FEBS Open Bio 2 (2012) 209–215

journal homepage: www.elsevier.com/locate/febsopenbio

The role of substrate specificity and metal binding in defining the activity and structure of an intracellular subtilisin

Michael Gamble [a], Georg Künze [a,1], Andrea Brancale [b], Keith S. Wilson [c], D. Dafydd Jones [a,*]

[a] School of Biosciences, Cardiff University, Cardiff CF10 3AT, UK
[b] School of Pharmacy, Cardiff University, Cardiff CF10 3NB, UK
[c] Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK

ARTICLE INFO

Article history:
Received 28 June 2012
Accepted 2 July 2012

Keywords:
Protease
Subtilisin
Substrate specificity
Metal binding
Binding simulations

ABSTRACT

The dimeric intracellular subtilisin proteases (ISPs) found throughout Gram-positive bacteria are a structurally distinct class of the subtilase family. Unlike the vast majority of subtilisin-like proteases, the ISPs function exclusively within the cell, contributing the majority of observed cellular proteolytic activity. Given that they are active within the cell, little is known about substrate specificity and the role of stress signals such as divalent metal ions in modulating ISP function. We demonstrate that both play roles in defining the proteolytic activity of *Bacillus clausii* ISP and propose the molecular basis of their effects. Enzyme kinetics reveal that one particular synthetic tetrapeptide substrate, Phe-Ala-Ala-Phe-pNA, is hydrolysed with a catalytic efficiency ~100-fold higher than any other tested. Heat-denatured whole proteins were found to be better substrates for ISP than the native forms. Substrate binding simulations suggest that the S1, S2 and S4 sites form defined binding pockets. The deep S1 cavity and wide S4 site are fully occupied by the hydrophobic aromatic side-chains of Phe. Divalent metal ions, probably Ca^{2+}, are proposed to be important for ISP activity through structural changes. The presence of >0.01 mM EDTA inactivates ISP, with CD and SEC suggesting that the protein becomes less structured and potentially monomeric. Removal of Ca^{2+} at sites close to the dimer interface and the S1 pocket are thought to be responsible for the effect. These studies provide a new insight into the potential physiological function of ISPs, by reconciling substrate specificity and divalent metal binding to associate ISP with the unfolded protein response under stress conditions.

1. Introduction

Proteases that function exclusively within the cell play a role in many different biological processes in both prokaryotes and eukaryotes [1]. An essential requirement is stringent regulation of proteolytic activity to prevent untimely degradation of vital cellular components. In eukaryotic organisms, non-membrane associated proteases, especially serine proteases, are located within specific compartments other than the cytosol so physically separating them from a large proportion of the proteome [1]. Bacteria have no such compartments so regulation of proteases free in the cytosol is vital. The intracellular subtilisin proteases (ISPs) represent a class of protease that function within the cytosol of Gram-negative bacteria.

The ISPs belong to the broader subtilisin family of serine endopeptidases [2] that perform a variety of important biological roles ranging from non-specific digestion for nutrient scavenging [2] to virulence factors [3–5] to prohormone processing [6]. The bacilli extracellular subtilisins (ESPs), such as Subtilisins BPN', Carlsberg and E, are some of the most studied of all the proteases and have proved to be valuable models to understanding the protein structure–function relationship. They have a broad substrate specificity that suits their role as scavenging proteases [7,8] and have been utilised extensively for biotechnological applications, especially as active ingredients in laundry detergents [9,10]. The ESPs have a characteristic primary structure comprised of an N-terminal secretion signal sequence adjoining the prodomain chaperone required for the correct folding of the mature catalytic domain [11–13].

The intracellular subtilisins (ISPs) are relatives of the ESPs that function within the confines of the cell of bacilli and related bacte-

Abbreviations: ESP, extracellular subtilisin; ISP, intracellular subtilisin (processed and active); proISP, unprocessed pro-form of ISP; pNA, *para*-nitroanilide; $proISP^{S250A}$, proISP with the active site S250A mutation; SEC, size exclusion chromatography; Suc, succinyl

* Corresponding author. Address: School of Biosciences, Main Building, Cardiff University, Cardiff CF10 3AT, UK.
E-mail address: jonesdd@cf.ac.uk (D.D. Jones).
URL: http://www.cardiff.ac.uk/biosi/contactsandpeople/stafflist/i-l/jones-da-fydd-dr-overview_new.html (D.D. Jones).
[1] Current address: Institute of Medical Physics and Biophysics, University of Leipzig, Germany.

2211-5463/$36.00
http://dx.doi.org/10.1016/j.fob.2012.07.001

ria [2,14–16]. They are responsible for the majority of observed cellular protease activity in *Bacillus subtilis* [17]. The ISPs have a distinctive primary structure (Fig. 1A) to that of their ESP relatives [18,19]; they have no signal sequence in keeping with their intracellular location, and the 60–80 residue ESP prodomain is replaced by a short N-terminal extension.

Recently determined crystal structures of the ISP from *Bacillus clausii* [18,19] allowed us to propose roles for the unique sequence features, such as the N-terminal extension. The catalytic domain of ISP has a similar fold to other subtilisins. However, while the ESPs are monomeric, ISP is dimeric with a C-terminal helical arm playing a major role in defining the dimer interface (Fig. 1A). In the full-length precursor of ISP (henceforth termed proISP), the N-terminal extension acts as an inbuilt inhibitor of ISP activity that binds back over the active site and regulates activity through an original combined mechanism [18]. The conserved LIPY/F motif in the ISP N-terminal extension is vital to regulation as it binds at the active site and the proline introduces a bulge that shifts the target scissile peptide bond out of reach of the catalytic serine. The LIPY/F motif also contributes to disruption of the catalytic triad conformation by shifting the serine and histidine triad residues apart. The N-terminal extension is slowly self-processed at a defined position over an extended time to generate mature active protease, known simply as ISP from hereon in [18]. Pro-segment processing results in a significant rearrangement of the substrate binding cleft, notably at the S1 site. The ISP substrate binding cleft is deeper and narrower than the more open ESP cleft (Fig. 1B) suggesting that ISP may require a more defined substrate-enzyme fit. However, little is known about the substrate preference of ISPs once processed and active. A high affinity metal ion binding site conserved amongst the ESPs and normally occupied by calcium is present in ISPs but is occupied by sodium in all observed crystal structures of proISP and ISP (Fig. 1C) [18,19]. The ISP crystal structure also reveals that two additional divalent metal ions bind each ISP protomer. One site binds a calcium ion and appears to organise a loop that contributes to formation of the S1 substrate binding cleft (Fig. 1C). The other site, which lies close to the dimer interface (Fig. 1C), is occupied by a strontium ion (close calcium mimic). The positioning of these additional metal binding sites are unique to the ISPs. As the divalent metals are not fully coordinated by the protein, it is unlikely they play a general structural role similar to that of the sodium ion (and calcium in the ESPs). Thus, their importance to ISPs currently remains unclear.

