The FEBS Journal

FEBS PRESS
science publishing by scientists

The LIPY/F-motif in an intracellular subtilisin protease is involved in inhibition

Journal:	*The FEBS Journal*
Manuscript ID	FJ-17-0989
Manuscript Type:	Regular Paper
Date Submitted by the Author:	18-Dec-2017
Complete List of Authors:	Bjerga, Gro Elin Kjæreng; Uni Research AS, Center for applied biotechnology Larsen, Øivind; Uni Research AS, Center for applied biotechnology Arsın, Hasan; Universitetet i Bergen Det Matematisk-naturvitenskapelige Fakultet, Department of Biology García-Moyano, Antonio; Uni Research AS, Center for applied biotechnology Williamson, Adele; UiT - The Arctic University of Norway, Department of Chemistry Leiros, Ingar; University of Tromsø, Dept. of Chemistry Puntervoll, Pål; Uni Research AS, Center for applied biotechnology
Key Words:	

Manuscript

The LIPY/F-motif in an intracellular subtilisin protease is involved in inhibition

Gro Elin Kjæreng Bjerga[1], Øivind Larsen[1], Hasan Arsın[2], Adele Williamson[3], Antonio García-Moyano[1], Ingar Leiros[3], Pål Puntervoll[1]

[1]Uni Research, Center for applied biotechnology, Thormøhlens gate 55, 5006 Bergen, Norway
[2]University of Bergen, Department of biology, Thormøhlens gate 53, 5006 Bergen, Norway
[3]Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

Corresponding author:
Gro Elin Kjæreng Bjerga
Telephone: (+47) 55 58 44 92 (office)
e-mail: gro.bjerga@uni.no
http://uni.no/

Abbreviations: ISP, intracellular subtilisin protease; ESP, extracellular subtilisin protease; IMAC, immobilized metal affinity chromatography; EDTA, ethylenediaminetetraacetic acid; DSC, differential scanning calorimetry; FITC, fluorescein isothiocyanate.

Enzymes, listed by EC-numbers: EC 3.4.21

Databases: Sequence data is available in the European Nucleotide Archive with the accession code XXX [TBA], and structural data is available in RCSB Protein Data Bank database under the accession number 6F9M.

Keywords: ISP, *Planococcus*, LIPY/F-motif, subtilisin, protease structure

Abstract

Intracellular subtilisin proteases (ISPs) have important roles in protein processing during the stationary phase in bacteria. Their protein degrading activity may have adverse effects inside a cell, but little is known about their regulatory mechanism. Until now, ISPs have mostly been described from *Bacillus* species, with structural data from a single homolog. This is the first study of a marine ISP originating from a phylogenetically distinct genus, *Planococcus* sp. The enzyme was successfully overexpressed in *E. coli,* and is active in presence of calcium, which is thought to have role in minor, but essential, structural rearrangements needed for catalytic activity. The ISP operates at alkaline pH and at moderate temperatures, and has a corresponding melting temperature around 60 °C. The high-resolution three-dimensional structure reported here, is the first representative ISP with an intact catalytic triad albeit in a configuration with an inhibitory pro-peptide bound. The pro-peptide is removed in other homologs, but the removal of the ISP pro-peptide in *Planococcus* sp. AW02J18 appears to be different, and possibly involves several steps. A first processing step is described here as the removal of two immediate N-terminal residues. Furthermore, the pro-peptide contains a conserved LIPY/F-motif, which was found to be involved in inhibition of the catalytic activity.

Introduction

ISPs have key roles in cell cycle regulation, specifically in protein recycling by processing proteins during transition to the stationary phase [1,2]. To prevent proteolysis that may be lethal to the cell, the activity of an intracellular protease must be tightly controlled. Although a potential ISP inhibitor protein has been identified [3,4], the primary mechanism of regulation is likely intrinsic [5,6]. In the precursor protein, an N-terminal pro-peptide of typically 16-20 residues binds across the active site and inhibits activity. As shown for a few homologs [6,7], the pro-peptide is released by intra-molecular maturation allowing the enzyme to act on exogenous substrates. ISPs are homodimeric [6], which contributes to making ISPs a structurally distinct family of subtilases. The catalytic domain of ISPs are homologous to those of other members of the Subtilisin superfamily, such as the extracellular subtilisin proteases (ESPs), which is a “Peptidase S8” domain in the Pfam classification [8].

Within this domain a catalytic triad, made up of an aspartate, a histidine and a serine, deprotonates the serine oxygen, and activates it for nucleophilic attack. Briefly, the nucleophile attacks unreactive carbonyl groups of the substrate, which ultimately leads to breakage of peptide bonds. Aside from homology within the catalytic domain, significant architectural differences exist. The N-termini of ESPs contain short leader sequences of about 20-30 residues for protein secretion [9], followed by a pro-domain of typically 60-80 residues [10,11], which is not conserved in sequence, but vital to their folding and function [12]. In an analogous manner to the ISP pro-peptide, the ESP pro-domain is processed intra-molecularly during maturation of the enzyme into an active conformation. The pro-domain has dual roles in acting as an inhibitor [13,14], and as a molecular chaperone that guides folding of the active enzyme [14–16].

The structure of ESP was first solved in 1969 [17], and has since been reported for several homologues [18,19] and a number of engineered mutants [20]. For ISPs, however, structural information is known from a single homologue, the *Bacillus clausii* ISP [5,6], with four structures reported (PDB

IDs: 2WVT, 2WWT, 2X8J and 2XRM), all from inactive mutants carrying catalytic Ser250 to Ala mutations. The four structures represent two activity states: the inactive state with the inhibitory pro-peptide binding and the active state without the pro-peptide bound.

In ISPs, the leader sequence and pro-domain of ESPs are replaced with a pro-peptide (also termed N-terminal extension). The pro-peptide binds across the active site, with residues Phe4-Leu6 forming a central β-strand of a three-stranded antiparallel β-sheet [6]. The pro-peptide also contains a LIPY/F motif, not found in ESPs. In *B. clausii* ISP this motif is involved in inhibiting the active site. Residues within the motif contribute to disruption of the conformation of the catalytic triad by shifting the catalytic Ser and His residues apart [5]. According to a standardized residue nomenclature for peptide binding to the active site [21], residues N-terminal to the scissile bond of the peptide substrate are termed P4, P3, P2, and P1, and those C-terminal to the bond are termed P1', P2', P3' and P4', where the scissile bond is between P1 and P1'. The corresponding sites in the enzyme are S4, S3, S2, S1, S1', S2', S3' and S4'. In *B. clausii* ISP, Leu6 and Ile7 correspond to P2 and P1 and are pointing inwards into the hydrophobic pocket at the S2 and S1 sites, respectively. Pro8 holds a unique position, which displaces the peptide bond between Ile7 (P1 site) and Pro8 (P1' site) out of reach of the active site Ser, whereas Tyr9 is occupying the S1' site. This particular Pro-induced "bridge" is unique in *B. clausii* ISP, and contrasts the scissile bond in ESPs, which is positioned to allow autoproteolytic processing. Altogether, the structure suggests that the residues in the pro-peptide are involved in blocking the active site serine [5,6].

Both ESPs and *B. clausii* ISP harbour a conserved high affinity metal-binding site occupied by a metal ion that serves a structural role [5,6,22,23]. The high affinity metal-binding site in ESPs is occupied by calcium [22,24], whereas in *B. clausii* ISP it is occupied by sodium [5,6]. In addition, *B. clausii* ISP has two unique binding sites for divalent metal ions, probably occupied by calcium ions, in each monomer: one close to the dimer interface and one in proximity to the active site. The latter is involved in ordering a loop that contributes to

formation of one of the binding sites (S1) involved in catalysis. Due to the processing of the pro-peptide and the positioning of calcium, the catalytic triad and substrate binding cleft is significantly rearranged, especially at the S1 binding site [5]. In a proposed model for ISP regulation [25], it was suggested that once a minor fraction of the pool of ISPs adopts an open conformation, calcium binding takes place and reshapes the S1 binding site, which ultimately releases the pro-peptide within this population and leads to a cascade of activation of other ISPs. The sequence of events and details of how the maturation precedes, in particularly the role of calcium, are not known.

This is the first study of an ISP from a marine isolate, *Planococcus* sp. AW02J18, which is from a related, but phylogenetically distinct genus to *B. clausii*. Here, we present biochemical data for the recombinant enzyme, showing it is active in presence of calcium, at alkaline pH and moderate temperatures. We furthermore present a high-resolution structure of the first ISP with an intact catalytic triad and an inhibitory pro-peptide bound across the active site. The structure supports previous findings and unique features of ISPs, such as its dimeric nature, sodium binding in the high-affinity metal-binding site and active site blocking by the pro-peptide. The processing of the pro-peptide appears however to be different from reported ISPs, possibly involving multiple processing steps. We also present mutagenesis data supporting an inhibitory role of the LIPY/F motif of the pro-peptide.

Materials and Methods

In silico identification of an intracellular subtilisin protease

The ISP sequence was identified from sequence-based mining of a marine bacterial isolate, *Planococcus* sp. AW02J18 (Table 1). This isolate was collected during expeditions in the coastal areas of Lofoten in 2009, and is stored in a bacterial collection at the University of Tromsø. The sampling procedure and collection has been presented elsewhere [26]. Genomic material was isolated for Illumina sequencing (MiSeq). Using a sequence-based approach, translated genomic sequences from a marine bacterial collection were mined for subtilisin-like proteases by searching for S08 family homologs against the MEROPS database [27]. The ISP candidate was identified in this data set, and the sequence has been deposited in European Nucleotide Archive with the accession code XXX [TBA].

The LIPY/F sequence conservation

Sequences homologous to *Planococcus* sp. AW02J18 ISP were identified using the UniProt blast search engine (default settings) against the UniRef90 database (UniProt release 2017_10) [28]. Sequence hit number 156, UniRef90_A0A136C445, was the first sequence to contain two motif mutations (LVNE) making the motif unlikely to be functional and was used to define the distance cut-off (expect value 4e-107; 57% sequence identify to *Planococcus* sp. AW02J18 ISP). Hence, the top 155 sequence hits were used to make a multiple sequence alignment (MAFFT, default settings) [29]. Three sequences were fragments that lacked the LIPY/F motif, and were manually removed (UniRef90: UPI00098840FB, UPI000590D2A7, UPI000689F3EC). The alignment containing the remaining 152 sequences was used to construct a sequence logo (default parameters) [30].

Sub-cloning of the *isp* gene to expression vectors

To facilitate enzyme expression we used our previously developed screening procedure for subtilisin-like serine proteases [31]. The *Planococcus* sp. AW02J18 ISP protein sequence was used as template for gene synthesis (GenScript), and the synthetic *isp* gene was codon-optimized to improve its expression in *E. coli*. The *isp* gene was synthesized with flanking *Sap*I sites,

and delivered in a customized *Sap*I-free pUC57 vector with kanamycin selection marker. The *isp* gene was sub-cloned from the delivery vector to a suite of expression vectors using a fragment exchange cloning method [32]. Construction of the expression vectors have been described previously [31].

