



RESEARCH ARTICLE

UPDATED The first structure in a family of peptidase inhibitors reveals an unusual Ig-like fold [v2; ref status: indexed, <http://f1000r.es/1nx>]

Daniel J Rigden¹, Qingping Xu^{2,3}, Yuanyuan Chang⁴, Ruth Y Eberhardt^{5,6}, Robert D Finn⁷, Neil D Rawlings^{5,6}

¹Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK

²Joint Center for Structural Genomics, La Jolla CA, 92037, USA

³Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park CA, 94025, USA

⁴Sandford-Burnham Institute, La Jolla CA, 92037, USA

⁵Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK

⁶European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK

⁷Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn VA, 20147, USA

v2 First published: 10 Jul 2013, 2:154 (doi: [10.12688/f1000research.2-154.v1](https://doi.org/10.12688/f1000research.2-154.v1))
Latest published: 23 Aug 2013, 2:154 (doi: [10.12688/f1000research.2-154.v2](https://doi.org/10.12688/f1000research.2-154.v2))

Abstract

We report the crystal structure solution of the Intracellular Protease Inhibitor (IPI) protein from *Bacillus subtilis*, which has been reported to be an inhibitor of the intracellular subtilisin Isp1 from the same organism. The structure of IPI is a variant of the all-beta, immunoglobulin (Ig) fold. It is possible that IPI is important for protein-protein interactions, of which inhibition of Isp1 is one. The intracellular nature of ISP is questioned, because an alternative ATG codon in the *ipi* gene would produce a protein with an N-terminal extension containing a signal peptide. It is possible that alternative initiation exists, producing either an intracellular inhibitor or a secreted form that may be associated with the cell surface. Homologues of the IPI protein from other species are multi-domain proteins, containing signal peptides and domains also associated with the bacterial cell-surface. The cysteine peptidase inhibitors chagasin and amoebiasin also have Ig-like folds, but their topology differs significantly from that of IPI, and they share no recent common ancestor. A model of IPI docked to Isp1 shows similarities to other subtilisin:inhibitor complexes, particularly where the inhibitor interacts with the peptidase active site.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
UPDATED		
version 2		report
published 23 Aug 2013	↑	↑
version 1		
published 10 Jul 2013	report	report

1 **Patricia C Weber**, Implex LLC USA

2 **Xiayang Qiu**, Pfizer Inc USA

Discuss this article

Comments (0)

Corresponding author: Neil D Rawlings (ndr@sanger.ac.uk)

How to cite this article: Rigden DJ, Xu Q, Chang Y *et al.* **The first structure in a family of peptidase inhibitors reveals an unusual Ig-like fold** [v2; ref status: indexed, <http://f1000r.es/1nx>] *F1000Research* 2013, 2:154 (doi: [10.12688/f1000research.2-154.v2](https://doi.org/10.12688/f1000research.2-154.v2))

Copyright: © 2013 Rigden DJ *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This work was supported by the Wellcome Trust (grant number WT077044/Z/05/Z: RYE and NDR); Howard Hughes Medical Institute (RDF); National Institutes of Health (NIH) Protein Structure Initiative (U54 GM094586-03: QX and R01GM101457: YC); National Science Foundation (IIS-0646708 and IIS-1153617: NDR, RYE, QX, YC). Portions of this research were conducted under the auspices of the Joint Center for Structural Genomics (JCSG: NIH U54 GM09458) and structural determination was carried out SSRL Molecular Biology Program at the Stanford Synchrotron Radiation Lightsource (SSRL) supported by the Department of Energy and NIH (P41 GM103393: QX and YC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of DOE, NSF, NIGMS, NCRR or NIH. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Competing interests: No competing interests were disclosed.