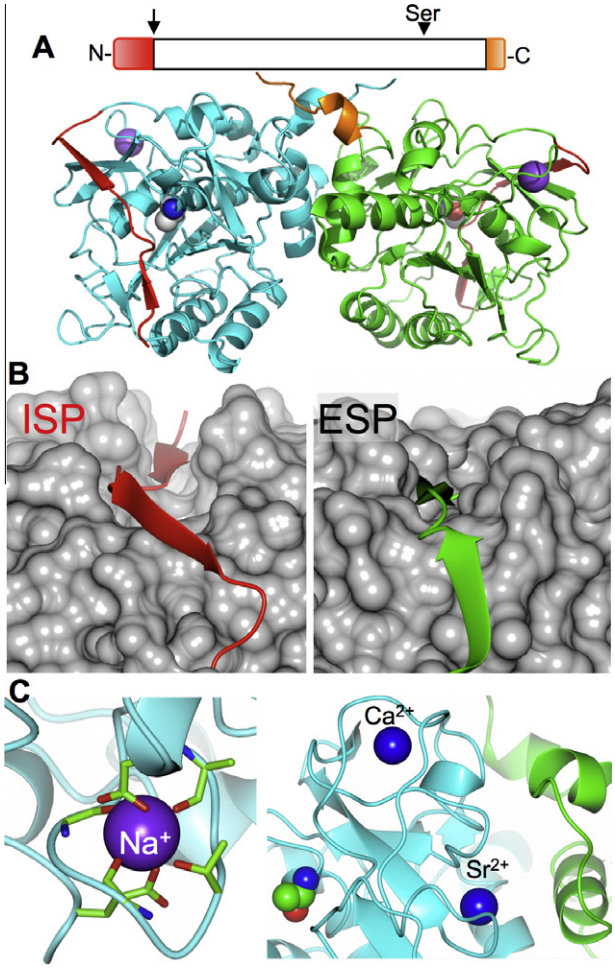

Fig. 1. Structure of ISP. (A) Schematic of primary and quaternary structure (PDB code 2x8j) with the N-terminal extension coloured red and the C-terminal dimerisation tail in one protomer coloured orange. Each protomer is coloured cyan and green. The Na^+ binding at the high affinity site is shown as a purple sphere and the active site alanine replacing the serine is shown as space-fill. (B) Substrate binding cleft of $proISP^{S250A}$ (left) and a representative ESP (BPN'; PDB code 1spb). The red ribbon represents the proISP N-terminal extension binding back across the active site. The green ribbon represents a section of the BPN' prodomain binding back across the active site. (C) Sodium (left panel) and calcium/strontium (right panel) binding sites in processed ISP (PDB code 2xrm). Each promoter is coloured as in A.

Here we investigate the substrate specificity of *B. clausii* ISP and the effect of divalent metal ions on ISP activity and structure. Enzyme kinetics using small tetrapeptide substrates revealed a strong preference for phenylalanine at the P1 and P4 positions. Molecular docking simulations confirmed that both the S1 and S4 sites provide a close fit for Phe. In the presence of low concentrations of the EDTA, enzyme activity was substantially reduced with accompanying effects on both the tertiary and quaternary structure of the enzyme implying a role for divalent metal ions in maintaining the protein in an active conformation.

2. Results

2.1. Enzyme kinetic characterisation of ISP

To investigate substrate specificity, ISP enzyme kinetics were measured using a variety of commercially available synthetic *para*-nitroanilide (pNA) linked tetrapeptide chromogenic substrates, many of which are known to act as effective substrates for ESPs. The substrate, Suc-P_4-P_3-P_2-P_1-pNA (where Suc is a succinyl group and P_n represents the individual amino acids according to the nomenclature of Schechter and Berger [20]) is abbreviated to the 4 core amino acid residues. The kinetic parameters are shown in Table 1. The order of ISP activity in terms of specificity constant (k_{cat}/K_M) is FAAF >> AAPF > AAPNle > AAPL ≈ AAPM ≈ AAVA ≈ AAPK. No activity was observed using the peptide substrates YVAD and AAPE, which have negatively charged residues at P_1. FAAF was by far the preferred substrate, with k_{cat}/K_M almost 100-fold higher than the next most efficient substrate, AAPF. The difference between the latter pair is largely due to K_M as the k_{cat} values are very similar (Table 1), indicating that substrate binding has been affected the greatest. The substrate specificity profile of ISP suggests it prefers largely hydrophobic substrates, especially with aromatics occupying P_1 and P_4. Using the peptides AAPF, AAPNle, AAPL, AAPM, AAPK and AAPE where P_2, P_3 and P_4 residues were constant (AAP), specificity at the P_1 position could be probed. The k_{cat} values were consistently low (0.2–0.7 s^{-1}), resulting in low k_{cat}/K_M values (0.8–1.4 mM s^{-1}). The exception was AAPF, where although a higher K_M was observed, the k_{cat} was substantially better (>15-fold) than the other substrates. This indicates that ISP prefers large hydrophobic residues at P_1, with Lys tolerated. The slightly higher K_M (~4-fold) for leucine at P_1 compared to its isomeric norleucine (Nle) form suggests that removing the branch point and extending the aliphatic side chain length by one carbon unit has marginally improved substrate recognition but the overall catalytic efficiencies are similar (Table 1). The pres-

Table 1
ISP Enzyme kinetics analysis.

Substrate Suc-P_4-P_3-P_2-P_1↓pNA	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
Phe-Ala-Ala-Phe-pNA	0.066 ±0.02	8 ±0.2	121
Ala-Ala-Pro-Phe-pNA	6.475 ±0.39	12 ±0.5	1.8
Ala-Ala-Pro-Nle-pNA	0.301 ±0.02	0.4 ±0.01	1.4
Ala-Ala-Pro-Leu-pNA	1.315 ±0.23	0.7 ±0.08	0.5
Ala-Ala-Pro-Met-pNA	0.960 ±0.25	0.5 ±0.05	0.5
Ala-Ala-Val-Ala-pNA	0.596 ±0.02	0.2 ±0.01	0.4
Ala-Ala-Pro-Lys-pNA	4.177 ±0.31	0.6 ±0.03	0.2
Ala-Ala-Pro-Glu-pNA	ND[a]	ND[a]	ND[a]
Tyr-Val-Ala-Asp-pNA	ND[a]	ND[a]	ND[a]

[a] ND, not determined due to no observed turnover.

ence of Pro at P_2 is unlikely to have had a major general effect, as AAVA with Val at P_2 is a poorer substrate for ISP than most of the AAPX-based substrates (Table 1).