Gene truncation and mutagenesis

Truncation constructs and mutants were prepared from the pUC57 template. Primers were designed to contain a *Sap*I-cloning site and a 15-20 bp gene-specific region targeting the desired truncation start. Primers in Table S1 were used to amplify the truncated ISP versions by PCR using Phusion polymerase. Gene fragments were purified, and cloned into the pINITIAL cloning vector by FX-cloning [31]. Plasmids were sequenced to confirm correct truncations. Point mutations were prepared by site-directed mutagenesis using primers in Table S1. Truncation constructs and mutants were sub-cloned into the p12 expression vector, as described above.

Small-scale expression and analysis of protein integrity

Small-scale recombinant expression was carried out according to the protocol described previously [31] in 4 mL culture volumes. Following expression, cells were collected and resuspended in 1 mL lysis buffer (50 mM Tris HCl pH 8.5, 50 mM NaCl, 0.25 mg/mL lysozyme, 10 % (v/v) glycerol). Lysis was completed by ultrasonication using two five-seconds pulses at 40-60 % amplitude with a CV-18 probe powered by an Ultrasonic Homogenizer 4710 (Cole Parmer). Lysates were cleared by centrifugation at 4600 x *g* for 20 minutes. Cleared lysate samples (representing soluble fraction) were analyzed by SDS-PAGE and immunoblot as described previously [31]. As background controls, lysates containing empty vector were used, herein termed GS due to the insertion of triple GS encoding sequence as a replacement of the *ccd*B gene in the expression vector [31].

Semi-quantitative analysis of recombinant protein in cleared extracts was performed in Image Lab 3.0 (BioRad). Target band intensities were extracted from image data of Coomassie-stained SDS-PAGE gels, and normalized to the total protein intensities in the lane excluding the target band intensities to adjust for variable growth rates and protein expression levels.

Large-scale expression

E. coli MC1061 cells containing the p1:ISP, p12:ISP or the p12:ISP-S251A (catalytic mutant) constructs were grown in 1 L terrific broth medium (1.2 % tryptone, 2.4 % yeast extract, 0.4 % glycerol, 17 mM KH_2PO_4 and 72 mM K_2HPO_4) supplemented with ampicillin (100 µg/mL) in 2.5 L Thomson's Ultra Yield™ flasks (Thomson Instrument Company). Protein expression was induced by 0.1% (w/v) *L*-arabinose overnight at 20 °C with 250 rpm shaking. Cells were collected by centrifugation (JLA-9.1000 rotor, Beckman) at 7500 x *g*, 30 min at 4 °C, and stored at -20 °C.

Protein purification

Frozen cell pellets from about 1 L culture were resuspended in 50 mM Tris HCl pH 7.5 at room temperature (RT, roughly around 20 °C), 150 mM NaCl and 0.25 mg/mL lysozyme. After incubation for 30 min at 37 °C and 250 rpm, the cell suspension was cooled on ice before sonication in a final concentration of 500 mM NaCl. Cell debris was removed by centrifugation at 20,000 x *g* for 20 min at 4 °C (JA-25.50 rotor, Beckman). The cleared lysate was loaded onto 2 x 5 mL HisTrap FF crude columns (GE Healthcare) equilibrated with 50 mM Tris HCl pH 7.5 (at RT), 500 mM NaCl and 10 mM imidazole. Bound proteins were eluted in the same buffer containing 800 mM imidazole. Fractions containing protein were pooled and dialyzed two times in Spectra Por® dialysis tubes (Spectrum Laboratories, Inc.) with 6-8 kilo dalton (kDa) MWCO against 1 L 20 mM Tris HCl pH 7.5 overnight at 4 °C. 1 mM $CaCl_2$ was added to a 50 µg/mL ISP solution and incubated overnight at RT during slow stirring, yielding what we herein term “matured ISP”. Protein solutions were concentrated using Amicon 10 kDa MWCO spin-filter columns (Merck) with buffer exchange to 50 mM Tris HCl pH 7.5, 50 mM NaCl and stored in aliquots at 4 °C at concentrations 80 mg/mL (WT) and 150 mg/mL (mutant). From 1 L expression culture yields of 0.2 g of purified matured ISP (no tags), and 0.4 g of the catalytic mutant (with C-terminal his-tag) were typically achieved. Purity was assessed by quantitative analysis in Image Lab 3.0 (BioRad), by extracting the target band intensities from image data of Coomassie-stained SDS-PAGE gels. MS analyses were performed at the

PROBE facility (University of Bergen, Norway). N-Terminal amino acid sequencing was carried out at Alta Bioscience (University of Birmingham, United Kingdom).

Casein-based activity assays

The protease fluorescent detection kit (Sigma-Aldrich) was used for routine detection of proteolytic activity as previously described [31,33]. Briefly, 10 µL lysate or 5 µM enzyme was assessed for activity on FITC-casein in 50 mM TrisHCl pH 8.5 (at RT), 50 mM NaCl, in absence or presence of 1 mM $CaCl_2$ in a total volume of 50 µL at 37 °C for 1 hour unless otherwise stated. For the mutants, activity was assessed using EnzChek™ Protease Assay Kit (ThermoFischer). 10 µg/mL BODIPY FL casein was prepared by resuspending the substrate in 50 mM Tris HCl pH 8.5 (at RT) and 50 mM NaCl. 12.5 µL of BODIPY-FL casein was used per reaction, with 10 µL cleared extract in 50 mM Tris pH 8.5 (at RT), 50 mM NaCl and 1 mM $CaCl_2$ in a final volume of 100 µL. Samples were incubated at 37 °C for 1 h, and fluorescence was read.

Determining the specific activity

Specific activity was determined using a protease colorimetric detection kit (Sigma-Aldrich). To avoid assay interference with amino groups from Tris, ISP was dialyzed against 25 mM borate/NaOH pH 8.2, 50 mM NaCl before assaying. Casein was solubilized in water at pH 8.3. One unit is defined as the amount of enzyme that will hydrolyze casein to produce color (as determined by addition of Folin-Ciocalteu's Reagent) equivalent to 1.0 µmole tyrosine per minute at pH 8.3 at 37 °C in presence of 10 mM $CaCl_2$.

Differential Scanning Calorimetry

Prior to Differential Scanning Calorimetry (DSC) measurements, aliquots of mature ISP at approximately 1 mg/mL were dialyzed into the following conditions overnight at 4 °C: 50 mM Hepes pH 8.0, 50 mM NaCl (DSC buffer); DSC buffer with 2 mM $CaCl_2$; DSC buffer with 1 mM ethylenediaminetetraacetic acid (EDTA). Thermal unfolding transitions were measured using a Nano-Differential scanning CalorimeterIII (Calorimetry

Sciences Corporation) from 5 to 75 °C with scan rates of 1 °C/s. Buffer from the final dialysis step was used as a reference. Data were analyzed using the NanoAnalyze software (TA Instruments).

Crystallization

Crystallization experiments were performed with a stock solution of purified mature ISP at 30 mg/mL in 50 mM TrisHCl pH 7.5 (at RT), 50 mM NaCl. Initial crystallization conditions were screened using the vapour diffusion sitting drop method set up by a Phoenix crystallization robot (Art Robbins Instruments). The plates were set up with 60 µl reservoirs solutions and sitting drops with equal amounts of reservoir solution mixed with protein stock solution in a total drop volume of 1 µl. The screens were incubated at 20 °C. Diffraction-quality crystals were obtained from six conditions, as outlined in Table S2.

X-ray data collection

Crystals grown in 0.25 M NH_4Ac, 21.73 % PEG 1500, 0.1 M Na-Citrate pH 4.0, were transferred through a cryoprotectant solution (crystallization conditions with 20 % (v/v) glycerol added, thereafter mounted in a nylon loop and flash-cooled in liquid N_2. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) beamline ID23EH1. The data were integrated by XDS/XSCALE [34], scaled and analyzed by programs in the CCP4 program suite [35] through autoPROC [36]. A summary of the data collection statistics is found in Table 2.

Structure determination

The crystal structure was solved by molecular replacement using MolRep in the CCP4 program package [35] with 2XRM [5] as search model (a representative structure of the homologous ISP from *B. clausii*). The initial refinement was executed in Refmac [37] followed by automated model improvement in Buccaneer [38]. The manual building was done in Coot [39] interspersed by cycles of refinement in Phenix [40] and resulted in final R_{cryst}/R_{free} values of 13.04/15.03. A summary of the refinement statistics is shown in Table S3. The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (www.rcsb.org) with the accession

code 6F9M. Figures presented in the results section were generated using Chimera [41].

Results

A new intracellular subtilisin protease with a conserved LIPY/F motif

A previously uncharacterized protease from *Planococcus* sp. AW02J18 was identified in an enzyme discovery initiative as a candidate for expression in *E. coli* (Table 1). According to sequence analysis, this protease contained a catalytic domain (Peptidase_S8/PF00082) as annotated by Pfam (residues 40-311, Figure 1A). Sequence analysis also revealed that it shared 53 % sequence identity to the previously described intracellular subtilisin protease (ISP) from *B. clausii* [6] (Figure S1). As expected from SignalP analysis, the ISP sequence does not contain a leader sequence to direct its export [42], and is thus predicted to have an intracellular localization. In stead, the *Planococcus* sp. AW02J18 ISP contains a short pro-peptide with a LIPY-sequence at the N-terminus, also identified in other homologs (Figure S1). Although the LIPY sequence has been reported as a conserved motif [6], evidence of its conservation has not previously been presented. To analyze the evolutionary conservation of the motif, sequences homologous to the *Planococcus* sp. AW02J18 ISP were collected. Using 152 UniRef90 sequences in a sequence alignment, we analyzed conservation of the motif in a context with two flanking residues on each side (eight residue window). A LIPY/F motif is derived from the alignment (Figure 1B). A hydrophobic leucine or valine, or in rare cases an isoleucine occurs at the first position. At the second position, the motif contains most often a hydrophobic isoleucine, but in certain sequences phenylalanine, leucine or valine. The third position is occupied by a highly conserved proline found in all but two sequences. This residue is structurally significant as part of the proline-induced “bridge” in *B. clausii* ISP, which positions the scissile bond between proline and the previous residue out of reach for autocatalysis. At the fourth position, an aromatic tyrosine, phenylalanine or in rare cases histidine occurs. At flanking positions of these four residues some consensus occurs, such as a charged residues at proximate positions to the LIPY/F motif, and hydrophobic residues at positions two residues upstream and downstream (Figure 1B). A four-residue motif can be expressed using the Prosite pattern syntax as [LVI]-[IFLV]-P-[YFH].