First published: 10 Jul 2013, 2:154 (doi: [10.12688/f1000research.2-154.v1](https://doi.org/10.12688/f1000research.2-154.v1))

First indexed: 29 Jul 2013, 2:154 (doi: [10.12688/f1000research.2-154.v1](https://doi.org/10.12688/f1000research.2-154.v1))

UPDATED Changes from Version 1

We thank the referees for their reports. Below is a summary of the changes made to version 1 in response to Dr. Webbers' report.

- It seems logical to the authors that the intracellular peptidase should be introduced first which then leads to the requirement for an inhibitor, so we would prefer not to change the order of paragraphs in the introduction.
- The NMR structure paper is preliminary, describing only the chemical shift. We have modified the text to indicate this, and added a section and a supplementary table to describe the close agreement between the solved tertiary structure and the beta-strands from the NMR structure.
- We have realised that the separation of the stereo panels in Fig 1A, as reproduced in the pdf version, is a long way short of the recommended 7cm. This has caused problems for some readers including Referee 2. We have prepared a new figure with greater separation between the stereo panels, and it is now displayed in a larger size, spanning two columns, in order to aid legibility.

See referee responses

Introduction

The Isp1 peptidase (also known as IspA) is the major intracellular peptidase in *Bacillus subtilis*, accounting for more than 80% of the intracellular azocasein- or azocollagen-degrading activity¹. It is a subtilisin homologue and a member of peptidase family S8. It is unclear what the physiological role of Isp1 is; it has been shown not to be important during sporulation where the majority of intracellular proteolysis occurs, even though intracellular proteolysis was correspondingly low when a defective mutant of Isp1 was created². Potential physiological substrates that have been identified include elongation factor Tu (cleavage within the GTP binding domain, PF00009) and ClpC, a negative regulator of genetic competence (with one cleavage occurring in a disordered region and another in a coiled-coil region)³. The peptidase is synthesized as a precursor with an N-terminal, 18-residue propeptide that blocks the active site. Release of the propeptide is probably by Isp1 itself, inter-molecularly⁴. The structures of both the precursor and the mature peptidase have been solved and show that a proline residue (Pro8) introduces a kink in the backbone that shifts the scissile bond in the propeptide away from the catalytic serine. Once the propeptide is released, the peptidase active site forms correctly⁴.

The activity of a powerful endopeptidase within the cell must be controlled in order to prevent unwanted proteolysis of cellular proteins, and in 1986 Nishino *et al.*⁵ described an inhibitor of Isp1 from *Bacillus subtilis* known as Intracellular Protease Inhibitor (IPI). This inhibitor was effective against a number of subtilisin homologues, using casein as a substrate following pre-incubation of peptidase and inhibitor, but was ineffective against cysteine, aspartic and metallopeptidases. A K_i for inhibition of the Isp1 peptidase was estimated to be in the order of 10^{-9} M⁶. The gene for the peptidase has been cloned⁷, and the deduced protein sequence showed, at the time, no similarity to any other protein. There are now 149 homologues in the Pfam family PF12690 (BsuPI). The family was included in the MEROPS database (as I22⁸), but subsequently removed, because it was not clear whether the proteins in the family

were inhibitors or competing substrates. In particular, the inhibition of intracellular peptidases Isp2 and Isp3 was "repressed" in the presence of the substrate casein⁶, which suggests the activity of competing substrates.

Structural similarity to other known peptidase inhibitors might argue for reinstatement of the family in the MEROPS database. However, most peptidase inhibitors are secreted and intracellular peptidase inhibitors are rare. The only examples are cystatins A and B from family I25, which inhibit cysteine peptidases⁹; calpastatin, which inhibits the intracellular peptidase calpain¹⁰; chagasin from the zooflagellate *Leishmania* (family I42), which also inhibits cysteine peptidases¹¹; three intracellular coagulation inhibitors from the horseshoe crab *Tachypleus* that are serpins from family I4¹²; and pinA from family I24, which is an inhibitor of the ATP-dependent serine endopeptidase Lon, but is of unknown structure¹³. Despite the inhibitors being intracellular, the known target peptidases are all extracellular, with the exceptions of calpain and endopeptidase Lon. Given the paucity of known intracellular peptidase inhibitors, it would not be a surprise if the fold of IPI were different from any known inhibitor structures, especially the secreted inhibitors which are stabilized by disulfide bridges, because intracellular proteins lack disulfides.

