

# Regulation of an intracellular subtilisin protease activity by a short propeptide sequence through an original combined dual mechanism

Michael Gamble<sup>a,1</sup>, Georg Künze<sup>a,1,2</sup>, Eleanor J. Dodson<sup>b</sup>, Keith S. Wilson<sup>b,3</sup>, and D. Dafydd Jones<sup>a,3</sup>

<sup>a</sup>School of Biosciences, Cardiff University, Cardiff CF10 3AT, United Kingdom; and <sup>b</sup>Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

Edited by Robert Huber, Max Planck Institute for Biochemistry, Planegg-Martinsried, Germany, and approved January 4, 2011 (received for review September 23, 2010)

**A distinct class of the biologically important subtilisin family of serine proteases functions exclusively within the cell and forms a major component of the *bacilli* degradome. However, the mode and mechanism of posttranslational regulation of intracellular protease activity are unknown. Here we describe the role played by a short N-terminal extension prosequence novel amongst the subtilisins that regulates intracellular subtilisin protease (ISP) activity through two distinct modes: active site blocking and catalytic triad rearrangement. The full-length proenzyme (proISP) is inactive until specific proteolytic processing removes the first 18 amino acids that comprise the N-terminal extension, with processing appearing to be performed by ISP itself. A synthetic peptide corresponding to the N-terminal extension behaves as a mixed non-competitive inhibitor of active ISP with a  $K_i$  of 1  $\mu$ M. The structure of the processed form has been determined at 2.6 Å resolution and compared with that of the full-length protein, in which the N-terminal extension binds back over the active site. Unique to ISP, a conserved proline introduces a backbone kink that shifts the scissile bond beyond reach of the catalytic serine and in addition the catalytic triad is disrupted. In the processed form, access to the active site is unlocked by removal of the N-terminal extension and the catalytic triad rearranges to a functional conformation. These studies provide a new molecular insight concerning the mechanisms by which subtilisins and protease activity as a whole, especially within the confines of a cell, can be regulated.**

posttranslational modification | protease regulation |  
proprotein activation | protease structure

Subtilisins are serine endopeptidases ubiquitous in nature, spread across the eubacteria, archaeobacteria, eukaryotes, and viruses (1). They play a variety of important biological roles that include specific posttranslational processing of hormones (2, 3), virulence and infection factors in various pathogens (4, 5), and as nonspecific digestive proteases (1). The distinctive primary structure features common to the majority of subtilisins, and typified by the bacilli extracellular subtilisins (ESP), comprise a N-terminal signal sequence with an adjacent prodomain required for correct folding of the mature, catalytic domain (6–8) (Fig. 1A). Both the signal sequence and the prodomain are posttranslationally removed, the latter autocatalytically. The bacilli ESPs have a broad substrate specificity that suits their role as scavenging proteases, and their robustness to harsh environments coupled with extensive protein engineering has resulted in their exploitation by industry for a variety of applications, in particular as an active ingredient in laundry detergents (6, 7, 9, 10). The ESPs have proved excellent paradigms for understanding key molecular features of proteins, including enzyme catalysis (11) and protein folding (6–8).

The intracellular subtilisins (ISPs) are a distinctive class found in many different bacilli and related bacteria (1, 12–17). They are the main component of the bacilli degradome, believed to account for approximately 80% of *Bacillus subtilis* intracellular

proteasome activity (18, 19). However, little is known regarding the crucial feature of how their activity is regulated posttranslationally within the cell, where control of protease activity is vital to prevent the untimely breakdown of crucial cellular protein components. This is exemplified by the harmful effects of intracellular expression of bacilli ESPs to the host cell (20).

The ISPs are close relatives of the bacilli ESPs, with 40–50% sequence identity (21). Despite this, their sequences have a number of distinctive features (Fig. 1A and B) that translate into tertiary and quaternary structural features unique amongst the subtilisins, including being dimeric (Fig. 1C) (21). Whereas a dimeric structure has also been observed for the multidomain tomato subtilase (22), dimerization occurs via a separate PA domain to the catalytic domain. In contrast, the catalytic domain together with a short C-terminal arm mediate dimerization in ISP (21). A key feature of the ISPs is the replacement of the classical “foldase” prodomain with a short (16–25 residue) N-terminal extension (Fig. 1A), which has no sequence homology with the ESP prodomains (Fig. 1B) and is not involved in the folding process (20). The LIPY/F sequence motif in the extension is strictly conserved (Fig. 1B), hinting at a common role. Posttranslational removal of the proISP extension has been reported in a number of organisms (13, 15, 16, 23) but the implications of this event remained unknown until the structure of the full-length ISP from *Bacillus clausii* was determined in our laboratories (21). The structure suggests that the extension acts as an inbuilt inhibitor of activity by binding back over and so blocking the active site. The LIPY/F motif plays a key role with the proline introducing a bulge that shifts the scissile peptide bond beyond the reach of the catalytic serine, preventing direct binding to the active site (Fig. 1D).

Here we describe the role of precise proteolytic processing of the N-terminal extension in regulating the activity of the *B. clausii* ISP. Comparison of the recently determined structure of the full-length protein (proISP<sup>S250A</sup>) with the processed form (N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup>) determined here reveals the original, dual approach by which this extension inhibits protease activity.

