembly: Spectroscopic Strategies in Thermodynamics and Kinetics. *In Methods in Enzymology.* Volume 455. Biothermodynamics Part A. Ed Johnson, M.L., Holt, J. M and Ackers, G.K. *Pages 1-39.* Elsevier Inc.

Wang, W., and Kollman, A. 2000. Free energy calculations in dimer stability of the HIV protease using molecular dynamics and a continuum solvent mel. *J Mol Biol* **303:** 567-582.

Warshel, A., Naray-Szabo, G., Sussman, F., and Hwang, J.K. 1989. How do serine proteases really work? *Biochemistry* **28:** 3629-3637.

Wells, J.A., Cunningham, B.C., Graycar, T.P., Estell, D.A., and Carter, P. 1987. On the evolution of specificity and catalysis in subtilisin. *Cold Spring Harb Symp Quant Biol* **52:** 647-652.

Wells, J.A., and Estell, D.A. 1988. Subtilisin-an enzyme designed to be engineered. *Trends Biochem Sci* **13:** 291-297.

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

Williamson, MS., Forde, J., Buxton, B., Kreis, M. 1987. Nucleotide sequence of barley chymotrypsin inhibitor-2 (CI-2) and its expression in normal and high-lysine barley. *Eur J Biochem.* **165**(1):99-106.

Wright, C.S., Alden, R.A., Kraut, J. 1969. Structure of subtilisin BPN' at 2.5 angström resolution. *Nature*. **221**(5177):235-42.

Wong, S.L., and Doi, R.H. 1986. Determination of the signal peptidase cleavage site in the preprosubtilisin of *Bacillus subtilis*. *J Biol Chem* **261:** 10176-10181.

Xu, S., Qin, S., and Pan, X.M. 2004. Thermal and conformational stability of Ssh10b protein from archaeon *Sulfolobus shibattae*. *Biochem J.* 1; **382**(Pt 2): 433–440.

Yabuta, Y., Hiroshi, H., Inouye, M., and Shinde, U. 2001. Folding pathway mediated by an intrmolecular chaperone. *J Biol Chem* **276:** 44427-44434.

Yamagata, Y., and Ichishima, E. 1995. A new alkaline serine protease from alkalophilic *Bacillus sp.* cloning,sequencing and characterisation of an intracellular protease. *Current Microbiology* **30:** 357-366.

Yang, M.Y., Ferrari, E., and Henner, D.J. 1984. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. *J Bacteriol* **160:** 15-21.

Yepes, M., Roussel, B.D., Ali, C., and Vivien, D. 2009. Tissue-type plasminogen activator in the ischemic brain: more than a thrombolytic. *Trends Neurosci.* **32**(1):48-55

Zhu, X., Ohta, Y., Jordan, F., and Inouye, M. 1989. Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* **339:** 483-484.

Appendix 1: Equilibrium folding model – Two-state model **($N_2 \leftrightarrow 2U$)**

In the model below the denaturant induced changes to the protein were measured using either fluorescence or circular dichroism that were directly related to mole fraction at each state e.g. native or unfolded protein.

Within the model the y values are the amplitudes of the spectroscopic signal using CD, delta ellipticity at 222 nm; or fluorescence intensity. y_N and y_U represent the folded native and unfolded y values and y_N^0 represents the y value in the absence of denaturant whereas and y_U^0 represents the y value for the unfolded protein. f_U and f_N represent the fractions of unfolded and native protein. $[P]_{TOT}$ represents the total protein concentration of the monomer. [D] represents the denaturant concentration. R represents the molar gas constant (1.987 $calmol^{-1}K^{-1}$). T represents the absolute temperature.

Mechanism - $N_2 \leftrightarrow 2U$ Equation 1

Where K_c is the equilibrium constant for the reaction.