Most serine peptidase inhibitors act as if they were super-substrates, binding so tightly to the active site that they are either not cleaved, or if cleavage occurs then the fragments are not released from the peptidase. There is a bond, known as the reactive bond, which occupies the peptidase active site with residues either side occupying the S1 and S1' binding pockets (in the nomenclature of Schechter & Berger, 1968¹⁴). This inhibitory mechanism is known as the standard or the Laskowski mechanism¹⁵. The residues that form the reactive bond will vary from inhibitor to inhibitor, according to the specificity of the peptidase that is inhibited. The chymotrypsin-like specificity of the Isp1 peptidase implies that the P1 residue in the reactive bond of the intracellular inhibitor should be a hydrophobic residue.

A preliminary NMR study, assigning chemical shifts to the *B. subtilis* intracellular peptidase inhibitor has been published¹⁶, which identified beta strands. We report the complete tertiary structure of the intracellular peptidase inhibitor from *Bacillus subtilis*.

Methods

Structural determination

The American Type Culture Collection (ATCC) provided the genomic DNA used to clone *ipi* (ATCC Number: 23857D-5). Protein production and crystallization of IPI was carried out by standard JCSG protocols¹⁷. The crystal was obtained using the vapor diffusion method in a sitting drop format where sitting drops composed of 100 nl protein solution mixed with 100 nl crystallization solution were equilibrated against a 50 μ l reservoir at 293 K. The crystallization reagent consisted of 48.5% polyethylene glycol 600, 0.1M CHES pH 9.7. Data were collected at wavelengths corresponding to the inflection and high energy remote of a selenium MAD (multi-wavelength anomalous dispersion) experiment at 100 K using a MARCCD 325 detector (Rayonix) at Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9_2. Data processing were carried out using XDS¹⁸ and the statistics are presented in [Supplementary table 1](#). The structure was determined by the MAD

method using programs SHELX¹⁹ and autoSHARP²⁰, and refinement was carried out using REFMAC²¹. The structure was validated using the JCSG Quality Control server (<http://smb.slac.stanford.edu/jcsg/QC>). Atomic coordinates and experimental structure factors to 2.6 Å resolution (PDB code: 3ISY) have been deposited in the Protein Data Bank (www.pdb.org)²².

Bioinformatics

Sequence conservation among homologues of IPI was mapped onto the crystal structure using ConSurf (<http://consurf.tau.ac.il/>)²³. The results were visualised with PyMOL (<http://www.pymol.org/>), which was also used for structural figures.

A homology model of *B. subtilis* intracellular proteinase, IspA, was created at the Swiss-model server²⁴ using the structure of processed, active intracellular protease from *Bacillus clausii* (PDB code 2XRM), around 50% identical to IspA, as the sole template. The high degree of sequence identity with the target and the small number of insertions and deletions between target and template (one six residue deletion and a single insertion of one residue, both readily accommodated) assured a high quality model. Additionally, the insertions and deletions lie distant from the catalytic site which was the main region of interest.

The new structure of IPI was docked to the model of IspA using three different webservers, GRAMM-X (<http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/>)²⁵, ZDOCK (<http://zdock.umassmed.edu/>)²⁶ and ClusPro 2.0 (<http://cluspro.bu.edu/>)²⁷.