The first two residues of the calcium-dependent ISP is processed

The full-length *isp* gene from *Planococcus* sp. AW02J18 was sub-cloned to a suite of expression vectors for heterologous expression. From SDS-PAGE analysis, we found that all recombinant constructs yielded soluble enzyme, but that solubility was further improved by use of fusion tags (Figure 1C). Since many serine proteases require calcium for proper folding and structural stability, activity was assessed on fluorescein isothiocyanate (FITC) conjugated casein in the absence or presence of calcium ions. Compared to extracts from strains carrying empty vectors, all recombinant enzymes were active, but required calcium for activity (Figure 1D). The p1-construct encoding an N-terminal deca-histidine (His) tag was chosen for in-depth characterization due to its potential to yield a recombinant enzyme that would mimic the native processed ISP, and ease downstream purification (Figure 1). In the absence of calcium, immobilized metal affinity chromatography (IMAC) was used for protein purification of His-ISP (approx. 38 kDa). In analogy to the ISP from *B. clausii*, the enzyme was incubated in presence of calcium to mature by autoproteolysis. From SDS-PAGE we obtained a “matured ISP”, with an expected lower mass (approx. 35 kDa), of 95% purity observed (Figure 3). Using this matured ISP, we found that increasing concentrations of calcium had a positive effect on activity (Figure 2A), whereas EDTA inactivated the ISP (Figure 2B). From SDS-PAGE analysis of the reaction products, we found that the enzyme was processed or degraded in presence of calcium (Figure 2B). In absence of calcium or in calcium-depleted reactions, enzymes were however persistent against proteolysis (Figure 2B), and could be stored for one month without any effect on activity (data not shown).

To further understand the processing, calcium chloride was added at various concentrations to the full-length recombinant enzyme (His-ISP) at a pH range 7.0-8.5. SDS-PAGE revealed that two processed ISP species less than 37 kDa were identified in presence of 1 mM $CaCl_2$ (Figure 3). Increasing the concentration of calcium chloride up to 10 mM led to further processing as well as the appearance of degradation products (i.e. fragments smaller than the 31 kDa peptidase domain). MS-analyses were performed on six protein

fragments after calcium-induced activation, with identification of ISP peptides in all samples (data not shown). N-terminal sequencing was performed on the two processed species immediately below 37 kDa (protein bands numbered 2-3, Figure 3), but data were only conclusive for the uppermost processed protein. In this protein starting on Asn3, the artificial N-terminal residues (His-tag and 3C protease site) and the two first native residues of the ISP (MK) were processed. Tag-removal was confirmed by immunoblot analysis and compared to a catalytic mutant designed by replacing the catalytic Ser251 with Ala (Figure S2). The processing of *Planococcus* sp. AW02J18 ISP appears to occur in multiple steps.

Planococcus sp. AW02J18 ISP operates at moderate temperatures and alkaline pH

To identify its optimal conditions for further activity assessments, the *Planococcus* sp. AW02J18 ISP was characterized with respect to the specific activity, temperature and pH optimum in casein assays (Figure 4). It was found to operate optimally at pH 11, but was active across pH 7.0-11.0, whereas no activity was observed below pH 6.0 (Figure 4A). Precipitation was observed at pH 4.0 in both citrate and acetate buffers, likely explained by an estimated pI around 4. The temperature optimum was found to be around 45°C (Figure 4B). No was found above 60°C, which indicates that the protein is destabilized at high temperatures. Using optimal temperature (45 °C) in alkaline conditions (pH 8.3) and 10 mM $CaCl_2$ the specific activity of the ISP was determined to be 13 ± 1 U/mg.

To determine the thermal unfolding temperature of ISP, DSC measurements were carried out (Figure 5). ISP unfolded as a single peak, which could be fitted to two two-state transitions with melting temperatures (Tm) separated by approximately 3.0 °C (Table 3). In the DSC data, the apparent Tm in absence of calcium and EDTA was around 60 °C, which is consistent with the data on temperature optimum and stability (Figure 5A). Addition of $CaCl_2$ increased the directly measured T_{max} by 1.7 °C, and the apparent Tm by up to 3.0 °C indicating that calcium has a stabilizing effect on the enzyme (Figure 5B). The presence of EDTA slightly increased the apparent Tm (Figure 5C). Repeat

scanning did not give rise to any subsequent unfolding transitions, indicating that ISP does not refold on the timescale used for this experiment; therefore the thermodynamics of unfolding were not analyzed further. No exothermic signals indicative of aggregation were present in the raw data (not shown), and no visible precipitate was observed suggesting that these data can be used in a comparative manner to understand the effect of EDTA and calcium on the system.

Structure of ISP with an intact catalytic triad and pro-peptide

ISPs are distinct from ESPs with regards to the N-terminal pro-peptide, their dimeric structure, and the sodium binding in the high affinity metal binding site [5,6], but details regarding their maturation are still unclear. To shed light on the latter, the crystal structure of the “matured ISP” from Asn3 to Asn310 was determined by X-ray crystallography to a resolution of 1.3 Å (Figure 6). In addition to being the second unique structure of an ISP, it is the first structure of an ISP with a native catalytic triad, and it represents the highest resolution structure of this enzyme family to date. The structure of the ISP is dimeric, and each monomer includes an almost intact pro-peptide (from Asn3) bound across the active site. There are two molecules of triethylene glycol (Peg3) symmetrically bound at the dimer interface distant from the active site (Figure 6A), which may be adducts of Peg 1500 during crystallization or introduced during recombinant expression. In three structures of ISP from *B. clausii*, similar molecules are bound in this region: a strontium ion and a tetraethylene glycol molecule bound in an overlapping position (PDB ID: 2XRM); three water molecules bound in the same region (PDB ID: 2WWT); and Peg3 (PDB ID: 2X8J) almost perfectly overlapping the conformation observed in the *Planococcus* sp. AW02J18 ISP structure.

The structure contains a catalytic core (residues 20-310) overlapping the Pfam assigned Peptidase_S8 domain (residues 40-311). The first two residues, two loop regions (residues 184-191 and 217-223), and the C-terminal 20 residues are not defined in the electron density. Superpositioning of *Planococcus* sp. AW02J18 ISP (chain A) with the catalytic mutant *B. clausii* ISP (PDB ID: 2X8J, chain A) gave an RMSD of 0.67 Å across 282 atom pairs

in an improved fit where far-apart residues are removed (across all 292 atom pairs of residues in the alignment: 1.205 Å), confirming that they have the same overall fold (Figure 6B). Superpositioning showed that catalytic triad residues are structurally conserved, although distances are slightly different in each monomer. Two distinct conformations were modelled in each monomer due to poor electron density: Monomer A, residues 248-252 (including the catalytic triad residue Ser251) and Monomer B, residues 16-20 (including parts of the pro-peptide). In monomer A the distances between Ser251Oγ and His86Nε2 is 3.20 and 3.58 Å, respectively, whereas the corresponding distance for monomer B measures to 3.82 Å (Figure 6C). Superpositioning with the structure representing the active state of *B. clausii* ISP (PDB ID: 2XRM) has a shorter distance, although only estimated, as both *B. clausii* structures are Ser251Ala mutants. One surface loop (residues 97-104) is different, probably reflecting an insertion in the *Planococcus* sp. AW02J18 ISP (Figure S1). Although the side-chains of some residues in this loop (residues Asp100, Glu101 and Glu102) are visible only at low contour levels, a sodium ion in each monomer was putatively identified and modelled in electron density as for the *B. clausii* ISP structures (PDB IDs: 2XRM and 2X8J).

The two loop regions that are disordered (residues 184-191 and 217-223) in *Planococcus* sp. AW02J18 ISP are ordered in the structure that simulates the active state of the *B. clausii* ISP. Residues from both loops are contributors in the coordination of a calcium ion, in *B. clausii* ISP, these are: Asp186 (sidechain; SC), Arg188 (mainchain; MC), Thr191 (MC), Glu193 (SC) and Thr221 (SC). In *Planococcus* sp. AW02J18 ISP the residues contributing with specific side chain contacts to the calcium ion are conserved, while one of the two unspecific main chain contacts are not conserved (Figure S1).

Mutations in the LIPY/F motif of the pro-peptide relieve inhibition

Removal of the first 18 residues of *B. clausii* ISP by calcium treatment or by truncation released an ISP enzyme in an active conformation [5]. The proteolytic site for cleavage is however not conserved among ISPs (Figure S1). As calcium seemed to improve activity (Figure 2), but also further

process the ISP (Figure 3), we aimed at identifying the second processing site for maturation. Despite repeated efforts, mass spectrometry and N-terminal sequencing of various protein species isolated from SDS-PAGE gels did not reveal other processing than the removal of the two first residues. As an alternative approach, we designed various constructs where the N-terminal region of the *Planococcus* sp. AW02J18 ISP was truncated (Figure 7A). To design a close mimic of the N-terminus of native and processed enzyme, a p12-based construct was chosen (ISP-His, 38 kDa). This facilitated the full-length ISP sequence and respective truncation mutants with C-terminal His-tags albeit with two artificial residues at the N-terminus of recombinant enzyme (MS, Figure 7A). A Leu6-truncation construct was designed to remove the first 5 residues, not affecting the LIPY-sequence, to assay potential detrimental effects of removal of the β1-strand of the antiparallel β-sheet required for structural stability (Figure 7B). An Arg10 truncation construct (i.e. starting at Arg10) was designed to remove the LIPY-sequence from the native N-termini, to release auto-inhibition induced by the motif. The Thr15-Arg20 truncations were designed to truncate the pro-peptide in search for an active enzyme that would mimic the processed *B. clausii* ISP. Truncations preceding Arg20 were considered to be destructive as these were anticipated to interfere with secondary structure elements in the core of the catalytic domain according to the *B. clausii* ISP structures [5,6]. Positions of ISP truncations are summarized in Figure 7. None of the truncations were expected to impair the high affinity metal-binding site or dimerization, as previous reports have identified the binding site and the dimer interface in other distant regions of the protein [25]. According to SDS-PAGE analysis recombinant enzymes were either not obtained or below our detection limits (data now shown). Growth of *E. coli* was not affected by recombinant expression, suggesting that active enzymes, if present, were not lost due to cell death. In case the recombinant enzymes were present at undetectable levels, the truncated enzymes were assessed in an activity assay, but found not to present activity (Figure 7B).