Author contributions: M.G., K.S.W., and D.D.J. designed research; M.G., G.K., and E.J.D. performed research; M.G., G.K., E.J.D., K.S.W., and D.D.J. analyzed data; and M.G., G.K., E.J.D., K.S.W., and D.D.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

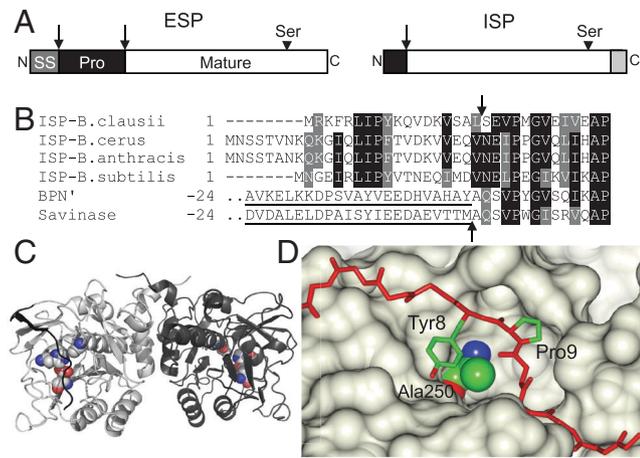
Data deposition: The crystallography, atomic coordinates, and structure factors have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID code 2xrm).

<sup>1</sup>M.G. and G.K. contributed equally to the work.

<sup>2</sup>Present address: Institute of Medical Physics and Biophysics, University of Leipzig, Leipzig D-04107, Germany.

<sup>3</sup>To whom correspondence may be addressed. E-mail: [jonesdd@cf.ac.uk](mailto:jonesdd@cf.ac.uk) or [keith@ysbl.york.ac.uk](mailto:keith@ysbl.york.ac.uk).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014229108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014229108/-DCSupplemental).



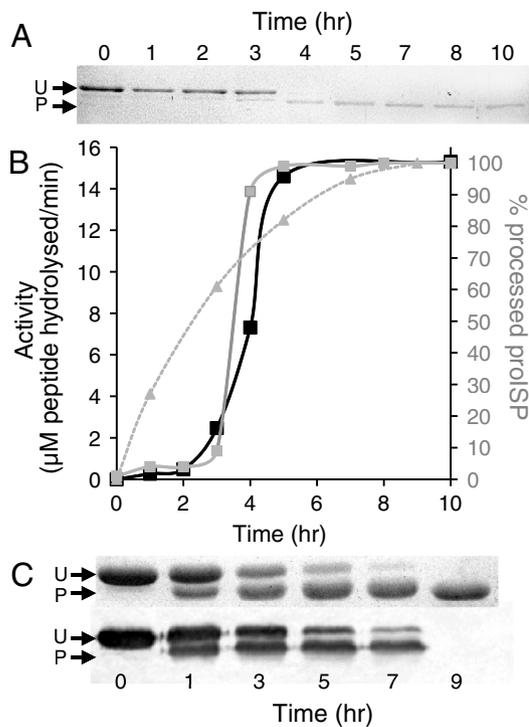
**Fig. 1.** Sequence and structural features of ESPs and ISPs. (A) General primary structure of ESP and ISP. SS and Pro refer to the signal sequence and prodomain region, respectively. Black and gray boxes represent the N- and C-terminal extensions, respectively, of the ISPs. Arrows indicate proteolytic cleavage positions. Approximate positions of the catalytic serine residues are indicated by triangles. (B) Partial sequence alignment of the N-terminal region of ISPs from various *Bacillus* species and two representative ESPs (BPN' from *Bacillus amyloliquifaciens* and Savinase from *Bacillus lentus*). A full sequence alignment has been published previously (21). Residues highlighted in black and gray are conserved and semiconserved, respectively. Residues underlined form part of the ESP prodomain sequence. Standard residue numbering based on mature BPN' is used for the ESPs. Arrows indicates the determined proteolytic cleavage position for the ISP from *B. clausii* that is part of this study and the two ESPs. (C) Recently determined structure of the dimeric ISP from *B. clausii* (ref. 21; PDB code 2WV7). Each monomer is shown in different shades of gray and the catalytic triad as space fill. (D) Binding mode of N-terminal extension. Close up of active site, with the core ISP structure shown in surface representation and the N-terminal extension shown in stick representation. The positions of residues Pro8, Tyr9, and Ala250 are indicated. Molecular structure diagrams were generated using CCP4mg (42).

**Results**

**Proteolytic Processing Regulates ISP Activity.** SDS-PAGE revealed proISP was processed to a smaller product after an initial lag phase in a time dependent manner (Fig. 2A and B). The N-terminal sequence of the larger species was (M)-R-K-F-R, identical to that expected for the full-length proISP with partial processing of the N-terminal methionine. For the smaller stable product it was S-E-V-P-M, corresponding to removal of the first 18 residues from the N-terminus of proISP (Fig. 1C). The measured molecular weight of the processed ISP (termed  $\Delta^{18}$ -ISP), as determined by mass spectrometry, was 31,548 Da. This is close to the theoretical molecular weight (31,554 Da) after the removal of the N-terminal extension (residues 1–18) and the unstructured C-terminal affinity tag (21) together with the lysine (Lys321) naturally resident at the ISP C-terminus (KLEHHHHHH). The excised C-terminal section is known to be unstructured and distant from the active site (21) so its removal is unlikely to influence the function or structure of the ISP.

Proteolytic processing resulted in activation of ISP (Fig. 2B). ProISP exhibited no activity toward the chromogenic tetrapeptide substrate Suc-Phe-Ala-Ala-Phe-pNA but after an initial lag phase of about 3 hr, activity rapidly increased to a maximum after 5 hr, which mirrored exponential processing of proISP. The final processed protein was stable with no loss of activity or further degradation even after 10 hr.