$K_c=[U]^2/[N_2]$ Equation 2

$[N_2]={}^1/_2[P]_{TOT}f_N$ Equation 3

$[U]^2=[P]_{TOT}^2 f_U^2$ Equation 4

Substitute equation 3 and 4

$K_c=2[P]_{TOT}f_U/f_N$ Equation 5

Rearrange for f_N

$f_N=2[P]_{TOT}f_U^2/K_c$ Equation 6

$f_N+f_U=1$ Equation 7

Substitute equation 6

$2[P]_{TOT}f_U^2/K_c+f_U=1$ Equation 8

Multiply by K_c

$2[P]_{TOT}f_U^2 + f_UK_c = K_c$ Equation 9

Quadratic

$2[P]_{TOT}f_U^2 + f_UK_c - K_c = 0$ Equation 10

Solve quadratic equation for f_U

$$fu = \frac{-K_c + \sqrt{K_c^2 + 8[P]_{TOT} K_c}}{4[P]_{TOT}}$$ Equation 11

The equilibrium constant (K_c) for the model is related to free energy (equation 12). Assuming that the free energy change is linearly dependent on denaturant concentration the free energy change in the absence of denaturant can also be calculated (equation 13) (Walters *et al.*, 2009).

Thus,

$\Delta G = -RT \ln K_c$ Equation 12

$\Delta G = \Delta G^{H_2O} - m[D]$ Equation 13

Combine 13 and 14 and solve for K_c,

$-RT \ln K_c = \Delta G^{H_2O} - m[D]$

$$\ln K_c = \frac{\Delta G^{H_2O} - m[D]}{RT}$$

$$K_c = \exp\left(\frac{\Delta G_D{}^{H_2O} - m\,[D]}{RT}\right)$$ Equation 14

Fitting equation

$y = y_N f_N + y_U f_U$ Equation 15

Where

$$y_N = y_N{}^0 + m_N[D]$$

$$y_U = y_U{}^0 + m_U[D]$$

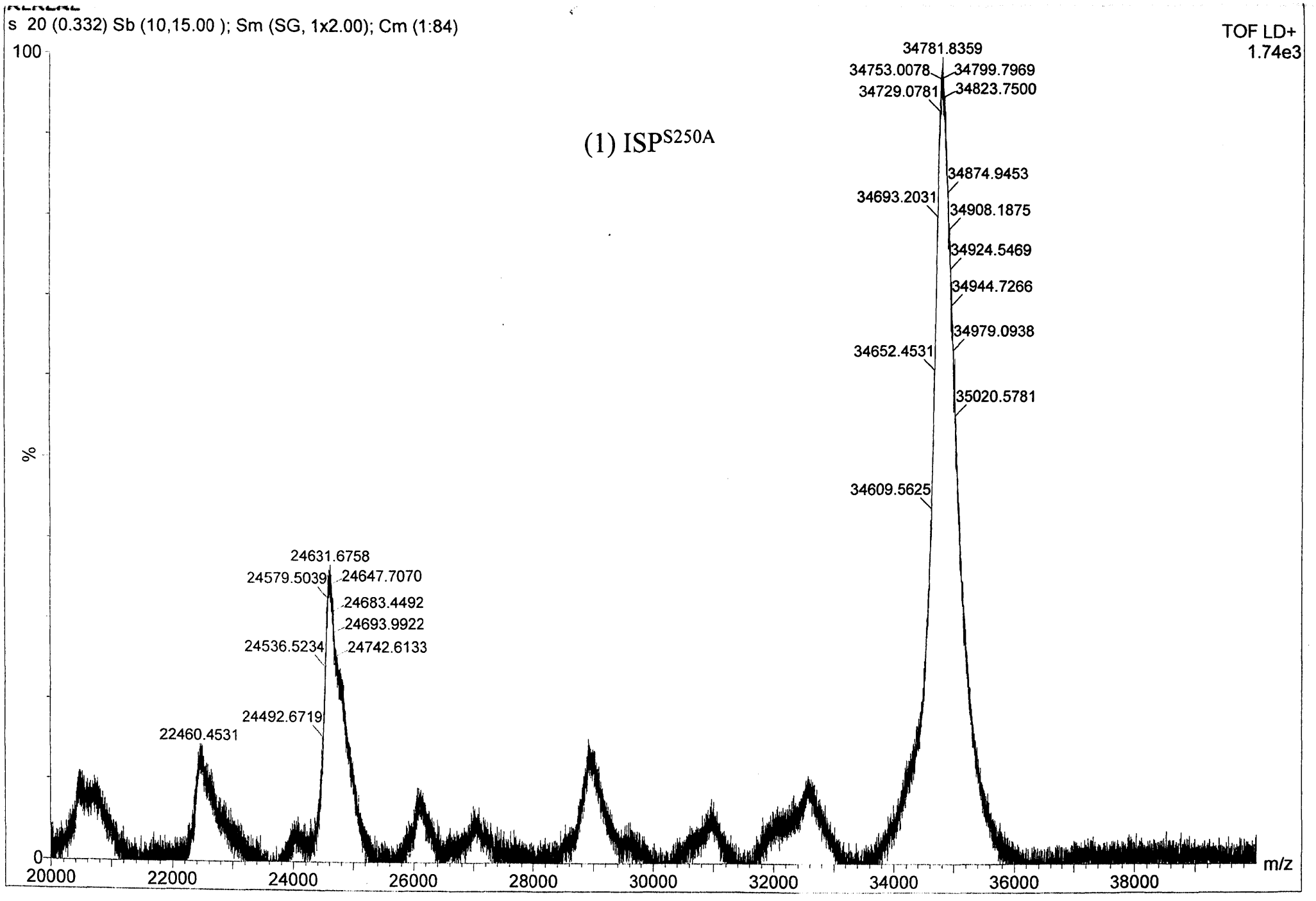
The values for ΔG^{H_2O} and m for each transition and $y_N{}^0$ and $y_U{}^0$ were allowed to vary as fitting parameters.