2.2. Simulating substrate docking with ISP

To investigate further the molecular basis of substrate preference, combined docking and molecular dynamic simulations of FAAF (in the form of $F_4A_3A_2F_1$–A_1' with pNA replaced with an alanine) were performed. The structure of ISP (PDB 2xrm) was modified to incorporate the active site serine (residue 250), as the structure had been determined with this residue mutated to alanine [18,19]. Substrate docking was combined with molecular dynamic simulations so the ISP–substrate complex structure can sample various molecular interactions rather than rely simply on docking to a static crystal structure. $F_4A_3A_2F_1$–A_1' bound in the narrow substrate binding cleft (Fig. 2A) and positioned the target peptide bond ($F_4A_3A_2F_1$↓A_1') for hydrolysis by Ser250 (Fig. 2B). The solvation free energy gain on substrate binding to ISP as calculated by PISA was −8.1 kcal/mol with an interface area of 520 Å^2 comprised of 32 ISP residues. The S1 site forms a narrow, deep, largely hydrophobic cavity that neatly houses the Phe at P_1. The site is predominantly comprised of residues from two loops; the looped turn constituted by residues Ala180-Pro197 that binds Ca^{2+} (Fig. 1C) and the first part of the Met152-Glu162 loop. Key residues forming the cavity include Gly155-Pro157, Ala180, Gly182, Asn183 and Ala195. The P_2 residue (A_2) occupies a smaller more open site comprised of residues Leu118, Ser153 and Leu154 and lies close to catalytic triad residues Asp49 and His86. The P_3 (A_3) side chain points towards the solvent but is flanked on one side by residues from the loop comprised of residues Lys116 to Glu125. The P_4 Phe binds at the larger, more open S4 cavity predominantly via a single face of the benzyl side-chain (Fig. 2A). Hydrophobic residues centred on the termini of the two partially exposed antiparallel α-helices make up the S4 site, with residues Met126, Ile129, Leu154, Leu163 and Pro197 being the major contributors. These residues are located at the opposite end of the substrate cleft from that of the active site serine (Fig. 2A). Pro197 contributes to both the S1 and S4 sites. In addition to side chain interactions, there is a hydrogen bond network between the main chain of the substrate and ISP resulting in the P_2–P_4 residues ($F_4A_3A_2$) forming a complementary antiparallel β-strand with ISP residues 122–124 (Fig. 2B).

Comparison with the structure adopted by the N-terminal extension in proISP reveals that P_2–P_4 adopt a similar structure but that P_1, $P_{1'}$ and the proISP-specific P_0 residues differed significantly (Fig. 2C). The latter difference is to be expected as the P_1 and P_0 (proline) residues of the prosequence bulge out and away from the catalytic serine (indeed the S1 site only becomes properly structured on maturation from proISP to ISP), whereas the docked substrate binds in a position primed for catalysis. Phe occupies the S4 site in both proISP and substrate docked ISP, with the aromatic side chain taking up a similar position (Fig. 2C).

Substrate docking simulations with the next best substrate, AAPF (in the form of $A_4A_3P_2F_1$–A_1' with pNA replaced with an alanine) reveals it binds in a similar manner to FAAF with the backbone trajectories of the two peptides being almost identical (Fig. 2D). This suggests that the presence of Pro at the P_2 site has little effect on mode of binding. Indeed, Pro at P_2 in AAPF has a slightly more significant contribution to solvation energy (−1.3 kcal/mol) than the P_2 Ala in FAAF (−0.92 kcal/mol). Overall, both the solvation free energy gain (−6.6 kcal/mol) and interface area (425 Å^2) are lower for AAPF compared to FAAF. The most noticeable difference relates to the S4 binding pocket, which remains largely unoccupied by the methyl side chain of the P_4 alanine residue (Fig. 2E). This is reflected in solvation energy difference of nearly 2 kcal/mol between Phe (−2.5 kcal/mol) and Ala (−0.6 kcal/mol) at the P_4 position.

2.3. Proteolytic digestion of whole proteins by ISP

To characterise ISP's ability to utilise whole proteins as substrates, proteolytic activity towards three-off-the-shelf proteins, malate dehydrogenase (MDH), lactate dehydrogenase (LDH) and bovine serum albumin (BSA), was investigated. All three proteins in their native form were poor substrates for ISP suggesting a high degree of substrate specificity (Fig. 3). Temperature denaturation of the substrate proteins prior to incubation with ISP resulted in markedly increased digestion indicating substrate sequence motifs hidden in the native form are now accessible. Both heat denatured

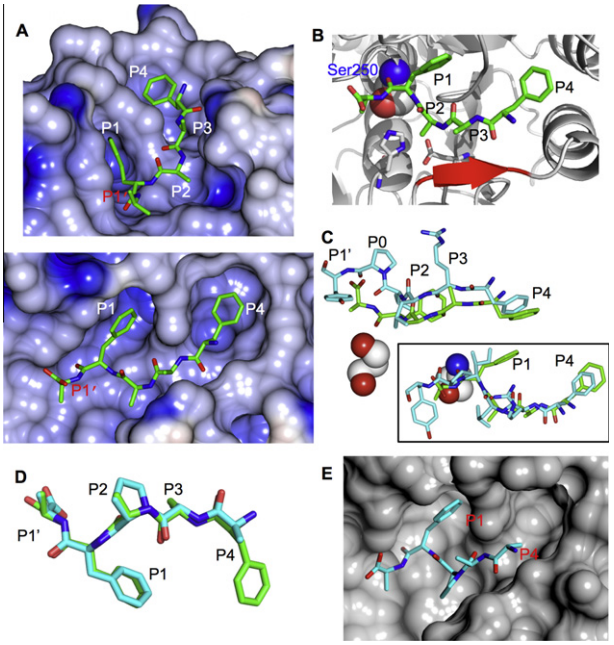

Fig. 2. Simulated substrate binding to ISP. (A) Electrostatic surface representation of ISP with the FAAF-A substrate shown as cylinders. The upper and lower panels represent two different views. (B) Ribbon representation of ISP with the catalytic serine shown as spheres, the Asp49 and His86 catalytic triad residues as grey cylinders and the $F^{P4}A^{P3}A^{P2}F^{P1}$–$A^{P1'}$ substrate as green cylinders. The β-strand formed antiparallel to the substrate is coloured red. (C) Overlay of modelled substrate (green) and pro-sequence (cyan; $F^{P4}R^{P3}L^{P2}I^{P1}P^{P0}$–$Y^{P1'}$) from proISP. The catalytic serine is shown as spheres. Structural alignment of residues equivalent to the mature form of the ISP was performed using CCP4mg [32]. (D) Overlay of the $F^{P4}A^{P3}A^{P2}F^{P1}$–$A^{P1'}$ (green) and $A^{P4}A^{P3}P^{P2}F^{P1}$-$A^{P1'}$ (cyan) substrates in the ISP-bound conformation. (E) Surface representation of ISP with $A^{P4}A^{P3}P^{P2}F^{P1}$-$A^{P1'}$ substrate shown as cylinders.