To investigate if ISP itself was responsible for processing, an inactive variant, proISP<sup>S250A</sup>, with the active serine mutated to alanine (21), was used as substrate for  $\Delta^{18}$ -ISP. ProISP<sup>S250A</sup> was initially cleaved at a single position (Fig. 2C), with the N-terminal sequence for the digested form matching that of



**Fig. 2.** Role of proteolytic processing in regulating ISP activity. (A) Time course of proISP processing. (B) Change in ISP activity (black squares; left y axis) measured by monitoring the hydrolysis of Suc-Phe-Ala-Ala-Phe-pNA compared with proISP processing (gray squares; right y axis) as determined by densitometry of the bands shown in (A) over time. The rate of proISP<sup>S250A</sup> processing by  $\Delta^{18}$ -ISP (triangle) determined by densitometry of the bands in (C) is also shown. (C) Time course of proISP<sup>S250A</sup> processing in the presence of active  $\Delta^{18}$ -ISP. Top panel, SDS-PAGE. (Lower) Western blot. U and P refer to the unprocessed and processed forms, respectively.

the processed wt ISP (SEVPM). The rate of processing of proISP<sup>S250A</sup> by  $\Delta^{18}$ -ISP followed a standard enzyme progression rate curve compared to the exponential processing characteristics observed for wt ISP alone (Fig. 2B). Western blotting using a primary antibody against the C-terminal hexahistidine tag confirmed that the affinity tag was still attached to ISP<sup>S250A</sup> during initial processing but was removed on prolonged incubation (Fig. 2C). Mass spectrometry suggested that the N-terminal extension together with the C-terminal tail sequence KLEHHHHHH was removed on full processing of proISP<sup>S250A</sup>, in line with the observations for wt ISP (theoretical mass of 31,538 Da versus measured mass of 31,547 Da). In the absence of active  $\Delta^{18}$ -ISP, there was no digestion of proISP<sup>S250A</sup> indicating that trace amounts of proteinases native to *Escherichia coli* were not responsible for the processing (Fig. S1).

**Proteolytic Processing of ISP Has Little Effect on Overall Tertiary and Quaternary Structure.** The circular dichroism (CD) spectra of proISP<sup>S250A</sup> and  $\Delta^{18}$ -ISP were almost identical (Fig. S24) and characteristic of proteins with a subtilisin-like fold (20, 21, 24). This indicates that there is no major change in the overall structure of ISP on processing.

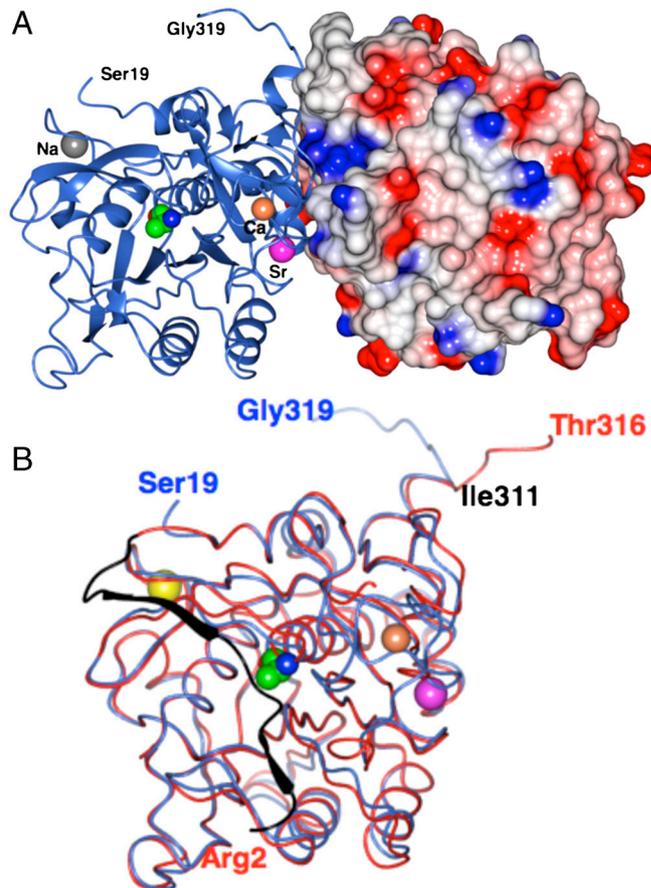
ISPs have a unique quaternary structure amongst the subtilisins in that they form dimers (21). Size exclusion chromatography revealed that  $\Delta^{18}$ -ISP eluted slightly later (0.2 mL) than proISP<sup>S250A</sup> (Fig. S2B), corresponding to an apparent molecular weight of 50 kDa, closer to that predicted for the dimer (63 kDa) than the monomer (31.5 kDa). This confirms that the dimeric quaternary structure observed is maintained in the processed form in solution.

**Refolding of Denatured  $\text{N}\Delta^{18}$ -ISP.**  $\text{N}\Delta^{18}$ -ISP was unfolded in 8 M urea, which totally abolished protease activity. On exchange back into native buffer, 88% of protease activity was restored, confirming that  $\text{N}\Delta^{18}$ -ISP had refolded to an active conformation (Fig. S3).

**N-terminal Extension Peptide Acts as an Inhibitor of ISP Activity.** A synthetic peptide,  $\text{N}\Delta 18$ , equivalent to the first 17 amino acids of ISP (lacking the N-terminal methionine) acted as an inhibitor in its own right when added to  $\text{N}\Delta^{18}$ -ISP, and substantially reduced proteolytic activity. The  $\text{IC}_{50}$  of the peptide was determined to be  $\sim 0.7 \mu\text{M}$  with Suc-Phe-Ala-Ala-Phe-pNA as the substrate (Fig. 3A). A more detailed kinetic analysis (Fig. 3B) revealed that the  $\text{N}\Delta 18$  inhibition data fitted poorly to a simple competitive inhibition model, but much better to mixed model with a significant noncompetitive inhibition contribution (see *Materials and Methods*). The calculated  $K_i$  of the  $\text{N}\Delta 18$  peptide was  $1(\pm 0.2) \mu\text{M}$ .