The LIPY/F-motif (residues 6-9 in *Planococcus* sp. AW02J18 ISP) is conserved in pro-peptides of ISPs (Figure S1). In *B. clausii* ISP the LIPY-

sequence is involved in binding the hydrophobic pocket at the active site, wherein Pro holds a critical position in displacing the scissile bond between Ile and Pro out of reach of the active site serine [6]. According to structural data on *Planococcus* sp. AW02J18 ISP (Figure 6B) and *B. clausii* ISP [6] the LIPY-sequence is involved in binding the active site, potentially having critical roles in inhibiting auto-proteolysis or cleavage of exogenous peptides. To investigate whether the LIPY/F-motif is required for inhibition, we designed point mutations in the motif by targeting the side chains of Leu6 and Ile7, which are protruding into the hydrophobic pocket. We designed Ala and Lys mutations at both sites and a double alanine mutant (substituting both positions with Ala). According to SDS-PAGE analysis, the Leu6Ala, and both Ile single mutants were successfully expressed, but gave lower yields than wild-type ISP (Figure 7C). Expression levels for the Leu6Lys single mutant and the double mutant were low, if any, and variation occurred in independent experiments. The ratio of soluble protein to expressed protein was generally higher for the mutants than for wild-type ISP (data not shown). Cleared lysates containing the wild-type ISP and mutants were assessed in an *in vitro* BODIPY-casein assay and compared to extracts from strains carrying the empty vector (Figure 7D). As expected, the wild-type ISP was found to be active upon calcium treatment as determined from an increase in fluorescent signal. Upon calcium addition, the Leu6Ala, and both Ile mutants showed a similar response, but mutants showed a higher than baseline level of activity even in the absence of calcium. No activity was detected for the Leu6Lys mutant, probably because it was not expressed. The double mutant was however found to be active, despite the low expression levels. The activity of the double mutant was similar both in absence and presence of calcium, albeit low. In all cases, EDTA prevented activity, likely by chelating calcium at one or several binding sites.

Discussion

An ISP from *Planococcus* sp. AW02J18 is herein characterized in terms of its catalytic activity, stability and structure. For recombinant expression, we explored the utility of N-terminal His, His-SUMO or His-MBP fusion tags to promote soluble expression of ISP, as previous data have shown that N-terminal tags can be used for both intracellular [1] and extracellular serine proteases [31]. Expression trials showed that all fusion constructs were soluble (Figure 1). The ISP was activated by addition of calcium (Figure 3). The assumption that ISP requires pro-peptide processing for activation, e.g. as in *B. clausii* ISP, allowed exploitation of its native protease activity for intrinsic tag removal. Indeed, the construct with an N-terminal his-tag facilitated creation of a matured ISP without artificial tags (Figure S2).

The ISP operates at moderate temperatures, with optimal conditions at 45 °C (Figure 4), and unfolds at about 60 °C (Figure 5). The organism of which this ISP originates, *Planococcus* sp. AW02J18, was isolated from a marine habitat, and is known to thrive at cold to moderate temperatures (data not shown). Although some ISPs are active at neutral pH [7], *Planococcus* sp. AW02J18 ISP, like the majority of ISPs [2,43–45], has optimal activity at alkaline pH (Figure 4). So far, one ISP has been structurally characterized, namely the ISP from *B. clausii*. This study provides structural information on a second unique ISP that originates from a phylogenetically and physiologically distinct genus [46]. The ISP crystallized mostly at acidic pH (Table S2), and calcium was not found in any of the crystal. The lack of activity (or processing) below pH 7.0 (Figure 4) may partly explain why structures are in the inactive conformation. Whether lack of crystals at conditions above pH 7.0 is caused by degradation or because the active state does not promote crystal growth is impossible to say. Processing is not induced by pH shift alone (Figure 3), but requires calcium as well. Both ISPs were found to crystallize in a dimeric state; thus, dimerization appears to be a generic feature of ISPs. The two monomers contained regions of poor electron density in proximity to each other. These are most likely partially flexible regions as a consequence of the structural reorganization caused by the insertion of the pro-peptide in the substrate-binding region. The C-terminal 20 residues were not defined in

electron density, while in two different crystal forms representing structures of ISP from *B. clausii* (PDB ID: 2X8J and 2WWT), these residues are stabilized through interactions with symmetry mates. According to sequence alignments, the C-terminal region is not conserved (Figure S1), but the reason for this region being flexible in the structure of *Planococcus* sp. AW02J18 ISP is not clear. Ultimately, the requirement and role of the C-terminal residues in folding and dimerization of ISPs remains unclear.

From studies of *B. clausii* ISP, divalent metal ions, possibly calcium, bind close to the S1 pocket [5,6]. In the crystals of *Planococcus* sp. AW02J18 ISP, calcium was not identified at any of the metal binding site. Two loop regions were not defined in the electron density of matured ISP, which is also the case for the *B. clausii* ISPs containing the intact pro-peptide (PDB IDs: 2X8J and 2WWT). These loop regions are however ordered in the *B. clausii* ISP structure that simulates the active conformation of the enzyme, albeit with a catalytic mutation (PDB ID: 2XRM). Residues from both loops contribute to the coordination of a calcium ion, and these residues are conserved in aligned sequences (Figure S1). This could indicate a specific role of calcium in the transition from inactive to active enzyme, not only for the *B. clausii* ISP, but also for other ISPs. Matured ISP from *Planococcus* sp. AW02J18 was active in presence of calcium, but susceptible to self-degradation (Figure 2 and 3). The fact that ISPs were not active without exogenous addition of calcium suggests that available metal binding sites were not occupied after production. Due to conservation of calcium-coordinating residues (Figure S1), and the need for high EDTA concentrations to inhibit activity (Figure 2), low affinity for calcium is likely not the case. DSC results suggests that additional calcium is only slightly stabilizing, and tightly bound calcium (removable with EDTA) is not essential for overall stability (Figure 5). DSC showed however that calcium does have a minor stabilizing effect; thus suggesting that the added calcium in our assays contribute to minor structural rearrangements.

It is likely that there are structural rearrangements, such as pro-peptide flip-out or removal, in order for the two loops to order and coordinate calcium. The IP residues of the LIPY/F motif in the pro-peptide are spatially close to

residues in one of the loops that need to be reoriented upon calcium binding. The two residues form hydrophobic interactions to the side chain of Phe195 in our inactive structure and probably hinder this reorienting into the active conformation (this side chain appears to be shifted almost 15Å in the active state).

It is likely that the pro-peptide in the ISP from *Planococcus* sp. AW02J18 is removed, in analogy to several Bacillus ISPs [5,7]. The removal of the ISP pro-peptide in *Planococcus* sp. AW02J18 appears to be different, and possibly involves several steps (Figure 3). In the first step the two first residues of the ISP (Met1, Lys2) are removed, as identified in the crystal and by N-terminal sequencing. Another product, which appears as the main product (around 30-35 kDa) at pH 8.5 in presence of 10 mM $CaCl_2$ (Figure 3), could possibly be functional. This product could in principle arise from processing of the C-terminal regions of the protein, too, which was not identified in the crystal. The N-terminal residues of this protein could not be identified. A truncation experiment was conducted to trim the pro-peptide in the hunt for the processing site. Two artificial residues (MS) are unavoidably added to the N-terminal end of these truncation constructs, which arise from fusion of the *isp* gene fragment to the start codon and the ligation seam added during sub-cloning (Figure 7A), and their negative interference on protein stability cannot be ruled out. Sequence analysis of *Planococcus* sp. AW02J18 ISP, reveals that it contains two Pro's in the transition from the pro-domain to the catalytic domain (Figure S1). Whereas Pro at the P2 site is likely accepted, Pro at the P1 is highly unlikely due to the preference of hydrophobic residues at the S1 site [25]. Multiple Pro's are normally not found to be present in sites for autoproteolysis by serine proteases [47], and the Pro's may instead serve a structural role [48]. This does not however rule out that other proteases, for example proline-specific endopeptidases, could process and remove the pro-domain in native conditions, or that processing site(s) are in other regions of the pro-peptide.

Although it has been found that the pro-peptide of *B. clausii* ISP has a role in inhibition, the contribution of the conserved residues within the LIPY/F-motif

has not been studied in detail. Due to the fact that Leu and Ile are conserved in the motif, and that the ISPs likely prefer hydrophobic amino acids at the S2 and S4 sites [25], we studied point mutations of Leu6 and Ile7 in *Planococcus* sp. AW02J18 ISP. Three of the four single point mutations, which resulted in increased activity - even in the absence of excessive calcium, indicates that Leu6 and Ile7 have substantial roles in inhibition and support the involvement of calcium during activation. A closer inspection of the structural context suggests that substitution of Leu6 with Ala likely reduced the hydrophobic interaction to the active site, and thus relieves the inhibition. Structural explanations for the Ile7 mutants were not conclusive due to their proximity to the flexible region (183-193), but it is likely that both mutations cause reduced interactions with the pro-peptide. We thus conclude that the pro-peptide, with the LIPY/F motif in a central position, is involved in inhibition. Our data is in line with the proposed ISP model [25], suggesting that calcium binding at the active site is prevented during pro-peptide inhibition.

Acknowledgements

We would like to thank Arne O. Smalås for sharing sequence data, Stefan Hauglid and Trine Carlsen for expert assistance on crystallization and support of staff during visit, and Hilde Eide Lien for excellent work on truncation and mutant studies. Provision of beam time at the European Synchrotron Radiation Facility (ESRF) is highly valued. The authors thank the Research Council of Norway for financial support (221568).

Author Contributions

GEKB designed the study, designed and supervised experiments and drafted the manuscript, ØL carried out expression, purification, and biochemical assays, HA set up crystallization trials and performed the DSC experiment, IL performed data collection, processed data, determined the structure and refined it, AGM performed phylogenetic analysis, AW supervised and designed the DSC experiment. PP supervised mutant design and performed bioinformatic analyses. All authors were involved in revision of the manuscript and approved the final version.

Competing Interest

None of the authors declare competing interests.

References

1 Lee AY, Goo Park S, Kho CW, Young Park S, Cho S, Lee SC, Lee DH, Myung PK & Park BC (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**, 3437–45.

2 Sheehan SM & 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.

3 Nishino T, Shimizu Y, Fukuhara K & Murao S (1986) Isolation and Characterization of a Proteinaceous Protease Inhibitor from Bacillus subtilis. *Agric. Biol. Chem.* **50**, 3059–3064.

4 Rigden DJ, Xu Q, Chang Y, Eberhardt RY, Finn RD & Rawlings ND (2013) The first structure in a family of peptidase inhibitors reveals an unusual Ig-like fold. *F1000Research* **2**, 154.

5 Gamble M, Künze G, Dodson EJ, Wilson KS & 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. U. S. A.* **108**, 3536–3541.

6 Vévodová J, Gamble M, Künze G, Ariza A, Dodson E, Jones DD & Wilson KS (2010) Crystal Structure of an Intracellular Subtilisin Reveals Novel Structural Features Unique to this Subtilisin Family. *Structure* **18**, 744–755.

7 Jeong YJ, Baek SC & Kim H (2017) Cloning and characterization of a novel intracellular serine protease (IspK) from Bacillus megaterium with a potential additive for detergents. *Int. J. Biol. Macromol.*

8 Finn RD, Mistry J, Tate J, Coggill P, Heger a., Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR & Bateman a. (2009) The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222.

9 Wong SL & Doi RH (1986) Determination of the signal peptidase cleavage

site in the preprosubtilisin of Bacillus subtilis. *J. Biol. Chem.* **261**, 10176–81.