MDH and BSA were substantially digested, with little full length MDH left after 60 min incubation (Fig. 3). LDH however was still a poor substrate for ISP after temperature denaturation with no digestion products observed.

2.4. Influence of EDTA on ISP activity and structure

ISP activity and structure are sensitive to the presence of the divalent metal ion chelator EDTA, despite a monovalent Na^+ occupying the high affinity site normally occupied by Ca^{2+} in the ESPs. The activity of processed ISP was measured in the presence of $CaCl_2$ or EDTA using Suc-FAAF-pNA as the substrate. The addition of 1 mM $CaCl_2$ resulted in only a small increase in ISP activity, suggesting that the majority of available divalent metal ion sites were already occupied (Fig. 4A). ISP activity was unaffected by 0.01 mM EDTA but increasing the concentration to $\geq$0.1 mM dramatically decreased activity (Fig. 4A). Processed active ISP is relatively stable to self-proteolysis (Fig. 4B). On addition of EDTA, marked degradation was observed (Fig. 4B) indicating that ISP becomes more susceptible to self-proteolysis. A control using an inactive variant of unprocessed ISP with the catalytic serine mutated to an alanine (proISP^{S250A}) analysed under the same conditions was not digested (Fig. 4B) and excludes the possibility that contaminating factors may cause the degradation.

As metal ions are not involved in the subtilisin catalytic mechanism but are an important structural element, the effect of EDTA on ISP structure was investigated. There were significant differences in the CD spectra of both PMSF inhibited ISP and proISP^{S250A} on addition of EDTA suggesting changes to the protein conformation had occurred (Fig. 5A). The troughs at 208 nm and, to a lesser extent, 222 nm were less negative suggesting that part of the helical character of the protease is lost in the presence of EDTA. However, the CD spectra suggested that the protein still retained a substantial amount of structure.

Metal ion removal also had a small but significant effect on the SEC elution profile of proISP^{S250A} (Fig. 5B). Rather than the expected increase in hydrodynamic volume upon a protein becoming less structured, the hydrodynamic volume decreased. The elution volume in the presence of calcium was 14.9 ml but in the presence of EDTA proISP^{S250A} eluted slightly later (15.1 ml). The apparent molecular weight in the presence of EDTA was calculated to be 50 kDa (down from 56 kDa), which is slightly closer to that of the monomer than dimer (35 kDa versus 70 kDa). To confirm that the change in apparent molecular weight was due to a reversible structural change and not other factors such as proteolysis, the peak fraction was collected and exchanged into buffer containing 1 mM $CaCl_2$, and then reanalysed by SEC. The elution volume of the buffer exchanged proISP^{S250A} returned to 14.9 ml (Fig. 5B).

3. Discussion

Proteases active within the confines of the bacterial cell need to be tightly regulated to prevent unwanted or untimely degradation of vital cellular components. Mechanisms of regulation can occur at multiple levels, ranging from coarse control such as gene expression, post-translational processing or modification to much finer control through substrate specificity. The bacterial ISPs are the only known class of subtilisins that function within the bacterial cytosol. Apart from their functional location, they differ from their secreted relatives in terms of their sequence features and molecular structure. Our recent structures have provided a basis for ISP regulation through post-translational proteolytic processing [18,19]. These studies suggested that the N-terminal extension inhibited activity through an unusual and unexpected mixed mechanism with a significant non-competitive component [18]. However, little is known about additional factors such as substrate specificity once the protease is active and stress signals such as calcium can modulate the activity of ISP. Both these molecular features are key to defining potential physiological roles for the ISPs.

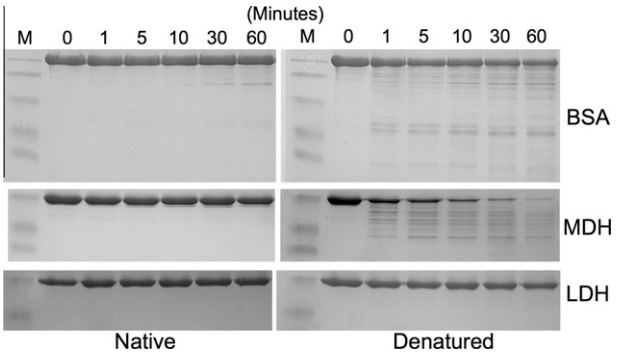

Fig. 3. Proteolytic digestion of protein substrates by ISP analysed by SDS–PAGE. Samples of ISP incubated with substrate were taken at the time points indicated. M represents molecular weight marker. The left and right hand gels show digestion patterns of the substrate before and after heat treatment, respectively.

3.1 Substrate specificity

The structure of *B. clausii* proISP^{S250A} in which the N-terminal extension binds back over the active site suggested a preference for hydrophobic amino acids as the S1, S2 and S4 pockets were