Results and discussion

Structure description

The crystal structure of an intracellular proteinase inhibitor (IPI, gene locus BSU11130) from *Bacillus subtilis* was determined to 2.6 Å resolution by the MAD method. Data collection, model and refinement statistics are summarized in [Supplementary table 1](#). The final model includes one molecule (residues 3–119), one tetraethylene glycol and 21 water molecules in the asymmetric unit. The structure is mainly composed of nine beta strands. Gly0 (which remained at the N-terminus after cleavage of the expression/purification tag), Met1 and Glu2 were disordered and not modeled. All the side chains except for Glu17, Lys60 and Lys97 were fully modeled. The Matthews coefficient (V_M^{28}) is 3.2 Å³ Da⁻¹ and the estimated solvent content is 60.9%. The Ramachandran plot produced by MolProbity²⁹ shows that 98.0% of the residues are in favoured regions, with no outliers. Two residues, Lys 86 and Glu 87, are flagged as having unusual Ramachandran plot properties by the PDB validation report although not with Molprobity. Electron density for them strongly supports their modelled conformations.

Validation report for PDB: 3ISY

1 Data File

<http://dx.doi.org/10.6084/m9.figshare.732739>

The structure of the IPI protein is similar to the immunoglobulin (Ig) fold. This common fold consists of a beta-sandwich formed of seven strands in two sheets with a Greek-key topology and proteins with

this fold are involved in a variety of functions, including cell-cell recognition, cell-surface receptors, muscle structure and the immune system. The limits of the B-strands seen in the structure coincide well with those determined by NMR¹⁶ ([Supplementary table 2](#)). At the strand level there is only a single disagreement: residues 62–65 were assigned as B-6 by NMR but in the crystal structure form an extended region that lacks the backbone hydrogen bonding characteristics to be defined by us as a strand. The structure of IPI shows an unusually extended loop on one edge of the Ig fold, seen on the right hand side in [Figure 1A](#), resulting in a wedge-like or prismatic overall shape. As [Figure 1B](#) shows, this unusual feature is conserved at the sequence level among members of the I22 family, suggesting that it may harbour a functional site.

From a comparison using the Dali website (http://ekhidna.biocenter.helsinki.fi/dali_server/)³⁰ the protein with the closest structural similarity is the RbmA protein from *Vibrio cholerae* (Z-score 10.0, Root-Mean-Square Deviation (RMSD) 2.9 Å, residues 38–152), which is one of the three major protein components of the biofilm matrix important for cell-to-cell contacts³¹. The A chain from coagulation factor XIII³² is also structurally similar, with z-scores in the range of 9.4–9.9 depending on the source species. The structural similarity with IPI is over residues 518–629 for human factor XIII, which is a domain from the transglutaminase, C-terminal Ig-like domain family (PF00927). Factor XIII is a transglutaminase important for stabilizing fibrin clots by cross-linking chains with isopeptide bonds³². These similarities probably reflect only the all-beta nature of the structures rather than any common physiological functions.

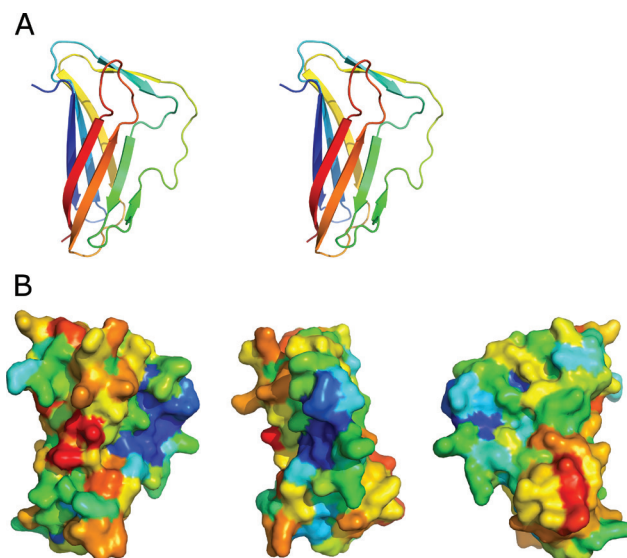


Figure 1. Overall structure of Intracellular Protease Inhibitor (IPI) reveal a wedge shape with a conserved surface. A) Stereo cartoon representation of IPI coloured from blue to red, N- to C-terminus. Protruding loops on the right hand side form a wedge shape. **B)** Sequence conservation amongst known homologues mapped to the protein surface with red indicating high conservation and blue low. The left panel shows IPI in the same orientation as in **A**), the other two show the results of successive 90 degree rotations about a vertical axis so that the wedge is seen end-on in the centre panel and the right panel shows the opposite face to the left panel.