**Structure of  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$  and Comparison with proISP $^{\text{S250A}}$ .** To identify the molecular events that occur on proISP processing and activation, the structure of  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$  was determined. The final model refined to 2.6 Å spacing comprises residues 19–319, plus a sodium, a calcium and a strontium ion (Fig. 4A), with 25 waters. The crystallographic statistics are provided in Table S1. Ser19 is now the N-terminal residue. The overall fold of the mature protomer is similar to the core of proISP, with an rmsd over 276 C $\alpha$  atoms of 0.79 Å (Fig. 4B). The dimeric structure is retained (Fig. 4A) with the subunit interface largely conserved, and the two subunits of  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$  related by a crystallographic axis. However, there are substantial differences in three regions on maturation (Fig. 5).

Firstly, residues 247–251 shift position by  $\sim 1.6$  Å resulting in the catalytic residue (S250A) moving closer to His86 to form a catalytic triad with similar geometry to that observed for ESPs such as BPN' (Fig. 5A). The position of the histidine side chain varies somewhat, lying further away from the alanine in the  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$  than in BPN', as might be expected from the loss of hydrogen bonding potential in the inactive mutant. A number of rearrangements in and around the S1 pocket are associated with the triad reordering (Fig. 5B). The binding site nomenclature follows that of Schechter and Berger (25); the peptide substrate are labeled P4, P3, P2, P1, and P1' with the respective subsites on the enzyme termed S4, S3, S2, S1, and S1' and cleavage occurring between P1 and P1'. The S1 binding pocket is better defined than in proISP $^{\text{S250A}}$  with the ordering of loop residues 183–188 and 216–225, which contribute directly to the formation of the S1 binding pocket in ESPs. The 216–224 region is similar in  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$  and BPN', but very different in proISP (Fig. 5B). The 180–199 region is also different between the two forms of ISP. In proISP $^{\text{S250A}}$ , there is no density for

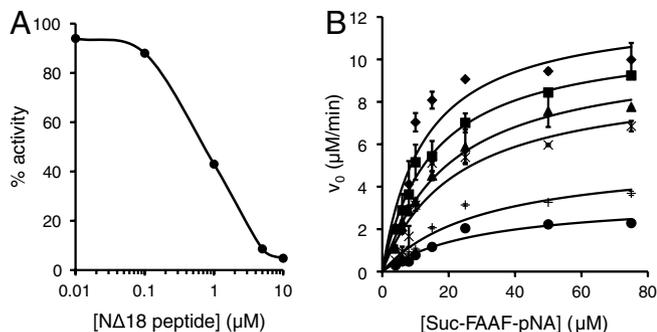


**Fig. 4.** Structure of  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$ . (A) Dimeric structure of  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$ . One subunit is shown as a ribbon, the other as an electrostatic surface. The positions of the N- (Ser19) and C-terminal (Gly319) residues are indicated. The catalytic Ser-to-Ala residue (colored by atom type), and the sodium (gray), calcium (orange), and strontium (magenta) ions are shown as spheres. (B) Superposition, using secondary structure matching (43), of  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$  (blue) and proISP $^{\text{S250A}}$  (red). The N- and C-terminal residues for each form are indicated and the N-terminal extension present only in proISP $^{\text{S250A}}$  is colored black. Metal ions colored as in (A).

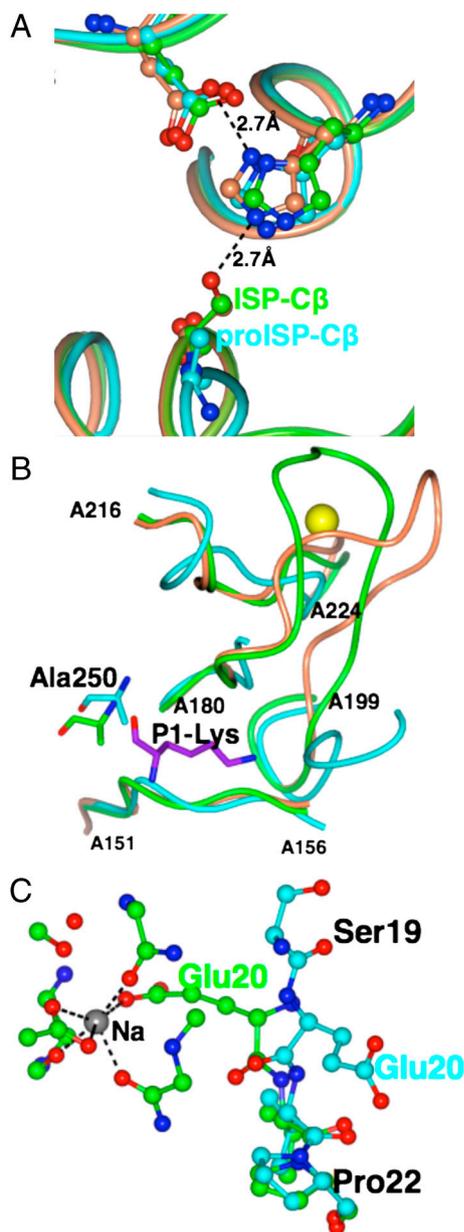
residues 184–194 (21) whereas this region is defined in  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$ . Within this loop there was an anomalous electron density peak whose height was half that at the  $\text{Sr}^{2+}$  site described below, which has been modeled as a calcium ion (Fig. 5B). No such calcium binding is observed to occur in proISP or BPN', with the equivalent region in BPN' having a substantially different structure (Fig. 5B).

Secondly, whereas the residues from Val21 onward have similar conformations in both  $\text{N}\Delta^{18}$ -ISP $^{\text{S250A}}$  and proISP $^{\text{S250A}}$ , the positions of Ser19 and Glu20 change on cleavage of the peptide bond following Leu18. This allows the side chain of Glu20 to rotate by roughly 180° from its position in proISP to act as the sixth ligand for the high affinity metal site (Fig. 5C). As in proISP, the site is occupied by a sodium rather than a calcium ion, with a coordination sphere very similar to that seen for calcium in the equivalent high affinity sites of the ESPs (21). The lack of any density at this site in the anomalous difference Fourier synthesis confirms that this is not a calcium ion.