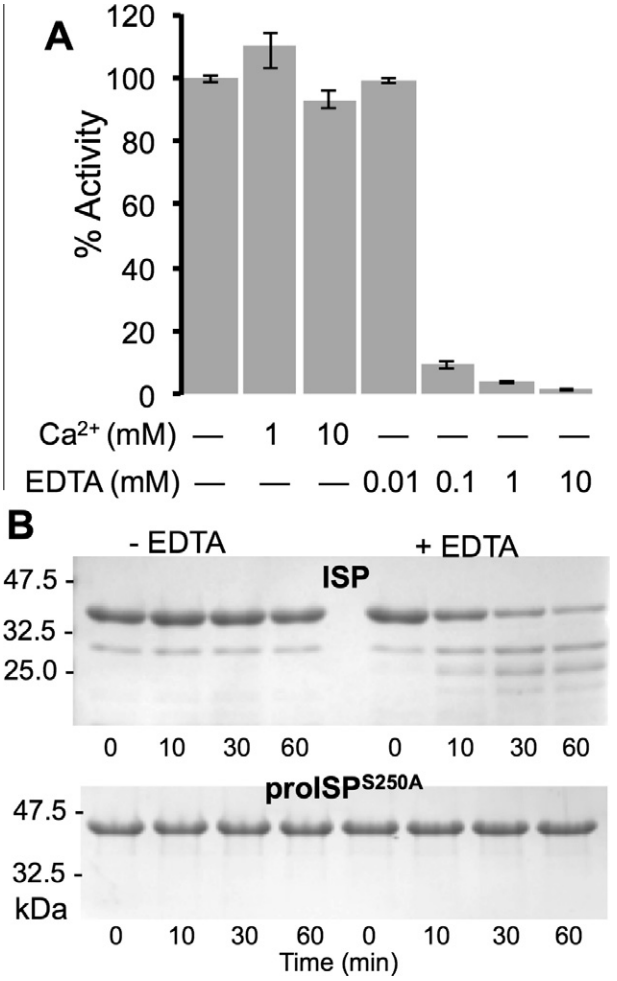

Fig. 4. Influence of EDTA on ISP function and degradation. (A) The dependence of ISP activity on the presence $CaCl_2$ and EDTA. The % activity was normalised to the activity of ISP in buffer without any added $CaCl_2$ or EDTA. (B) Stability to self-proteolytic digestion in the absence and presence of EDTA. Top panel represent active ISP and bottom panel inactive proISP^{S250A}.

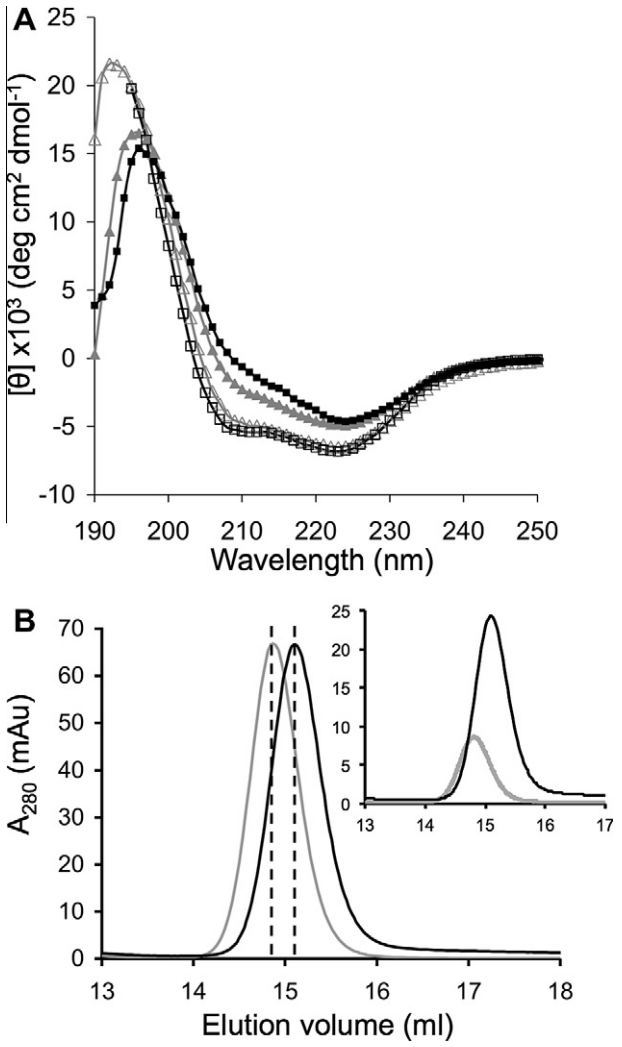

Fig. 5. Influence of EDTA on ISP structure. (A) Circular dichroism spectra of proISP^{S250A} (black line and squares) and the PMSF inhibited processed active ISP (grey line and triangles) at 25 °C in the presence of 1 mM $CaCl_2$ (open symbols) or 1 mM EDTA (closed symbols). (B) Elution profile at 25 °C of proISP^{S250A} in the presence of either 1 mM $CaCl_2$ (grey line) or 1 mM EDTA (black line) in the chromatography buffer. The vertical dashed lines represent the peak elution volume for each sample. Inset is the elution profile in the presence of EDTA (black) and the collection of these fraction and reapplication of the sample after the addition of $CaCl_2$.

occupied by Ile, Leu and Phe, respectively, with Leu and Ile part of the LIPY/F motif conserved amongst the intracellular subtilisins [19]. However, this is a catalytically inactive conformation and the S1 site in particular undergoes significant rearrangement on N-terminal processing [18]. A preference for hydrophobic amino acids is also implied by the ability of whole proteins to act as a substrate for ISP (Fig. 3). Native proteins were poor substrates for ISP indicating that any recognition motifs are not visible to ISP but increased proteolytic digestion was observed on heat denaturation when hydrophobic core residues are likely to become exposed.

The enzyme kinetics studies confirmed the importance of hydrophobic sequences in defining ISP substrate specificity as amongst the pNA tetrapeptides substrates tested, hydrophobic ones proved to be the best. In particular, the enzyme kinetics revealed that FAAF was by far the best substrate, with a catalytic efficiency nearly 100-fold higher than any other tested. This implies that large hydrophobic residues such as Phe at both P_1 and P_4 are especially important to substrate specificity, as other substrates with a hydrophobic P_1 residue and an alanine at P_4 are relatively poor in comparison (Table 1). Analysis of the crystal structure suggests that the S1 and S4 sites will be the prime determinants of substrate binding. Binding simulations with $F_4A_3A_2F_1–A_1'$ support this and outline the importance of Phe at the P_1 and P_4 positions in substrate recognition. Both the S1 and S4 binding pockets are compatible with Phe at P_1 and P_4 (Fig. 2A). This is especially notable at the S4 site where the main interaction with the substrate is via one face of the benzyl side-chain. The substrate docking simulations reveal that the S4 site remains largely unoccupied when P_4 is alanine (Fig. 2E). Preference for Phe at the S4 site is also evident from crystal structures of both proISP^{S250A} [19] and ISP^{S250A} [18]. In proISP^{S250A}, Phe4 from the inhibitory N-terminal extension occupies the S4 site (Fig. 2C). In ISP^{S250A}, a crystal artefact results in Phe315 from an adjacent dimer in the crystal unit cell occupying the S4 site. Thus the combined enzyme kinetics, binding simulations and crystal structure suggest that the ISPs have a strong preference for Phe at the P_4 position. Based on the structures of ISP and substrate binding simulations, the P_3 residue is unlikely to play a major role in substrate specificity. There is no apparent binding pocket for the P_3 residue and, in common with other subtilisins [2], the side chain points towards the solvent. The S2 pocket is less distinct and more exposed compared to S1 and S4. Structural data suggest that ISP can tolerate a wide variety of residues at P_2. Substrate docking simulations suggest both Pro and Ala at P_2 bind in a similar manner and from the available crystal structure data both Leu (Leu6 from proISP^{S250A}) and Glu (Glu317 from an adjacent dimer interaction with ISP^{S250A}) can occupy this S2 site. Thus, the S2 site may not be vital to defining substrate specificity compared to the S1 and S4 sites.