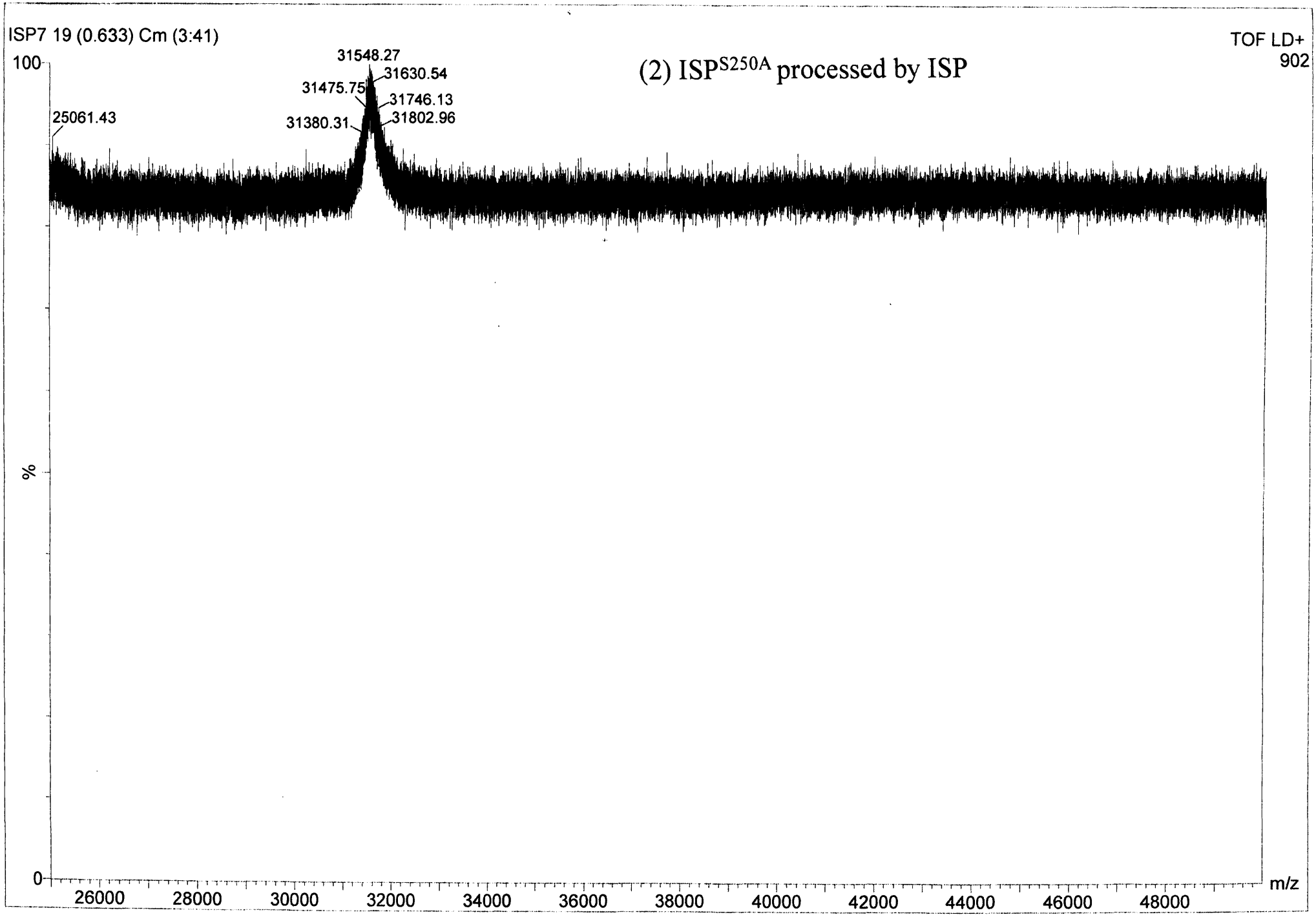
MALDI-TOF ANALYSIS (Table 5.1)