Thirdly, the strongest feature in the anomalous difference Fourier is at a site close to the 2-fold axis, where there was a PEG molecule in proISP (21). We have modeled this as a strontium ion, consistent with the composition of the crystallization buffer, again with evidence for adjacent binding of PEG. Its significance is unclear, as it does not appear to change the interactions across the subunit interface when compared with proISP.



**Fig. 3.** Inhibition of ISP by the  $\text{N}\Delta 18$  peptide. (A) Determination of  $\text{IC}_{50}$ . (B) Enzyme kinetics in the absence (diamond) and presence of 0.5 (square), 1 (triangle), 2 (x), 3 (+), and 5  $\mu\text{M}$  (circle) of  $\text{N}\Delta 18$  peptide.



**Fig. 5.** Structural changes to ISP on processing. (A) Overlay of the catalytic triads. The C $\beta$  atom of the catalytic serine equivalent residue, Ala250, is indicated for both proISP<sup>S250A</sup> and N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup>. The hydrogen bond between the serine and histidine in the catalytic triad is shown for BPN'. The carbon atoms and backbone worm depiction of N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup>, proISP<sup>S250A</sup> and the ESP BPN' (PDB code 1TO2) are colored green, cyan and coral, respectively. (B) Structural rearrangements of loops forming the S1 pocket on processing. Structures colored as in (A). The catalytic Ser-to-Ala is shown as stick representation. Lys318 from a separate N $\Delta$ -ISP18 homodimer is colored purple. Calcium ion is shown as a sphere (yellow). (C) Repositioning of Glu20 on processing to form a fully coordinated metal binding site. The carbon atoms of N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup> and proISP<sup>S250A</sup> are colored green and cyan respectively.

In the absence of strontium, it is possible that this site is occupied by a calcium ion.

Removal of the N-terminal extension in the isolated N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup> dimer can be presumed to expose the catalytic triad to solvent. However in the crystal, the C-terminus from residue 311 onward rearranges, with residues 313–319 packing against a symmetry related dimer. Whereas this is clearly a result of crystal packing and is not likely to have any physiological significance, it is of interest that these residues act as a substrate mimic, taking the position of residues P1'–4 in the 3-stranded  $\beta$ -sheet that

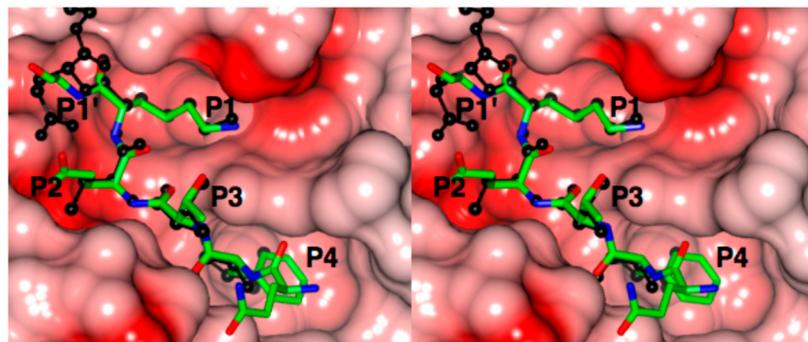
anchors proteinaceous inhibitors and substrates into the channel leading to the catalytic site (Fig. 6). The backbone atoms of residues 315–319 in N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup> overlap closely with those of the CI2 inhibitor in the BPN' complex (and indeed with those in other precursors and inhibitors). The side chains, in spite of differences in sequence, fit snugly into the four subsites, P1–4, with the lysines in P1 superposing particularly closely. In addition sites P2–4 overlap well with those in the proISP N-terminal extension (21).

## Discussion

Their atypical cellular location and unusual sequence features mark the bacterial ISPs as distinctive members of the subtilisin family. The work presented here together with the recently determined structure of proISP<sup>S250A</sup> (21) has enabled us to assign functional roles to these unique sequence elements so confirming their physiological relevance, especially in terms of regulating proteolytic activity. Given that the sequence features are shared amongst the ISPs (Fig. 1C), our findings for the *B. clausii* enzyme are relevant to the ISPs as a whole, and will have an impact on our molecular understanding of subtilisins in general and mechanisms by which protease activity within the cell is regulated. Regulation of proteases active within the cell is of paramount importance to prevent untimely degradation of vital cellular components. The N-terminal extension of the ISPs containing the conserved LIPY/F motif is a notable deviation from the secreted subtilisins (Fig. 1), and we have established here that it plays an important functional role by regulating proteolytic activity (Figs. 2 and 3).

Analysis of the proISP structure provides an obvious mechanism by which the N-terminal extension inhibits activity as it binds back across and blocks the active site (Fig. 1D); cleavage between Leu18 and Ser19 activates wt ISP (Fig. 2). However, the manner by which active site blocking is achieved differs compared to other proteases. When the N-terminal extension binds back over the active site in proISP, a key feature is the deviation of the main-chain away from the active site induced by the conserved LIPY/F motif contributing to the formation of the proline bulge (Fig. 1D). The positioning of Pro8 is critical as it points away from the body of the protein so moving the scissile peptide bond out of reach of the catalytic serine. This is in contrast to the mode of binding of the classical prodomains from ESPs (26–28) and the subtilisin-like proteins kumamolisin (29, 30) and fervidolysin (31) over the active site, which retain the target peptide bond in a position to allow autocatalytic processing. Interactions between subtilisins and classical proteinaceous inhibitors (32, 33) such as CI2 (34) are normally restricted to the active site and substrate binding region whereas the N-terminal extension of ISP forms distinct interactions with regions of the protein beyond the substrate binding pocket. The biochemical data (Fig. 2) combined with the structure of N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup> (Figs. 4 and 5) reveal that proteolytic cleavage of the N-terminal extension removes the physical block to the active and substrate binding site that will thus allow entry of a potential protein substrate.