Substrates containing negatively charged amino acids at P_1 showed no measurable hydrolysis. Lys at P_1 was tolerated, even though the AAPK substrate was the poorest of those whose kinetic parameters could be determined. In the ISP^{S250A} structure, residues 315–318 from an adjacent subunit in the crystal occupy the P_4–P_1 sites, with Lys318 occupying the P_1 site [18]. This crystal artefact confirms that Lys has the potential to bind at S1, with the amine of the Lys interacting with the carboxyl of Glu184. This explains why acidic P_1 residues such as Asp or Glu cannot act as substrates. However, it also confirms that Lys is not preferred at P_1 as the majority of the hydrophobic S1 site remains unoccupied suggesting non-optimal binding.

Amongst the bacilli subtilisins, the ISPs appear to have a more defined substrate range than their secreted relatives. A more limited substrate range may be considered a perquisite when active within the cytosol, especially as expression of active ESP within the cell in is known to be lethal [21]. For example, the catalytic efficiencies of savinase from *B. lentus* towards FAAF, AAVA and AAPF are very similar [22] whereas there is up to 300-fold difference for ISP, with FAAF being by far the preferred substrate. Savinase also has significant activity towards both YVAD and AAVA [22], both poor substrates for ISP. BPN′ from *B. amyloliquefaciens* has a catalytic efficiency for AAPK 200-fold higher than for ISP [23]. Transplantation of ISP substrate binding regions into a savinase scaffold has a significant effect on activity towards all tested substrates, resulting in a strong preference for FAAF [24].

3.2 Divalent metal binding

It was generally thought calcium was important for ISP activity [16,21,25] and is supported in this study (Fig. 4) but the mechanism of action was unknown. Before the advent of detailed structural knowledge of the ISPs, it was presumed that Ca^{2+} occupied the conserved high affinity site (or A site); removal of Ca^{2+} from this high affinity site was thought to be the mechanism of action, through a general change in protein structure and stability. However, recently determined structures of *B. clausii* proISP^{S250A} and ISP^{S250A} reveal that Na^+ not Ca^{2+} binds at the high affinity site (Fig. 1C) [18,19]. Therefore, the ef-

fect of calcium on function must be exerted through some other mechanism. The structure of ISP^{S250A} provided evidence that Ca^{2+} bound at additional sites that could be important for the structure–function relationship (Fig. 1C): formation of the S1 binding site and close to the dimer interface [18]. In different crystal forms of proISP^{S250A}, an unidentifiable metal ion appeared to bind close to the dimer interface but its occupation with Ca^{2+} could not be defined due to the limited resolution of the analysis [19]. Furthermore, the S1 site of proISP^{S250A} was disordered with no metal bound. The role of a divalent metal ion, probably calcium, in promoting a catalytically competent ISP structure is strongly indicated here (Figs. 4 and 5). The presence of a divalent metal ion chelator knocked out activity (Fig. 4A), reduced proteolytic stability (Fig. 4B) and altered both the tertiary (Fig. 5A) and quaternary structure (Fig. 5B). Given the combined effects on activity and structure coupled with the evidence from structural studies that show ISP binds divalent metal ions [18], it is unlikely that EDTA is having a non-specific, metal-independent effect on activity. Rather, it is more likely that divalent metal ion removal by EDTA is influencing protein structure and thus function. The drop in apparent MW of proISP^{S250A} by 6 kDa in the presence of EDTA suggests that the protein has become a more compact dimer or a less structured monomer. CD spectra suggest that the protein is losing features of helical structure in the presence of EDTA, providing support for the latter. Indeed, helical regions including the important C-terminal arm (Fig. 1A) comprise the dimer interface. The divalent metal ions are only partially coordinated through the protein suggesting metal ion affinity is lower compared to sodium coordination at the high affinity site (Fig. 1C). This may explain why relatively low concentrations of EDTA (0.1 mM) are detrimental to ISP activity. Ca^{2+} dissociation from the high affinity site of ESPs is known to be very slow even in high concentrations of EDTA due to a large kinetic barrier [26] and is thus not compatible with our observations here. Therefore, Ca^{2+} removal by EDTA at the additional sites and not the predicted high affinity site is the most likely explanation for the observed changes in both structure and function.

3.3 Model for ISP regulation within the cell

We propose a model for regulation of ISP activity within the bacterial cell that integrates all the currently available information (Fig. 6). The change in role of divalent metals from a major structural determinant to a potential molecular switch ligand has physiological implications in terms of control over ISP activity, pre- and post-processing. Production of the nascent ISP under low intracellular [Ca^{2+}] traps the protein in a structurally compromised inactive monomeric state. Binding of sodium rather than calcium at the high affinity site may be critical as it could help maintain the protein in the inactive monomeric state without gross unfolding so allowing calcium to have more subtle yet important structural influence on activity. Increase in [Ca^{2+}] on influx into the cell converts the protease structure to an active conformation. Calcium has been proposed to act as a signalling molecule during cellular stress [27,28]. Under resting conditions, cellular *B. subtilis* [Ca^{2+}] is thought to be $\sim 10^{-7}$ M and increases in response to stress [27,29]. Increasing [Ca^{2+}] may result in the initial conversion of the partially structured monomeric to the dimeric proISP, through binding of calcium at the dimer interface. However, as the crystal structure of proISP has a disordered S1 site and no metal ion bound, there has to be a secondary calcium binding event either pre or post N-terminal extension processing. One route may be via a sparsely populated minor "open state" of proISP whereby binding of calcium at the S1 site may disturb binding of the N-terminal extension across the active site and lift inhibition of protease activity by the pro-sequence. Alternatively, transient dissociation of the N-terminal extension may facilitate calcium binding to the S1 site. The "open state" conformational change is stabilised by replacement of the water molecule that coordinates Na^+ at the high affinity site with a glutamate (Glu20), as observed in the processed ISP [18]. The small amount of active ISP will then instigate a cascade of proISP conversion to ISP; exponential activation of ISP through self processing has been observed previously [18]. Further calcium binding at the S1 site of the processed ISP reconstitutes the active form of the protease.