10 Wells JA, Ferrari E, Henner DJ, Estell DA & Chen EY (1983) Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin in Bacillus subtilis. *Nucleic Acids Res.* **11**, 7911–25.

11 Vasantha N, Thompson LD, Rhodes C, Banner C, Nagle J & Filpula D (1984) Genes for alkaline protease and neutral protease from Bacillus amyloliquefaciens contain a large open reading frame between the regions coding for signal sequence and mature protein. *J. Bacteriol.* **159**, 811–819.

12 Bryan PN (2002) Prodomains and protein folding catalysis. *Chem. Rev.* **102**, 4805–16.

13 Power SD, Adams RM & Wells J a (1986) Secretion and autoproteolytic maturation of subtilisin. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3096–100.

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

15 Ikemura H, Takagi H & Inouye M (1987) Requirement of pro-sequence for the production of active subtilisin E in Escherichia coli. *J. Biol. Chem.* **262**, 7859–7864.

16 Zhu XL, Ohta Y, Jordan F & Inouye M (1989) Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* **339**, 483–484.

17 Wright CS, Alden RA & Kraut J (1969) Structure of subtilisin BPN' at 2.5 angström resolution. *Nature* **221**, 235–42.

18 Neidhart DJ & Petsko GA (1988) The refined crystal structure of subtilisin Carlsberg at 2.5 A resolution. *Protein Eng.* **2**, 271–6.

19 Betzel C, Klupsch S, Papendorf G, Hastrup S, Branner S & Wilson KS (1992) Crystal structure of the alkaline proteinase Savinase from Bacillus lentus at 1.4 A resolution. *J. Mol. Biol.* **223**, 427–45.

20 Wells JA & Estell DA (1988) Subtilisin - an enzyme designed to be engineered. *Trends Biochem. Sci.* **13**, 291–7.

21 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.

22 Bode W, Papamokos E & 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.

23 Bryan PN, Rollence ML, Pantoliano MW, Wood J, Finzel BC, Gilliland GL, Howard AJ & Poulos TL (1986) Proteases of enhanced stability: characterization of a thermostable variant of subtilisin. *Proteins* **1**, 326–34.

24 Pantoliano MW, Whitlow M, Wood JF, Rollence ML, Finzel BC, Gilliland GL, Poulos TL & Bryan PN (1988) The engineering of binding affinity at metal ion binding sites for the stabilization of proteins: subtilisin as a test case. *Biochemistry* **27**, 8311–7.

25 Gamble M, Künze G, Brancale A, Wilson KS & Jones DD (2012) The role of substrate specificity and metal binding in defining the activity and structure of an intracellular subtilisin. *FEBS Open Bio* **2**, 209–215.

26 De Santi C, Altermark B, de Pascale D & Willassen N-P (2016) Bioprospecting around Arctic islands: Marine bacteria as rich source of biocatalysts. *J. Basic Microbiol.* **56**, 238–53.

27 Rawlings ND, Waller M, Barrett AJ & Bateman A (2014) MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, 503–509.

28 The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169.

29 Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–80.

30 Crooks GE, Hon G, Chandonia J-M & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–90.

31 Bjerga GEK, Arsın H, Larsen Ø, Puntervoll P & Kleivdal HT (2016) A rapid solubility-optimized screening procedure for recombinant subtilisins in E. coli. *J. Biotechnol.* **222**, 38–46.

32 Geertsma ER & Dutzler R (2011) A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272–8.

33 Twining SS (1984) Fluorescein isothiocyanate-labeled casein assay for proteolytic enzymes. *Anal. Biochem.* **143**, 30–4.

34 Kabsch W (2010) XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32.

35 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–42.

36 Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T & Bricogne G (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 293–302.

37 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240–55.

38 Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 1002–11.

39 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–32.

40 Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–21.

41 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–12.

42 Petersen TN, Brunak S, von Heijne G & Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–6.

43 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**, 357–66.

44 Kirimura J, Shimizu A, Kimizuka A, Ninomiya T & Katsuya N (1969) Contribution of peptides and amino acids to the taste of foods. *J. Agric. Food Chem.* **17**, 689–695.

45 An S-Y, Ok M, Kim J-Y, Jang M-S, Cho Y-S, Choi Y-L, Kim C-H & Lee Y-C (2004) Cloning, high-level expression and enzymatic properties of an intracellular serine protease from Bacillus sp. WRD-2. *Indian J. Biochem. Biophys.* **41**, 141–7.

46 Yoon J-H, Kang S-J, Lee S-Y, Oh K-H & Oh T-K (2010) Planococcus salinarum sp. nov., isolated from a marine solar saltern, and emended description of the genus Planococcus. *Int. J. Syst. Evol. Microbiol.* **60**, 754–8.

47 Kim JC, Cha SH, Jeong ST, Oh SK & Byun SM (1991) Molecular cloning and nucleotide sequence of Streptomyces griseus trypsin gene. *Biochem. Biophys. Res. Commun.* **181**, 707–13.

48 Vanhoof G, Goossens F, De Meester I, Hendriks D & Scharpé S (1995) Proline motifs in peptides and their biological processing. *FASEB J.* **9**, 736–44.

Tables

Table 1: **Information on the ISP candidate**

ORF tag	PRTUIT00110
Isolate (NCBI taxonomy ID)	*Planococcus* sp. AW02J18 (1379956)
Isolate origin (Degrees Lat Long)	Lofoten (68.5025473N°, 015.0046585E°)
Isolate source (depth in meters)	Biota (135)
Length (aa)	329
SignalP leader sequence	No
Pfam domain (name)	Peptidase_S08
Pfam domain (aa)	40-311
Closest MEROPS hit (ID)	MER324776
Closest MEROPS organism	*Planococcus donghaensis*
Identity (%)	72.0

Table 2: Data collection and processing statistics. Values in parentheses are for the outermost shell.

Diffraction source/Beamline	ESRF ID23EH1
Wavelength (Å)	0.97625
Detector	Q315R CCD (ADSC)
Crystal-to-detector distance (mm)	214.77
Rotation range pr. image (°)	0.1
Total rotation range (°)	130
Space group	$P2_12_12_1$
a, b, c (Å)	70.17, 85.18, 104.58
α, β, γ (°)	90, 90, 90
Mosaicity (°)	0.2
Resolution range (Å)	66.05-1.298 (1.32-1.298)
Total No. of reflections	675238 (36457)
No. of unique reflections	142753 (7606)
Completeness (%)	92.2 (99.1)
Multiplicity	4.7 (4.8)
$<I/\sigma(I)>$	16.4 (2.2)
$R_{p.i.m.}$	0.026 (0.412)
Wilson B-factor ($Å^2$)	17.39

Table 3: Thermal denaturation measured by DSC. ΔH_{cal} (calorimetric enthalpy), ΔS (entropy of unfolding) and T_{max} are calculated directly from the unfolding transition. ΔH_{vH} and Tm are derived from fitting two two-stated scaled models to each transition after subtraction of buffer scans and a sigmoidal baseline.

Treatment	ΔH_{cal}	Tmax	ΔS	$\Delta H_{vH}1$	Tm1	$\Delta H_{vH}2$	Tm2	ΔD
None	120.1	60.7	0.356	105.0	57.2	209.1	61.0	0.49
2 mM $CaCl_2$	145.1	62.4	0.4211	142.0	60.1	266.0	62.8	0.70
EDTA	125.0	61.1	0.345	111.7	58.0	215.2	61.4	0.39

Figures

Figure 1

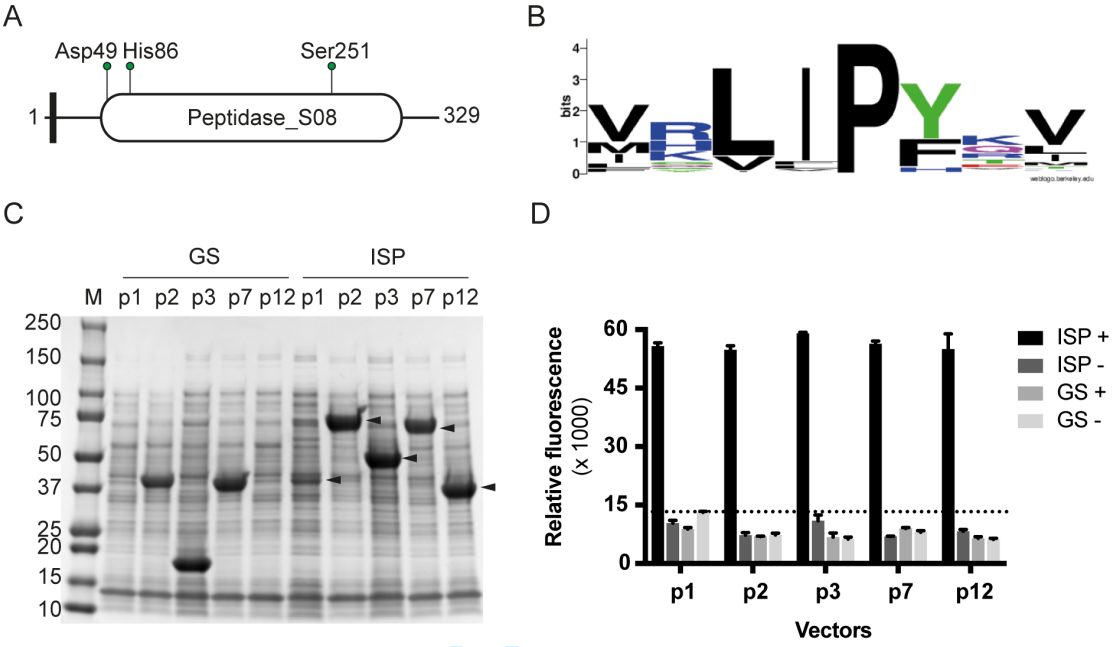

Figure 1 Overexpression and activity assessment of the recombinant ISP. A) A cartoon of the ISP architecture drawn to scale. Black box, LIPY/F motif; oval circle, Peptidase_S08 Pfam domain (PF00082); green pins point to residues involved in catalysis (catalytic triad). **B)** Sequence logo showing the evolutionary conservation of the LIPY/F motif based on an alignment with 152 ISP sequences. **C)** ISP constructs were produced from multiple vectors, and cleared lysates were inspected on SDS-PAGE for the presence of soluble overexpressed proteins. Arrows indicate soluble ISP proteins. Fusion partners from the various vectors are: p1, N-terminal His-tag; p2, N-terminal His-tag and MBP; p3; N-terminal His-tag and SUMO protein; p7, N-terminal MBP and C-terminal His-tag; p12; C-terminal His-tag. Empty vector controls (GS) will produce fusion partners only, wherein MBP and SUMO can be observed on SDS-PAGE. M, BioRad's Precision Plus Protein™ Dual Color Standard. **D)** Cleared lysates (see C for details), including empty vector controls (GS) were assayed over night with FITC-casein, in the presence (+) or absence (-) of 1 mM $CaCl_2$. Dotted line indicates the highest data point for background measurements (in the absence of calcium).