The folds of chagasin and amoebiasin (both members of inhibitor family I42) have also been described as Ig-like^{33,34}. From a Dali pairwise comparison, the structures of IPI and chagasin (PDB code 2NNR) are similar but distant (Z-score 2.2, RMSD 4.2 Å; structural alignment over 110 residues). The structures of IPI and amoebiasin (PDB: 3M86) are also similar (Z-score 3.0, RMSD 4.5 Å, over 111 residues). Figure 2 shows a secondary structure comparison of IPI, chagasin and amoebiasin. Chagasin and amoebiasin have similar topologies, but the order of the beta strands differs significantly between these two I42 family members and IPI. This implies that IPI is not closely related to the I42 family, and that the similar folds may have been acquired independently by convergent evolution.

Docking

In order to explore the hypothesis that IPI is an inhibitor of the intracellular protease IspA, structure docking was carried out. As an experimental structure for *B. subtilis* was not available, a homology model was created. The ~50% sequence identity shared with available templates assures a high quality model: at this level of similarity an RMSD error of around 1 Å is expected³⁵. Since it is often difficult to pick out the true solution from the multitude of poses typically resulting from a docking experiment, we applied three cross-checks to improve confidence in our results. First, we required that the interface of the inhibitor with protease involve the conspicuously conserved wedge structure (Figure 1B), which is indicated as the functional site of the inhibitor structure. We chose not to incorporate this information in the docking runs, although several servers allow this, in order to reserve it as an independent check on the results. Secondly, we required that a solution sterically

block the catalytic site cleft since this, rather than allosteric mechanisms, is the predominant mode of protease inhibition. Thirdly, we sought docking poses that were independently proposed by different algorithms; although simple structural compatibility is a principal criterion for all docking servers, they differ in their scoring functions so that poses jointly flagged by different methods may be considered more reliable.

Comparison of the top 10 results from each of the docking servers used highlighted a family of poses satisfying our criteria. This contained solutions ranked 1, 4 and 5 from ZDOCK, the second solution from GRAMM-X and the third from ClusPro (Supplementary figure 1).

As shown in Supplementary figure 1, these poses occlude the catalytic site cleft of the enzyme by inserting the conserved IPI wedge. They vary slightly in their rotation about an axis running along the cleft, and to a minor extent in their translation along it, but can be considered as a cluster of solutions.

It is well known that inhibitors typically inhibit cognate enzymes through interaction of extended regions³⁶ and, furthermore, that different families of inhibitors acting on the same class of enzyme can exhibit highly similar, but convergently evolved interaction features^{36–38}. Remarkably, we find that the region at the edge of the IPI wedge structure could interact with subtilisin in a manner strongly reminiscent of inhibition modes of other subtilisin family inhibitors. As Figure 3A shows, in a representative of the cluster of docking solutions, solution 5 from ZDOCK, residues 62–68 of IPI superim-