(1) ISP^{S250A}

(2) ISP^{S250A} processed by ISP

(3) Processed ISP


s 20 (0.332) Sb (10,15.00); Sm (SG, 1x2.00); Cm (1:84)
TOF LD+
1.74e3
100
%
0
(1) ISP^S250A
34781.8359
34753.0078
34799.7969
34729.0781
34823.7500
34874.9453
34693.2031
34908.1875
34924.5469
34944.7266
34979.0938
34652.4531
35020.5781
34609.5625
24631.6758
24579.5039
24647.7070
24683.4492
24693.9922
24536.5234
24742.6133
24492.6719
22460.4531
20000
22000
24000
26000
28000
30000
32000
34000
36000
38000
m/z



(2) ISP^{S250A} processed by ISP

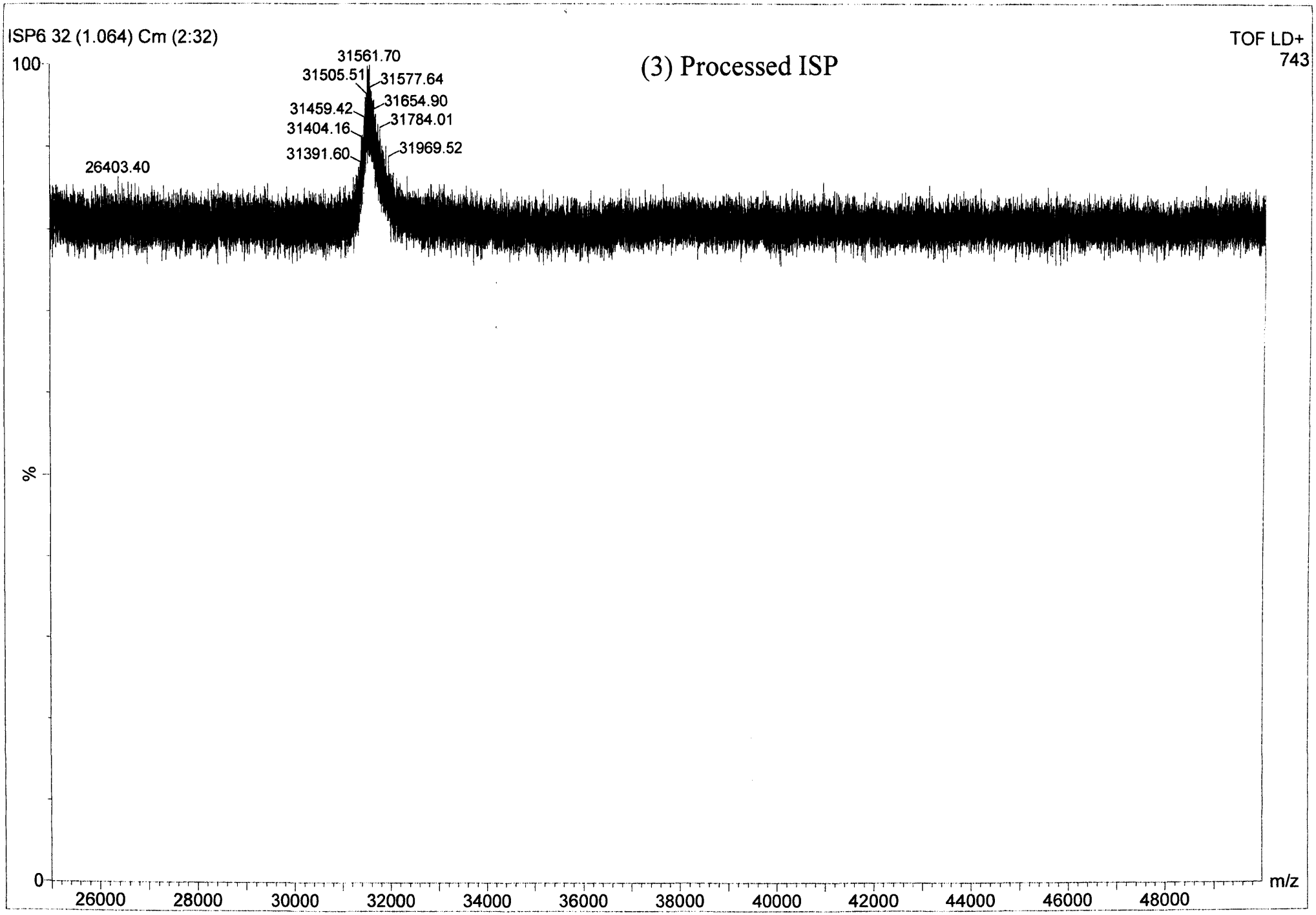

ISP6 32 (1.064) Cm (2:32)
TOF LD+
743
(3) Processed ISP
100
%
0
31561.70
31505.51
31577.64
31654.90
31459.42
31784.01
31404.16
31969.52
31391.60
26403.40
26000
28000
30000
32000
34000
36000
38000
40000
42000
44000
46000
48000
m/z

Table 1. Crystallographic statistics.

Data statistics	Crystal 1	Crystal 2
Beamline	ESRF ID14-1	ESRF ID14-1
Wavelength (Å)	0.9762	0.93
Space group	$P3_1$	$P3_1$
Cell parameters (Å)	a = 121.67, c = 106.23	a = 119.83, c = 106.20
Resolution range (Å)	60.8 – 2.68 (2.83-2.68)	47.3 – 2.45 (2.58-2.45)
Number of observations	198,401 (29,667)	238,937 (33,575)
Unique reflections	49,303 (7,217)	62,796 (9,185)
Completeness (%)	99.9 (100.0)	100.0 (99.9)
I/I(σ)	11.6 (2.0)	11.6 (1.2)
Average multiplicity	4.0 (4.1)	3.8 (3.6)
R_{merge} (%)[b]	14.4 (70.0)	18.4 (66.0)
Protomers in asymmetric unit	6	6
Refinement statistics:[c]		
Protein atoms	13285	13168
Waters	20	321
Sodium ions	6	6
PEG	-	-
Glycerol	-	-
Free reflections (%)	5	5
$R_{cryst}=\Sigma \lvert F_o-F_c \rvert / \Sigma F_o$(%)	16.7	18.8
Free R factor (%)	25.0	26.9
Twinning Fraction	h,k,l: 0.72 -k,-h,-l: 0.28	-