Once the precursor is processed to the manture enzyme and the substrate binding sites are fully formed, ISP can then fulfil its cellular function. Most soluble cytosolic proteins normally bury a large proportion of their hydrophobic residues on folding. Stress conditions can result in an increase in the amount of unfolded and non-functional proteins, which need to be dealt with. ISP is produced during the stationary phase [30], a time when the cell is under stress and nutrients becoming limiting. We speculate that the ISPs may play a role in the bacilli version of the unfolded protein stress response, specifically targeting signature motifs such as stretches of exposed large hydrophobic residues for digestion and recycling.

4. Materials and methods

Peptide substrates. Synthetic substrates succinyl-XXXX-*para*-nitroanilide (Suc-XXXX-pNA with XXXX representing either FAAF, AAPF, AAPNle (Nle – Norleucine), AAPL, AAPM, AAVA, AAPK, AAPE and YVAD) were purchased from Sigma–Aldrich (Poole, Dorset, UK) or Bachem (Bubendorf, Switzerland). Peptides were dissolved in either water or DMF before being diluted into buffer.

ISP enzyme kinetics. The various forms of ISP were purified by Ni^{2+} affinity chromatography as described previously [18,19]. The Michaelis–Menten kinetic parameters K_M, k_{cat} and k_{cat}/K_M were determined for each of the Suc-XXXX-pNA substrates. The concentrations of ISP used for each peptide were 0.02 μM (FAAF), 0.2 μM (AAPF) and 1 μM for the remaining peptides. Enzyme rates were measured in buffer containing 50 mM Tris–HCl pH 8.0, 0.5 M ammonium sulphate and 1 mM $CaCl_2$ at 25 °C. Enzymatic activity was determined by measuring the increase in absorbance at 405 nm, due to hydrolysis of the amino acid–pNA peptide bond, using a Varian Cary 50 Bio uv/vis spectrophotometer. The initial rates measured as A_{405}/min were converted into velocity (μM^{-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 the GraphPadPrism software.

4.1 Substrate docking simulations

All molecular modelling studies were performed on a MacPro dual 2.66 GHz Xeon running Ubuntu 10.04. The ISP structure used was PDB code 2XRM, with Ala250 mutated *in silico* back to Ser. Hydrogen atoms were added to the protein using the protonat 3D function of Molecular Operating Environment (MOE) (www.chemcomp.com). The peptide structure was built using MOE and minimized using the AMBER99 force field until a rmsd gradient of 0.05 kcal mol^{-1} $Å^{-1}$ was reached. The docking simulations were performed using FlexX (www.biosolveit.de). The best scored conformation that presented the scissile bond between F_1 and A_1' in proximity of Ser250 was kept for further studies. Molecular dynamics was then performed with Gromacs 4.5 [31] on the protease/peptide complex. The structure was solvated in a periodic octahedron simulation, providing a minimum of 9.0 Å of water between the protein surface and any periodic box edge and appropriately neutralised. The system was minimized and a position restrained molecular dynamic run was carried out for 150 ps, at 300 K in a NTP environment, using the GROMOS96 force field. Finally, a 5 ns production simulation was conducted, using the same conditions.

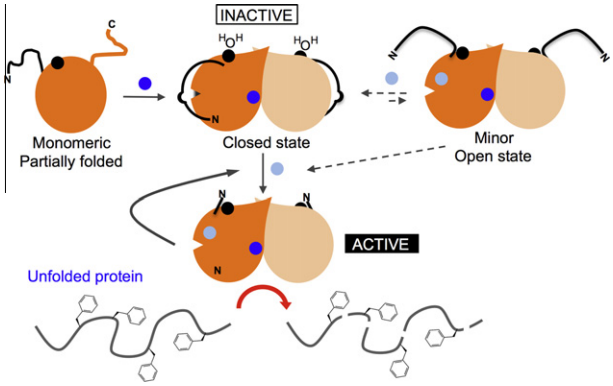

Fig. 6. Proposed model for the regulation of ISP activity. Initially, proISP (orange) is a partially folded, inactive monomer with the high affinity metal site occupied with sodium (black circle). On binding a divalent metal ion (probably calcium; dark blue circle) proISP converts to a folded dimeric protein with the N-terminal extension (black curved line) binding over and thus blocking the active site. The bulge represents the "proline bulge" that shifts the scissile bond beyond the reach of the catalytic serine. Posttranslational cleavage of the 18 residue pro-sequence by ISP activates the protease, with a second calcium ion (light blue) helping to form the S1 binding pocket. In order for ISP to activate itself, it is proposed that a small population of proISP adopts an "open" conformation in which the pro-sequence no longer binds across the active site. Formation of the open conformation may occur with the aid of calcium binding at the S1 site and Glu20 replacing water as the final ligand to bind sodium at the high affinity site. The small population of active ISP then initiates further processing and activation proISP via an intermolecular mechanism leading to a cascade of ISP activation. The active ISP then proceeds to digest protein substrates with exposed hydrophobics, such as phenylalanine.

Proteolysis of protein substrates. Digestion of bovine serum albumin (BSA), malate dehydrogenase (MDH) and lactate dehydrogenase (LDH) (all from Sigma–Aldrich) as substrates was performed by the addition of 2 μM ISP (2 μM) to 38 μg of each substrate in 50 mM Tris–HCl pH 8.0, 1 mM calcium chloride and 0.5 M ammonium sulphate. The substrate protein digestion was analysed before and after heat treatment at 95 °C for 5 min. Aliquots were taken at time intervals of 0, 1, 5, 10, 30 and 60 min and digestion analysed by reducing 12.5% (w/v) SDS–PAGE.

Effect of EDTA, $CaCl_2$ and urea. The calcium dependency of ISP proteolytic activity was measured as above but with the buffer (50 mM Tris–HCl pH 8.0, 0.5 M ammonium sulphate) supplemented with either $CaCl_2$ (1 mM) or ethylenediaminetetraacetic acid (EDTA) (0.01 mM, 0.1 mM or 1 mM). Rate of Suc-FAAF-pNA (50 μM) hydrolysis was measured as described above using 0.1 μM of ISP. To measure EDTA dependent self-proteolytic stability, ISP or $proISP^{S250A}$ was incubated in the presence or absence of 50 mM EDTA and aliquots taken at time intervals of 0, 10, 30 and 60 min and analysed using reducing 12.5% (w/v) SDS–PAGE. Circular dichroism spectroscopy and size exclusion chromatography (using a Superdex 200 10/300 GL column) were performed as described previously [18,19] with the solutions supplemented with concentrations of $CaCl_2$ or EDTA as outlined in the text. For the denaturation studies, the buffers were supplemented with the appropriate amount of urea indicated in the main text and samples left to equilibrate for at least 5 h.

Acknowledgements

M.G., G.K., and D.D.J. thank Cardiff University for financial support. M.G. was funded by a Cardiff School of Biosciences studentship. G.K. was an Erasmus exchange student at Cardiff University. We also thank Novozymes A/S for continuing support of K.S.W.