The observed change in enzyme kinetics in the presence of the N-terminal extension peptide N $\Delta$ 18 (Fig. 3) unexpectedly suggested a mixed noncompetitive rather than a competitive mode of inhibition. Comparison of the structure of proISP with that of N $\Delta$ <sup>18</sup>-ISP<sup>S250A</sup> determined here provides an explanation and reveals that the N-terminal extension regulates protease activity via an additional mechanism: disruption of the catalytic triad and S1 binding pocket. Leu6 and Tyr9 within the N-terminal extension point toward the loop carrying the catalytic serine mutation resulting in displacement of the main-chain atoms of the loop containing Ala250 by  $\sim 1$  Å compared to other subtilisins, including those with the active site serine residue mutated to alanine (26, 30). This results in Ala250 being displaced by  $\sim 1.6$  Å from the position required for the formation of a catalytic



**Fig. 6.** Substrate binding pocket of  $N\Delta^{18}$ -ISP<sup>S250A</sup>. The bound section of the C-terminal region (residues 315–319) of a protomer from a separate dimer unit is colored by atom type. The BPN' Cl2 complex (PDB 1to2) was superposed on  $N\Delta^{18}$ -ISP<sup>S250A</sup> using SSM, and the binding of the inhibitor is shown in black. The mode of binding is closely similar in the P4-P1 sites.

triad. Removal of the N-terminal extension allows Ala250 to reposition enabling the native serine residue to make a productive interaction with His86 and form the catalytic triad (Fig. 5A). The S1 binding pocket also becomes structured and fully formed on maturation (Fig. 5B), facilitating substrate binding (Fig. 6). Together with the physical blocking of the active site by the N-terminal extension, disruption to both the catalytic triad and S1 substrate binding pocket are likely to give rise to the mixed inhibition mechanism indicated by the enzyme inhibition kinetics (Fig. 3).

It appears that the N-terminal extension of the ISPs is not involved in folding (Fig. S2) or its processing dependent on conformations accessed during folding (Fig. 2) thus playing a different role from that of the prodomain of ESPs (6). It does however allow the folded enzyme to be tightly regulated within the cell until such a time that its activity is required. This allows finer control over proteolytic activity compared to the coarser folding-dependent mechanism of the ESPs that only requires the protease to be inactive until it is secreted from the cell. Whereas it has previously been proposed that intracellular inhibitors of ISPs exist to regulate proteinase activity (35), we believe that the probable primary mode of regulation is through the N-terminal extension. This is supported by the observation that whereas the *E. coli* genome does not encode a known ISP homologue, high levels of *B. clausii* proISP can be produced intracellularly in *E. coli* without detriment to the cell.

Cleavage of the N-terminal extension at similar positions has been observed for ISPs from other species (13, 15, 16) and together with conservation of the LIPY/F motif (Fig. 1C) suggests a common role for the N-terminal extension. However, the N-terminal cleavage point between Leu18 and Ser19 is on a separate face of the enzyme distant from the active site (Fig. 4). As the whole of the extension makes extensive contacts with the rest of the protein, it is difficult to envisage its autocatalytic removal comparable to the prodomain processing in ESPs. We provide clear evidence that ISP itself is likely to be the main agent for the precise processing of the extension (Fig. 2) but this is likely to be by an inter—rather intramolecular event. Other structural families of serine proteases, such as trypsin, may also be activated through precise intermolecular proteolytic cleavage of their inactive precursor (e.g., trypsinogen) but the manner in which cleavage activates the protease is different to that of the ISP (36, 37). Intermolecular processing requires at the very least the presence of low levels of active ISP to initiate an activation cascade. This would explain the exponential activation of ISP *in vitro* (Fig. 2) and the lag of several hours between the production of transcripts and the appearance of proteolytic activity *in vivo* (12, 38).

It is not yet clear how the initial, low-level activity of ISP required to start the processing cascade is attained. Given that Pro8 introduces a kink at the position where the N-terminal extension interacts with the active site preventing hydrolysis of the peptide bond, it is unlikely that there are any intermediate autocatalytic processing events of the N-terminal extension

occurring prior to its full removal. However, a model involving metal ion binding has been proposed (21). Calcium plays a central role in signalling events as part of bacterial stress responses (39, 40). Changes in cytosolic  $Ca^{2+}$  may activate ISP through major structural changes that may disrupt interactions between the N-terminal extension and the rest of the protein leading to release of the LIPY motif from the active site. It is unlikely to involve the classical high affinity calcium binding site as both proISP<sup>S250A</sup> and  $N\Delta^{18}$ -ISP<sup>S250A</sup> bind sodium at this site but with crucial differences. In proISP<sup>S250A</sup>, binding of the N-terminal sequence back over the active site holds Glu20 in a position that prevents it from occupying the sixth metal ion coordination position. Upon processing, Glu20 is no longer constrained and flips by  $\sim 180^\circ$  to occupy the sixth coordination position (Fig. 5C). However, the presence of two distinct metal sites that bind calcium and the calcium mimic strontium in  $N\Delta^{18}$ -ISP<sup>S250A</sup>, suggest that calcium binding outside of the high affinity site may instigate the required structural changes. One of these sites is close to and helps organise the S1 binding pocket and the second is close to the dimer interface. This “open” conformation may be further stabilized by Glu20 now being able to occupy the final coordination position at the conserved metal binding site. Whereas the alternative open conformation is not observed in the crystal structure of proISP, it may represent a minor yet critical component of the ISP population required to initiate activation. Thus, metal ion binding may provide an additional mechanism by which ISPs are regulated.