R.m.s. deviations from ideal geometry (target values in parentheses)		

Papers

JBC Presubmission

Novel mechanism for regulating the activity of an intracellular subtilisin

Michael Gamble[1], Jitka Vévodová [2], Antonio Ariza[2], Eleanor Dodson[2], Keith S. Wilson[2] and D. Dafydd Jones[1]

[1] School of Biosciences, Cardiff University, Cardiff CF10 3AT. [2] Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, UK;

Address correspondence: D. D. Jones, School of Biosciences, Main Building, Cardiff University, Cardiff, CF10 3AT, UK. Email: jonesdd@cf.ac.uk; K .S. Wilson. Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, UK Email: keith@ysbl.york.ac.uk

Structure Presubmission

Crystal structure of an intracellular subtilisin reveals a dimeric structure unique amongst the subtilisin family of proteases

Authors: Jitka Vévodová [1]#$, Michael Gamble[2]$, Antonio Ariza1, Eleanor Dodson1, D. Dafydd Jones[2]* and Keith S. Wilson[1]*.

[1] Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, UK

[2] School of Biosciences, Cardiff University, Cardiff CF10 3AT.

$ These authors contributed equally to the work.

* Corresponding authors. K .S. Wilson. Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, UK Email: keith@ysbl.york.ac.uk; D. D. Jones, School of Biosciences, Main Building, Cardiff University, Cardiff, CF10 3AT, UK. Email: jonesdd@cf.ac.uk.

Seminar

Structure, function and folding in a novel class of intracellular serine protease from *Bacillus clausii*

Molecular Cell Biology group, Cardiff University. 23 April 2007

Posters Year one and two internal assessments and at Postgraduate research days, Health Hospital, Cardiff University 2007 and 2008.