References

[1] Bond, JS and Butler, PE (1987) Intracellular proteases. Annu. Rev. Biochem. 56, 333–364.
[2] Siezen, RJ and Leunissen, JA (1997) Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci. 6, 501–523.
[3] Paton, AW, Beddoe, T, Thorpe, CM, Whisstock, JC, Wilce, MC, Rossjohn, J, Talbot, UM and Paton, JC (2006) AB5 subtilase cytotoxin inactivates the endoplasmic reticulum chaperone BiP. Nature 443, 548–552.
[4] Withers-Martinez, C, Jean, L and Blackman, MJ (2004) Subtilisin-like proteases of the malaria parasite. Mol. Microbiol. 53, 55–63.
[5] Wong, W, Wijeyewickrema, LC, Kennan, RM, Reeve, SB, Steer, DL, Reboul, C, Smith, AI, Pike, RN, Rood, JI, Whisstock, JC, et al. (2011) S1 pocket of a bacterially derived subtilisin-like protease underpins effective tissue destruction. J. Biol. Chem. 286, 42180–42187.
[6] Rockwell, NC and Thorner, JW (2004) The kindest cuts of all: crystal structures of Kex2 and furin reveal secrets of precursor processing. Trends Biochem. Sci. 29, 80–87.
[7] Bryan, PN (2000) Protein engineering of subtilisin. Biochim. Biophys. Acta 1543, 203–222.
[8] Wells, JA and Estell, DA (1988) Subtilisin–an enzyme designed to be engineered. Trends Biochem. Sci. 13, 291–297.
[9] Gupta, R, Beg, QK and Lorenz, P (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. Appl. Microbiol. Biotechnol. 59, 15–32.
[10] Maurer, KH (2004) Detergent proteases. Curr. Opin. Biotech. 15, 330–334.
[11] Bryan, PN (2002) Prodomains and protein folding catalysis. Chem. Rev 102, 4805–4816.
[12] Eder, J and Fersht, AR (1995) Pro-sequence-assisted protein folding. Mol. Microbiol. 16, 609–614.
[13] Shinde, U and Inouye, M (2000) Intramolecular chaperones: polypeptide extensions that modulate protein folding. Semin. Cell Dev. Biol. 11, 35–44.
[14] Strongin, AY, Abramov, ZT, Yaroslavtseva, NG, Baratova, LA, Shaginyan, KA, Belyanova, LP and Stepanov, VM (1979) Direct comparison of the subtilisin-like intracellular protease of Bacillus licheniformis with the homologous enzymes of *Bacillus subtilis*. J. Bacteriol. 137, 1017–1019.
[15] Strongin, AY, Gorodetsky, DI, Kuznetsova, IA, Yanonis, VV, Abramov, ZT, Belyanova, LP, Baratova, LA and Stepanov, VM (1979) Intracellular serine proteinase of *Bacillus subtilis* strain Marburg 168. Comparison with the homologous enzyme from *Bacillus subtilis* strain A-50. Biochem. J. 179, 333–339.
[16] Strongin, AY, Izotova, LS, Abramov, ZT, Gorodetsky, DI, Ermakova, LM, Baratova, LA, Belyanova, LP and Stepanov, VM (1978) Intracellular serine protease of *Bacillus subtilis*: sequence homology with extracellular subtilisins. J. Bacteriol. 133, 1401–1411.
[17] Burnett, TJ, Shankweiler, GW and Hageman, JH (1986) Activation of intracellular serine proteinase in *Bacillus subtilis* cells during sporulation. J Bacteriol 165, 139–145.
[18] Gamble, M, Kunze, G, Dodson, EJ, Wilson, KS and Jones, DD (2011) Regulation of an intracellular subtilisin protease activity by a short propeptide sequence through an original combined dual mechanism. Proc. Natl. Acad. Sci. USA 108, 3536–3541.
[19] Vevodova, J, Gamble, M, Kunze, G, Ariza, A, Dodson, E, Jones, DD and Wilson, KS (2010) Crystal structure of an intracellular subtilisin reveals novel structural features unique to this subtilisin family. Structure 18, 744–755.
[20] Schechter, I and Berger, A (1968) On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. Biochem. Biophys. Res. Commun. 32, 898–902.
[21] 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.
[22] Tindbaek, N, Svendsen, A, Oestergaard, PR and Draborg, H (2004) Engineering a substrate-specific cold-adapted subtilisin. Protein Eng. Des. Sel. 17, 149–156.
[23] Wells, JA, Cunningham, BC, Graycar, TP and Estell, DA (1987) Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering. Proc. Natl. Acad. Sci. USA 84, 5167–5171.
[24] Jones, DD (2011) Recombining Low Homology, Functionally Rich Regions of Bacterial Subtilisins by Combinatorial Fragment Exchange. PLoS One 6, e24319.
[25] Tsuchiya, K, Ikeda, I, Tsuchiya, T and Kimura, T (1997) Cloning and expression of an intracellular alkaline protease gene from alkalophilic Thermoactinomyces sp. HS682. Biosci. Biotechnol. Biochem. 61, 298–303.
[26] Bryan, P, Alexander, P, Strausberg, S, Schwarz, F, Lan, W, Gilliland, G and Gallagher, DT (1992) Energetics of folding subtilisin BPN'. Biochemistry 31, 4937–4945.
[27] Herbaud, ML, Guiseppi, A, Denizot, F, Haiech, J and Kilhoffer, MC (1998) Calcium signalling in *Bacillus subtilis*. Biochim. Biophys. Acta 1448, 212–226.
[28] Norris, V, Grant, S, Freestone, P, Canvin, J, Sheikh, FN, Toth, I, Trinei, M, Modha, K and Norman, RI (1996) Calcium signalling in bacteria. J. Bacteriol. 178, 3677–3682.
[29] Dominguez, DC (2004) Calcium signalling in bacteria. Mol. Microbiol. 54, 291–297.
[30] Sheehan, SM and Switzer, RL (1990) Intracellular serine protease 1 of *Bacillus subtilis* is formed in vivo as an unprocessed, active protease in stationary cells. J. Bacteriol. 172, 473–476.
[31] Hess, B, Kutzner, C, Van Der Spoel, D and Lindahl, E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. 4, 435–447.
[32] Potterton, L, McNicholas, S, Krissinel, E, Gruber, J, Cowtan, K, Emsley, P, Murshudov, GN, Cohen, S, Perrakis, A and Noble, M (2004) Developments in the CCP4 molecular-graphics project. Acta Crystallogr. D Biol. Crystallogr. 60, 2288–2294.