The C-terminal residues from 311 onward move significantly from their position in proISP suggesting they are not required to maintain the dimer interface. In  $N\Delta^{18}$ -ISP<sup>S250A</sup>, residues 315–319 from one protomer bind within the active P4 to P1' groove of a symmetry related dimer. This is likely to be the result of crystal packing on exposure of the active and substrate binding sites on N-terminal removal. Although this may not represent a true protease-substrate interaction, this serendipitous event provides a striking insight into the precision of substrate recognition. Residues at the C-terminus bind in a similar fashion to that observed for the Cl2-BPN' complex (Fig. 6), most notably the P1 lysine residue.

The present study coupled with the recently determined structure of the full-length, unprocessed proprotein provides a timely insight into the molecular details encoded by the sequence features unique to the ISPs. This crucially includes a detailed description of the mechanism by which ISP proteolytic activity can be controlled posttranslationally within the cell, which differs from modes of inhibition observed for other subtilisins, and to our knowledge, other serine proteinases. Thus, uncovering the molecular features of ISP provides important insights that extend our understanding of the structure, function, and folding relationship within the large and important subtilisin family of serine proteases and the mechanism by which protease activity can be regulated within the confines of the cell.

## Materials and Methods

**Proteolytic Activity and Inhibition of ISP.** ISP was produced as outlined in *SI Materials and Methods*. To assess the role of N-terminal processing on activity, full-length proISP was incubated at 20 °C in buffer A (20 mM TrisHCl pH 8.0, 0.2 M ammonium sulphate, 1 mM calcium chloride) and aliquots taken at intervals between 0 and 10 hr and frozen in liquid nitrogen. ISP processing was analyzed by SDS-PAGE. Enzymatic activity was measured at 405 nm using a Varian Cary Bio UV/vis spectrophotometer with 50  $\mu$ M of the chromogenic substrate Suc-Phe-Ala-Ala-Phe-pNA and 0.25  $\mu$ M ISP. Processing of proISP<sup>S250A</sup> by activated  $\Delta$ 18-ISP was demonstrated by incubating 40  $\mu$ M proISP<sup>S250A</sup> and 1  $\mu$ M  $\Delta$ 18-ISP in buffer A at 25 °C. Aliquots were taken at time intervals between 0 and 9 hr and analyzed by SDS-PAGE. The relative amounts of unprocessed and processed protein were determined by densitometry of SDS-PAGE gels using the Labworks™ image analysis software. Western blotting, mass spectroscopy and refolding of  $\Delta$ 18-ISP were performed as described in *SI Materials and Methods*.

**Inhibition of ISP.** To determine the inhibitory effect of the N-terminal extension on ISP activity, a synthetic peptide, termed  $\Delta$ 18, corresponding to the N-terminal extension of ISP (RKFLIPYKQVDKVSAL) was synthesized by Activotec, Comberton. ISP activity was determined as described above over a  $\Delta$ 18 peptide concentration range of 0.01–10  $\mu$ M with 50  $\mu$ M Suc-Phe-Ala-Ala-Phe-pNA and 0.25  $\mu$ M of processed ISP. The IC<sub>50</sub> was determined by calculating the  $\Delta$ 18 concentration at which 50% of ISP activity was lost. To determine the inhibition mechanism, the activity of ISP was measured using 20 nM active ISP and Suc-Phe-Ala-Ala-Phe-pNA over a concentration range spanning 4–75  $\mu$ M. The  $\Delta$ 18 peptide concentration varied from 0.5 to 5  $\mu$ M. To determine the mode of inhibition and calculate  $K_i$ , least squares

nonlinear regression was used fit the data to the mixed inhibitor equation  $v = (V_{\max}/(1 + [I]/(\alpha K_i)) [S]) / (K_m(1 + [I]/K_i)/(1 + [I]/(\alpha * K_i)) + [S])$  using the GraphPad Prism 5 software (<http://www.graphpad.com>). The  $\alpha$  value reports on the mechanism of inhibition; large  $\alpha$  values are indicative of competitive inhibition whereas a value of 1 suggests pure noncompetitive inhibition. The value determined here was 2.7.

**Structure Determination of  $\Delta$ 18-ISP<sup>S250A</sup>.** A detailed description of the structure determination (including quaternary structure) of  $\Delta$ 18-ISP<sup>S250A</sup> is provided in *SI Materials and Methods*. Briefly, crystals grown from 20% PEG 3350, 0.1 M BisTrisPropane pH 6.5 and 0.2 M sodium bromide and soaked for 5 min in mother liquor supplemented with 0.5 M SrCl<sub>2</sub> and 23% (v/v) glycerol as a cryoprotectant, were used for X-ray crystallography. Data were collected to a resolution of 2.6 Å. The structure was solved using the CCP4 (41) suite of programs with one protomer of proISP (PDB 2x8j) used as a search model. The final R and  $R_{\text{free}}$  values were 19.8 and 28.1% respectively (Table S1). The X-ray data and the coordinates have been deposited in the Protein Data Bank (PDB) with the code 2xrm.

**ACKNOWLEDGMENTS.** 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. The authors thank the European Synchrotron Radiation Facility and to Diamond Light Source for access to beamlines and support by staff during visits, and the Bochtler group in Cardiff for help with preparation of crystals. We thank Johan Turkenburg and Sam Hart for assistance with data collection. We also thank Novozymes A/S for continuing support of K.S.W.