Figure 2

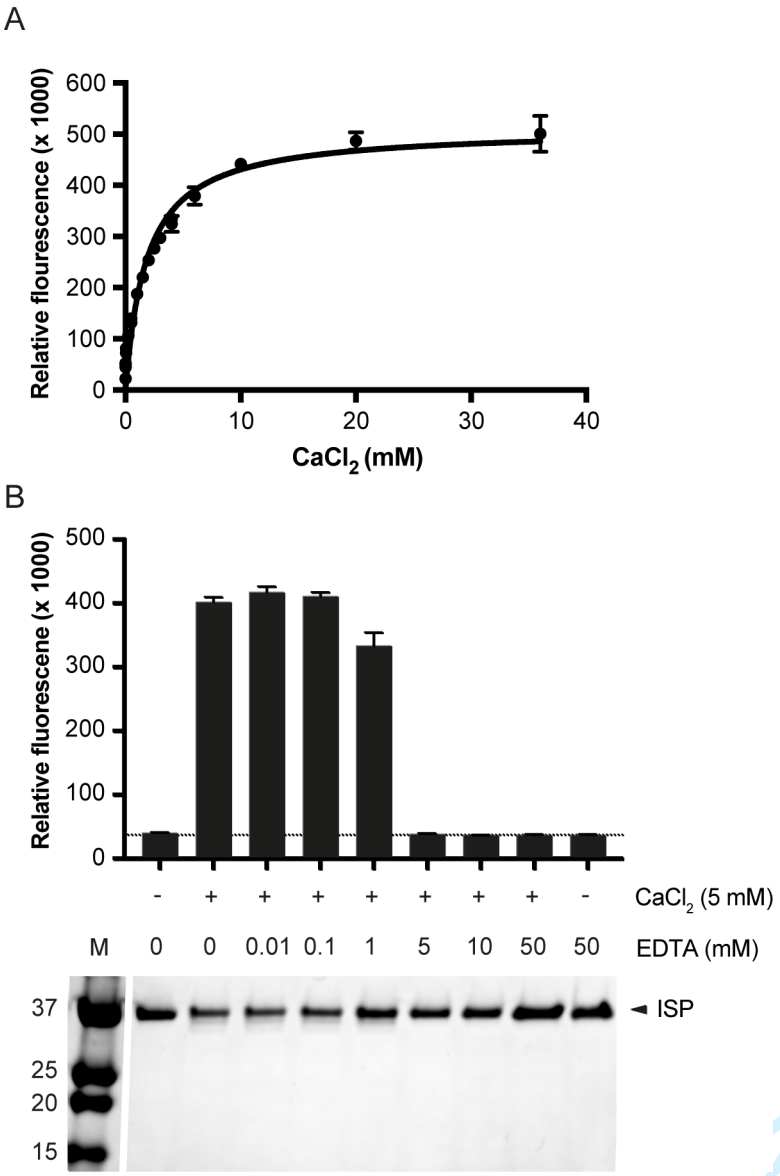

Figure 2. ISP activity and stability. A) Using the FITC-casein assay, 4 µM purified and matured ISP (p1 construct) was incubated with increasing concentration of $CaCl_2$ at 37 °C for 1 hour. 50% activity is achieved with 2.5 mM $CaCl_2$. **B)** The activity of 1 µg ISP (as in A) was measured in the presence or absence of 5 mM $CaCl_2$ and concentrations of EDTA up to 50 mM (upper panel). Dotted line represents the average of buffer (37248 units) in presence of 5 mM $CaCl_2$ and 5 mM EDTA. Lower panel shows SDS-PAGE containing 0.5 µg purified and matured ISP treated with $CaCl_2$ and EDTA as in the activity assay. M, BioRad's Precision Plus Protein™ Dual Color Standard.

Figure 3

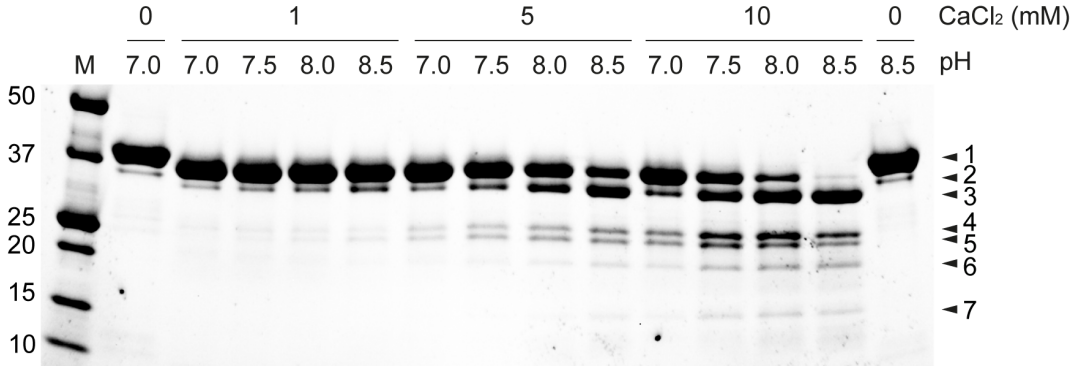

Figure 3 Processing of the ISP. The His-ISP protein construct (p1-construct, numbered 1 in the right panel) was used to investigate calcium-induced maturation. 1, 5 or 10 mM $CaCl_2$ was added to 2 µg enzyme at pH range 7.0-8.5 at room temperature for incubation overnight before analysis on SDS-PAGE. Indicated proteins (numbered 1-7) were all confirmed by MS. Some proteins (numbered 2-3) were analysed by N-terminal sequencing in addition. Theoretical mass of matured ISP (numbered 2), 35 kDa. M, BioRad's Precision Plus Protein™ Dual Color Standard.

Figure 4

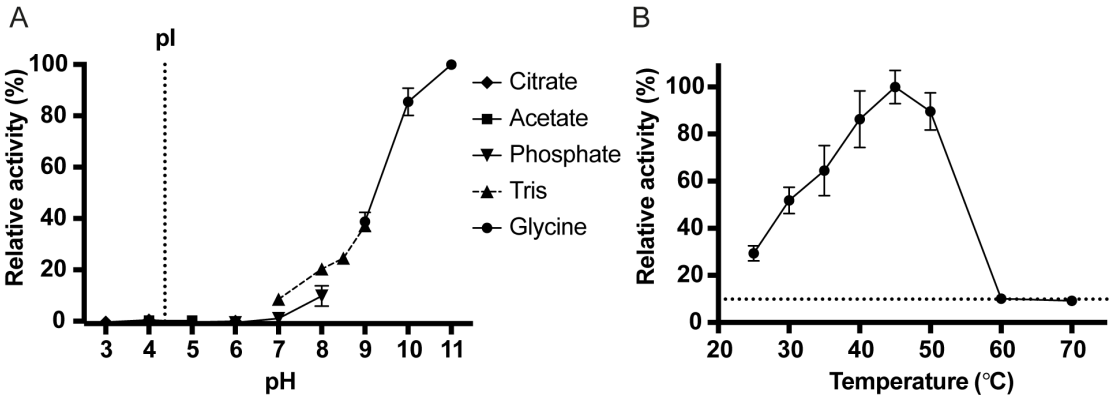

Figure 4. pH and temperature optimum of ISP. A) Using the FITC-casein assay, the activity of 5 µM purified and matured ISP (p1 construct) at pH 3-11 was measured in the initial rate of the reaction at 37 °C. Background from buffer was subtracted and data was made relative to measurement data at pH 11. Citrate buffer was used for pH 3.0–6.0 (diamonds), Acetate buffer for pH 4.0-6.0 (square), Sodium Phosphate buffer from pH 6.0–8.0 (down-pointing triangles), Tris-HCl buffer for pH 7.0–9.0 (up-pointing triangles, dotted line between points) and Glycine buffer (circles) for pH 9.0–11.0. Error bars represent deviation between two replicas in one experiment. The pI of the ISP is estimated to approximately 4.37 (vertical dotted line). **B)** Activity of 5 µM purified and matured ISP (p1 construct) was monitored across a temperature range of 25-70C, background subtracted and made relative to the measured data at 45°C. $CaCl_2$ was added immediately before assaying. The assay took place for 1 hour at the respective temperatures. Error bars represent deviation between data points from three independent experiments. The horizontal dotted line represents the highest background measurement.

Figure 5

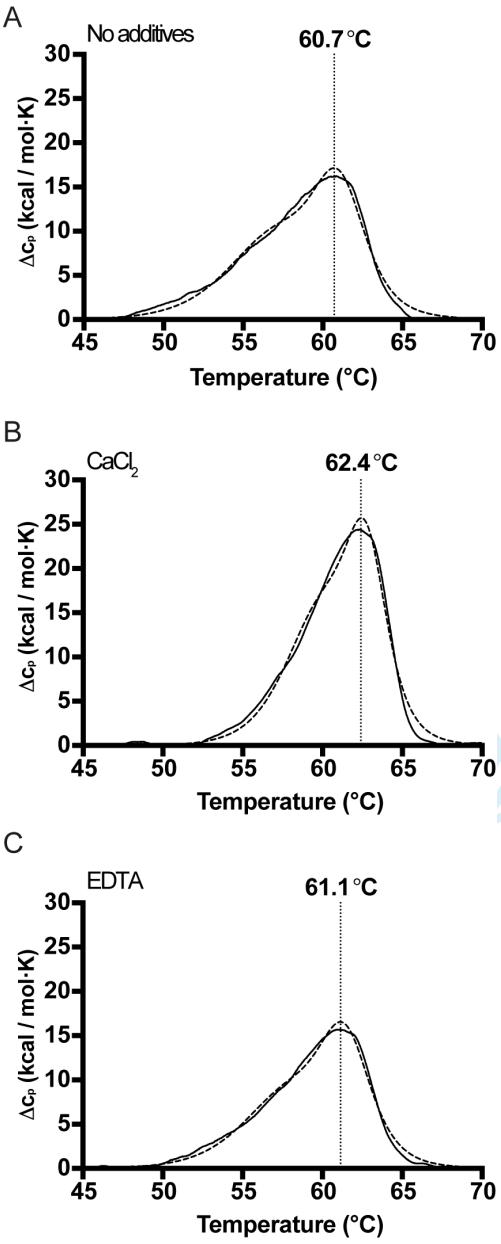

Figure 5. Thermal unfolding transitions of matured ISP. Unfolding was measured in metal-depleted ISP in three conditions; A) without additives, B) in presence of 2 mM $CaCl_2$, and C) in presence of 1 mM EDTA. Representative thermograms are shown after subtraction of buffer scans and fitting of a sigmoidal baseline (solid lines). The sum of the two two-state models fitted to each thermogram is shown with dashed lines. The Tm of the higher temperature transition is indicated with a vertical dotted drop-line for comparison.

Figure 6

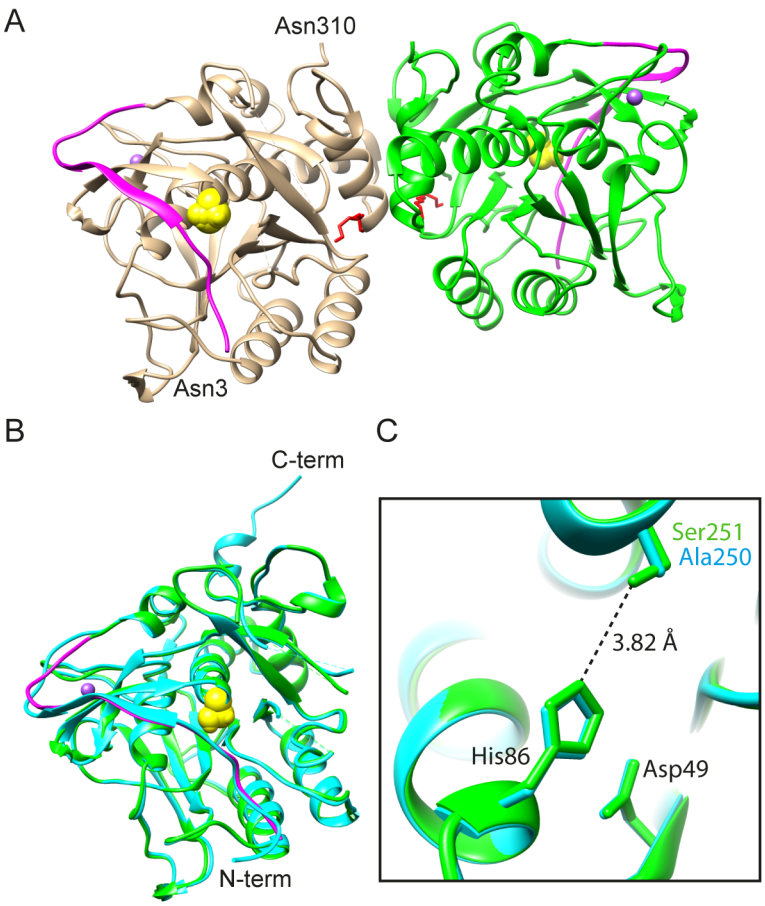

Figure 6. Structure of *Planococcus* sp. AW02J18 ISP. A) Dimer presented in ribbon. The pro-peptide (residues 3-20) is shown in magenta in both monomers (chain A in tan, chain B in green). The N-terminal and C-terminal residues are labelled in monomer A. The catalytic Ser251 is shown as a yellow sphere. The PEG molecules in the dimer interface are shown in red. **B)** Superposition of the ISP from *Planococcus* sp. AW02J18 (green, chain B) on the ISP template from *B. clausii* (blue, PDB ID: 2X8J, chain A). Ser251 in ISP from *Planococcus* sp. AW02J18 is shown as a yellow sphere, and its pro-peptide (residues 3-20) is shown in magenta. **C)** The catalytic triad of *Planococcus* sp. AW02J18 ISP (green, chain B) and the catalytic mutant of *B. clausii* ISP (cyan, chain A). Distance (Å) between Ser251 and His86 in *Planococcus* sp. AW02J18 ISP is given as a dashed line.

Figure 7

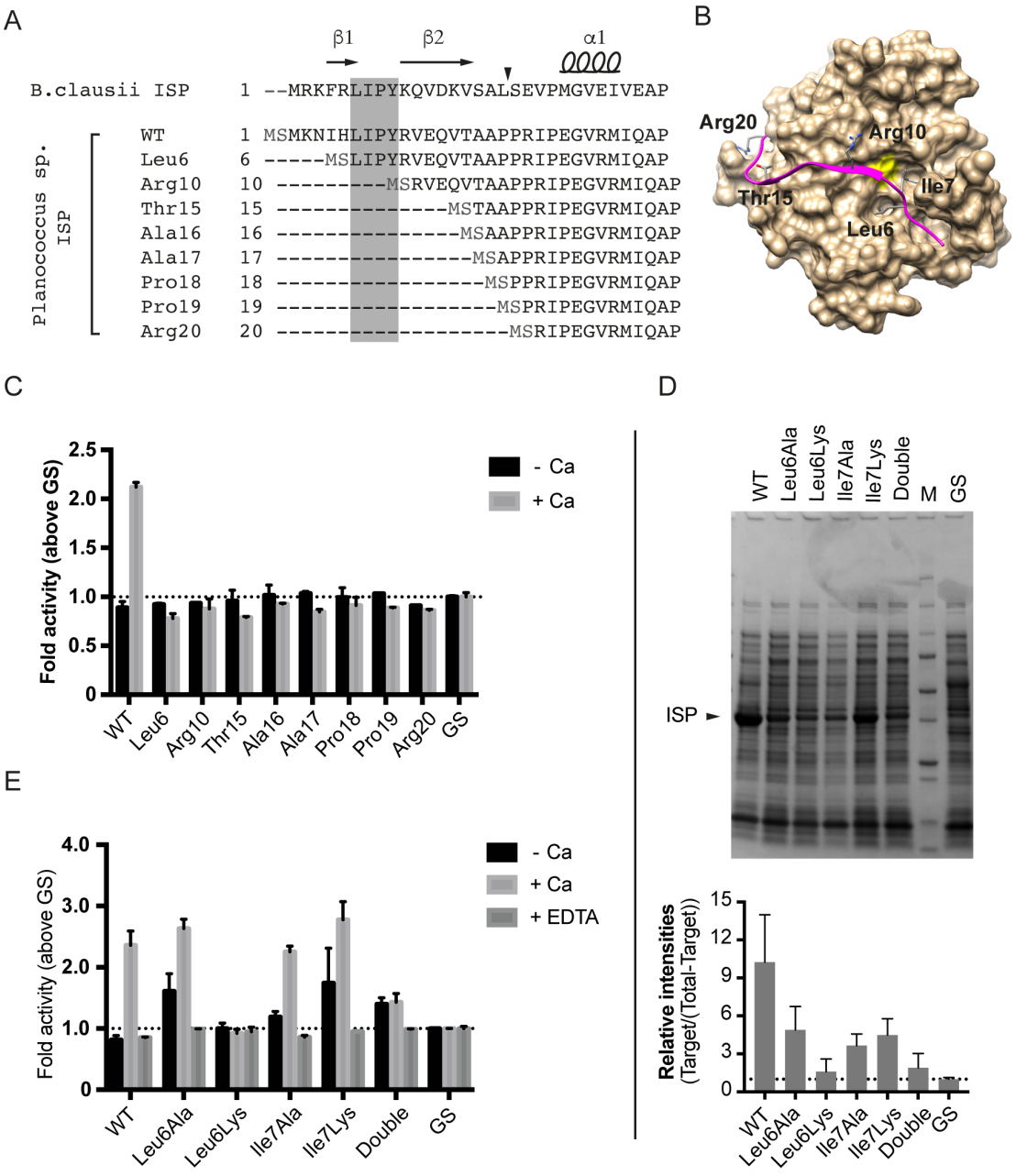

Figure 7. Truncating ISP at the N-termini. A) Alignment of the N-terminal region of ISP and the various truncated versions (indicated by starting residue given in three-letter ambiguity codes and their sequential numbers). A grey box indicates the LIPY/F-motif. Beta-strands (β), alpha helix (α), and arrow that points to the site of maturation refers to information from *B. clausii* ISP (PDB ID: 2X8J). Residues in light grey (MS) are added to the recombinant enzymes. **B)** Solvent-accessible surface of *Planococcus* sp. AW02J18 ISP is shown (tan, chain A) with the catalytic serine residue in yellow. The pro-peptide (residues 3-20) is shown in magenta, with the Leu6 and Ile7 residues

of the LIPY/F motif coloured by atom. Additionally, defining residues used in the truncation experiment are indicated. **C)** Cleared lysates containing wild-type ISP-His or truncated versions (p12 constructs), were screened for activity against BODIPY-FL-casein in the absence and presence of 1 mM $CaCl_2$ (+Ca) for 1 hour at 37 °C. Fluorescence was normalized to optical density of expression cultures, to account for any growth effects. Expression from empty vectors (GS) was used as background, and samples were calculated as fold above control. Error bars represent standard deviation between parallels in two experiments. **D)** A representative SDS-PAGE analysis of cleared lysates containing wild-type ISP-His (WT) or mutant versions (double, both Leu6 and Ile7 mutated to Ala). M, BioRad's Precision Plus Protein™ Dual Color Standard; and GS, extracts with empty vector. Arrow points to the recombinant ISP variants. The lower panel shows intensities of target ISP proteins relative to the target corrected total lane intensity. Intensity data arise from two independent experiments. **E)** Cleared lysates analysed as in B, in absence or presence of 1 mM $CaCl_2$ (Ca) or with 1 mM EDTA.

Supporting information

Supplementary item	Title	Citation
Table	Table S1: Primers used in this study	Table S1
Table	Table S2: Crystallization conditions	Table S2
Table	Table S3: Structure determination and refinement statistics	Table S3
Figure	Figure S1. Sequence alignment of *Planococcus* sp. AW02J18 ISP and homologs	Figure S1
Figure	Figure S2. Purity and activity of *Planococcus* sp. AW02J18 ISP and a catalytic mutant	Figure S2

Supplementary data

The LIPY/F-motif in an intracellular subtilisin protease is involved in inhibition

Gro Elin Kjæreng Bjerga[1], Øivind Larsen[1], Hasan Arsın[2], Adele Williamson[3], Antonio García-Moyano[1], Ingar Leiros[3], Pål Puntervoll[1]

[1] Uni Research, Center for applied biotechnology, Thormøhlens gate 55, 5006 Bergen, Norway

[2] University of Bergen, Department of biology, Thormøhlens gate 53, 5006 Bergen, Norway

[3] Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

Corresponding author:

Gro Elin Kjæreng Bjerga

Telephone: (+47) 55 58 44 92 (office)

e-mail: gro.bjerga@uni.no

http://uni.no/

Table S1: Primers used in this study

Primer name	Sequence (5'-3')	Purpose
T0034-S251A-F	CTGTCGGGTACCGCTATGGCTACGC	Mutagenesis
T0034-S251A-R	GCGTAGCCATAGCGGTACCCGACAG	Mutagenesis
T0034-L06-FX-F	ATATATGCTCTTCTAGTCTGATCCCGTATCGTGTGG	Truncation
T0034-R10-FX-F	ATATATGCTCTTCTAGTCGTGTGGAACAGGTTACCG	Truncation
T0034-T15-FX-F	ATATATGCTCTTCTAGTACCGCCGCCCCGCCGCG	Truncation
T0034-A16-FX-F	ATATATGCTCTTCTAGTGCCGCCCCGCCGCGTATTCC	Truncation
T0034-A17-FX-F	ATATATGCTCTTCTAGTGCCCCGCCGCGTATTCCG	Truncation
T0034-P18-FX-F	ATATATGCTCTTCTAGTCCGCCGCGTATTCCGG	Truncation
T0034-R20-FX-F	ATATATGCTCTTCTAGTCGTATTCCGGAAGGCGTCCG	Truncation
T0034-L6A-F	GAAAAACATTCATGCGATCCCGTATCG	Mutagenesis
T0034-L6A-R	CGATACGGGATCGCATGAATGTTTTTC	Mutagenesis
T0034-I7A-F	CATTCATCTGGCCCCGTATCGTG	Mutagenesis
T0034-I7A-R	CACGATACGGGGCCAGATGAATG	Mutagenesis
T0034-L6AI7A-F	ACATTCATGCGGCCCCGTATCG	Mutagenesis
T0034-L6AI7A-R	CGATACGGGGCCGCATGAATGT	Mutagenesis
T0034-L6K-F	GAAAAACATTCATAAGATCCCGTATCG	Mutagenesis
T0034-L6K-R	CGATACGGGATCTTATGAATGTTTTTC	Mutagenesis
T0034-I7K-F	CATTCATCTGAAACCGTATCGTG	Mutagenesis
T0034-I7K-R	CACGATACGGTTTCAGATGAATG	Mutagenesis

Table S2: Crystallization conditions

Crystal	Crystallization condition	Image
1	0.1 M Sodium Citrate pH 5, 0.51 % Ethylene glycol, 19.94 % PEG 4000	
2	0.17 M Na-K-Phosphate, 27.25 % PEG MME 2000	
3	0.1 M Sodium Citrate pH 5.5, 24.18 % PEG 6000	
4	0.1 M Sodium Citrate pH 4, 0.25 M Ammonium Acetate, 21.73 % PEG 1500	
5	0.09 M Ammonium Formate, 0.1 M Sodium Citrate, 20.01 % PEG 1500	
6	0.1 M Phosphate Citrate Buffer pH 4.5, 0.2 M LiSO4, 20 % PEG 1500	

Table S3: Structure determination and refinement statistics.

Resolution range (Å)	44.563-1.298
Completeness (%)	92.19
No. of reflection, working set	142740
No. of reflection, reference set	1992
Final R_{cryst}	13.04
Final R_{free}	15.03
MolProbity score	1.315
Clashscore	3.35
No. of non-H atoms	
Protein#	4548
Water	549
Other*	30
Total	5127
R.m.s. deviations	
Bonds (Å)	0.007
Angles (°)	0.962
Average B factors (Å^2)	
Overall	23.57
Protein	21.87
Water	37.20
Other*	32.01
Ramachandran plot (%)	
Preferred	96.87
Allowed	2.61
Outliers	0.52

#Two molecules pr. asymmetric unit

*Two molecules of Acetate, Na and triethylene glycol (Peg 3) were identified in the electron density and modeled. These occupy similar positions around the two protein molecules in the asymmetric unit.

Figure S1

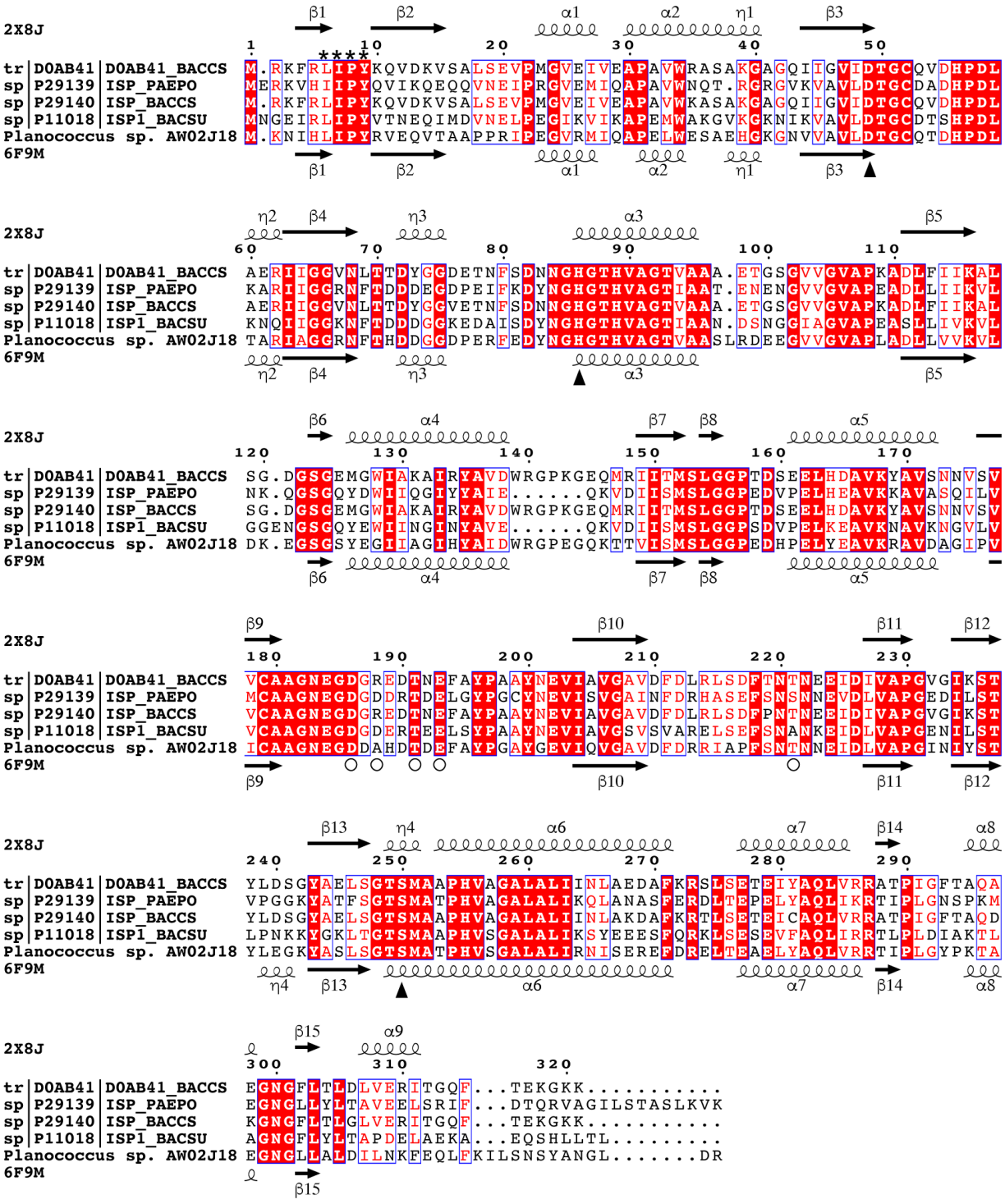
2X8J
tr|D0AB41|D0AB41_BACCS
sp|P29139|ISP_PAEPO
sp|P29140|ISP_BACCS
sp|P11018|ISP1_BACSU
Planococcus sp. AW02J18
6F9M

Figure S1. Sequence alignment of *Planococcus* sp. AW02J18 ISP and homologs. Alignment of ISPs from *B. clausii* (UniProt ID: D0AB41, P29140), *Paenibacillus polymyxa* (UniProt ID: P29139), *B. subtilis* sp. 168 (UniProt ID: P11018) and *Planococcus* sp. AW02J18. Secondary structure annotations are retrieved from *B. clausii* (PDB ID: 2X8J, chain A) and *Planococcus* sp. AW02J18 ISPs (PDB ID: 6F9M). Arrows and spirals indicate β-strands and α-helices, respectively. Asterisks point to the conserved LIPY/F motif. Conserved residues are shown in red, wherein red backgrounds indicate identical residues and red letters indicate similar residues. Positions indicated with filled triangles represent the catalytic triad (Asp49, His86, Ser251 in *Planococcus* sp. AW02J18). Aligned sequences were made with MAFFT and rendered by ESPript3.0 [47].

Figure S2

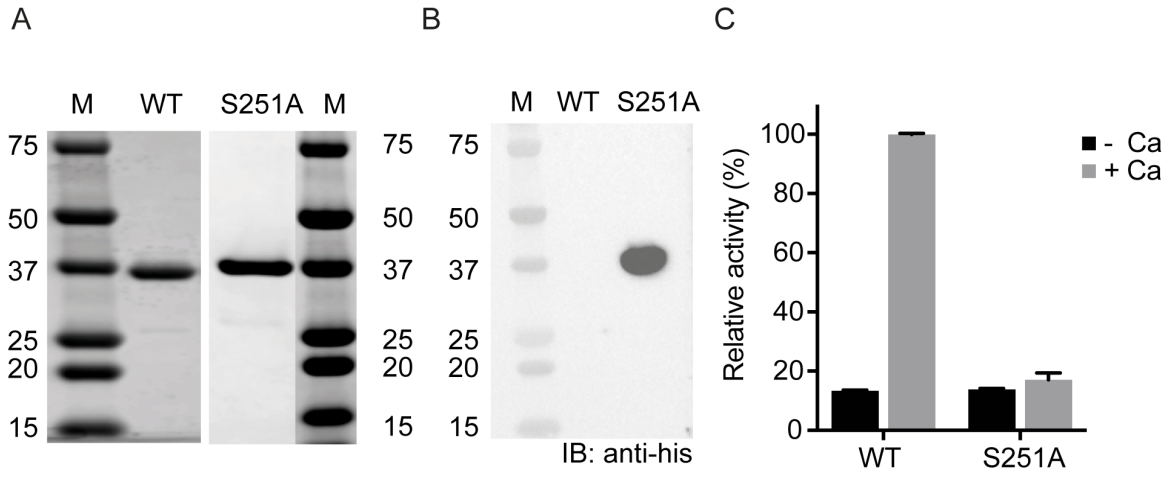

Figure S2. Purity and activity of *Planococcus* sp. AW02J18 and a catalytic mutant. A) Purity of matured His-ISP from *Planococcus* sp. AW02J18 (WT, p1-construct) and the catalytic ISP-His mutant (S251A, Ser251Ala; p12-construct) after purification as analyzed by SDS-PAGE. Both proteins are >95% pure, as identified by quantitative analysis. M, BioRad's Precision Plus Protein™ Dual Color Standard. **B)** Immunoblot analysis of the proteins in B. A standard immunoblot analysis using 1:3000 dilution of anti-His antibody was used to identify the presence of His-tags in proteins (as in A). M, as in A. **C)** The activities of matured ISP (WT) and the catalytic mutant (as in A) from crude extracts (both p12-constructs) were compared in the FITC-casein assay in the absence and presence of 1 mM $CaCl_2$ to verify that the serine mutation abolishes the catalytic power.