- Siezen RJ, Leunissen JA (1997) Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Sci* 6:501–523.
- Rockwell NC, Krysan DJ, Komiyama T, Fuller RS (2002) Precursor processing by kex2/ furin proteases. *Chem Rev* 102:4525–4548.
- Rockwell NC, 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.
- Paton AW, et al. (2006) AB5 subtilase cytotoxin inactivates the endoplasmic reticulum chaperone BiP. *Nature* 443:548–552.
- Withers-Martinez C, Jean L, Blackman MJ (2004) Subtilisin-like proteases of the malaria parasite. *Mol Microbiol* 53:55–63.
- Bryan PN (2002) Prodomains and protein folding catalysis. *Chem Rev* 102:4805–4816.
- Eder J, Fersht AR (1995) Pro-sequence-assisted protein folding. *Mol Microbiol* 16:609–614.
- Shinde U, Inouye M (2000) Intramolecular chaperones: Polypeptide extensions that modulate protein folding. *Semin Cell Dev Biol* 11:35–44.
- Bryan PN (2000) Protein engineering of subtilisin. *Biochim Biophys Acta* 1543:203–222.
- Wells JA, Estell DA (1988) Subtilisin—An enzyme designed to be engineered. *Trends Biochem Sci* 13:291–297.
- Carter P, Wells JA (1988) Dissecting the catalytic triad of a serine protease. *Nature* 332:564–568.
- Sheehan SM, Switzer RL (1990) Intracellular serine protease 1 of *Bacillus subtilis* formed in vivo as an unprocessed, active protease in stationary cells. *J Bacteriol* 172:473–476.
- Strongin AY, et al. (1979) Direct comparison of the subtilisin-like intracellular protease of *Bacillus licheniformis* with the homologous enzymes of *Bacillus subtilis*. *J Bacteriol* 137:1017–1019.
- Strongin AY, et al. (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.
- Strongin AY, et al. (1978) Intracellular serine protease of *Bacillus subtilis*: Sequence homology with extracellular subtilisins. *J Bacteriol* 133:1401–1411.
- Tsuchiya K, Ikeda I, Tsuchiya T, Kimura T (1997) Cloning and expression of an intracellular alkaline protease gene from alkalophilic Thermoactinomyces sp H5682. *Biosci Biotech Biochem* 61:298–303.
- Yamagata Y, Ichishima E (1995) A new alkaline serine protease from alkalophilic *Bacillus* sp.: Cloning, sequencing, and characterization of an intracellular protease. *Curr Microbiol* 30:357–366.
- Burnett TJ, Shankweiler GW, Hageman JH (1986) Activation of intracellular serine proteinase in *Bacillus subtilis* cells during sporulation. *J Bacteriol* 165:139–145.
- Orrego C, Kerjan P, Manca de Nadra MC, Szulmajster J (1973) Ribonucleic acid polymerase in a thermosensitive sporulation mutant (ts-4) of *Bacillus subtilis*. *J Bacteriol* 116:636–647.
- Subbian E, Yabuta Y, Shinde U (2004) Positive selection dictates the choice between kinetic and thermodynamic protein folding and stability in subtilases. *Biochemistry* 43:14348–14360.
- Vevodova J, et al. (2010) Crystal structure of an intracellular subtilisin reveals novel structural features unique to this subtilisin family. *Structure* 18:744–755.
- Ottmann C, et al. (2009) Structural basis for Ca<sup>2+</sup>-independence and activation by homodimerization of tomato subtilase 3. *Proc Natl Acad Sci USA* 106:17223–17228.
- Koide Y, Nakamura A, Uozumi T, Beppu T (1986) Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. *J Bacteriol* 167:110–116.
- Eder J, Rheinhecker M, Fersht AR (1993) Folding of subtilisin BPN': Characterization of a folding intermediate. *Biochemistry* 32:18–26.
- Schechter I, 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.
- Gallagher T, Gilliland G, Wang L, Bryan P (1995) The prosegment-subtilisin BPN' complex: Crystal structure of a specific "foldase". *Structure* 3:907–914.
- Ruan B, London V, Fisher KE, Gallagher DT, Bryan PN (2008) Engineering substrate preference in subtilisin: Structural and kinetic analysis of a specificity mutant. *Biochemistry* 47:6628–6636.
- Tanaka S, et al. (2007) Crystal structure of unautoprocessed precursor of subtilisin from a hyperthermophilic archaeon: Evidence for Ca<sup>2+</sup>-induced folding. *J Biol Chem* 282:8246–8255.
- Comellas-Bigler M, et al. (2002) The 1.4 Å crystal structure of kumamolysin: a thermostable serine-carboxyl-type proteinase. *Structure* 10:865–876.
- Comellas-Bigler M, et al. (2004) 1.2 Å crystal structure of the serine carboxyl proteinase pro-kumamolysin; structure of an intact pro-subtilase. *Structure* 12:1313–1323.
- Kim JS, Kluskens LD, de Vos WM, Huber R, van der Oost J (2004) Crystal structure of feravidolysin from *Fervidobacterium pennivorans*, a keratinolytic enzyme related to subtilisin. *J Mol Biol* 335:787–797.
- Bode W, Huber R (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 204:433–451.
- Laskowski M, Qasim MA (2000) What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim Biophys Acta* 1477:324–337.
- Radisky ES, Koshland DE, Jr (2002) A clogged gutter mechanism for protease inhibitors. *Proc Natl Acad Sci USA* 99:10316–10321.
- Shiga Y, Yamagata H, 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.
- Huber R, Bode W (1978) Structural basis of the activation and action of trypsin. *Accounts Chem Res* 11:114–122.
- Stroud RM, Kossiakoff AA, Chambers JL (1977) Mechanisms of zymogen activation. *Annu Rev Biochem* 6:177–193.
- Ruppen ME, Van Alstine GL, Band L (1988) Control of intracellular serine protease expression in *Bacillus subtilis*. *J Bacteriol* 170:136–140.
- Herbaud ML, Guiseppi A, Denizot F, Haiech J, Kilhoffer MC (1998) Calcium signalling in *Bacillus subtilis*. *Biochim Biophys Acta* 1448:212–226.
- Norris V, et al. (1996) Calcium signalling in bacteria. *J Bacteriol* 178:3677–3682.
- CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
- Potterton L, et al. (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D* 60:2288–2294.
- Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* 60:2256–2268.