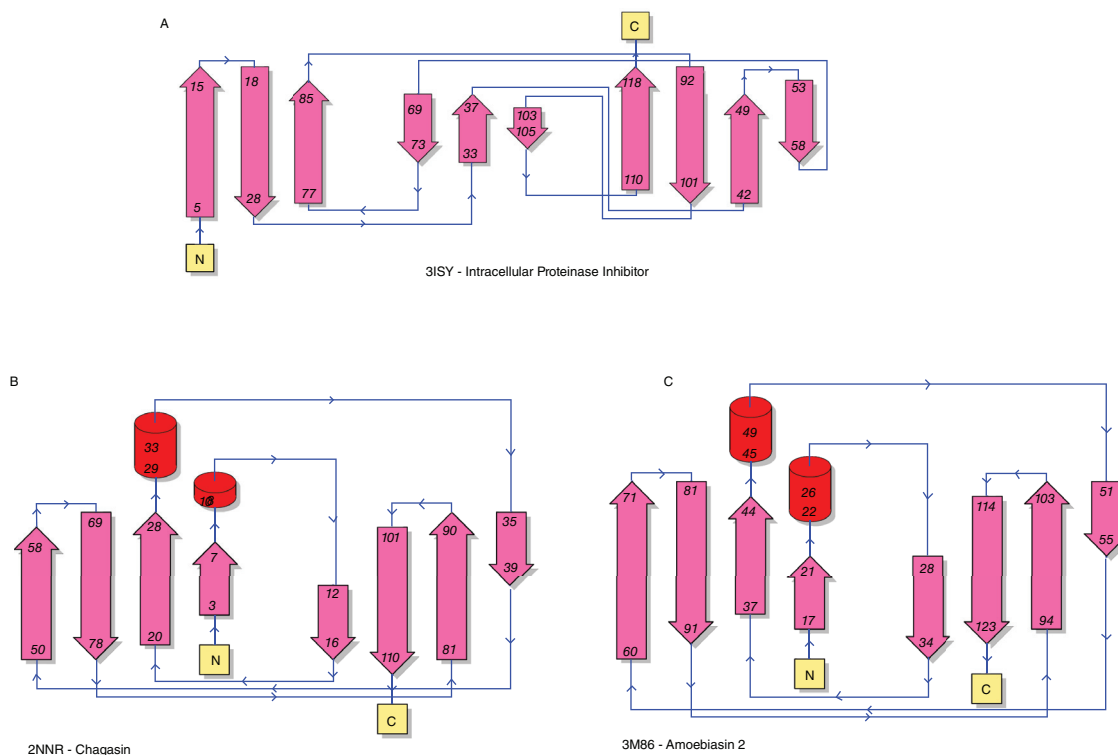


Figure 2. Secondary structure topologies of Intracellular Protease Inhibitor (IPI) and other peptidase inhibitors with an Ig-like fold. Topology diagrams are from PDBSum⁴³. The labels indicate the PDB accession number and the protein name.

pose well on strongly similar inhibitory stretches from five distinct inhibitor classes, especially considering the potential for small conformational changes on binding and the fact that a model was used for the docking. Incidentally, the experimentally observed propeptide⁴, while lying across the catalytic site cleft in the same direction, is different to both IPI and the other four inhibitors in Figure 4.

The comparison between the matching inhibitory regions also revealed suggestive sequence similarities with IPI. As shown in Figure 3B, amino acids in IPI abide by trends evident in the other inhibitors. Most notably three of four previous inhibitors have a central Thr residue, also present in IPI. The following position is the P1 residue. Ordinarily, substrate specificity would dictate conservation here among inhibitors but, as Figure 3B shows, the sequence varies surprisingly at this position, including a Glu in 3QTL, similar to the Gln borne at this position by IPI. The previous two positions and the last are conserved hydrophobic in all sequences; again IPI conforms to the consensus.

Sequence and structure similarities

IPI homologues are found in both bacteria and archaea, but no homologues have been detected in eukaryotes. All bacterial homologues are from Gram-positive bacteria (Firmicutes), with the exception of

homologues from the Gemmatimonadetes bacterium *Gemmatimonas aurantiaca*, and the Gammaproteobacteria *Shewanella oneidensis* and *Kangiella koreensis*. All three of these organisms also possess subtilisin homologues. A subtilisin homologue from *Gemmatimonas aurantiaca* (UniProtKB C1AD56) is predicted to be intracellular.

All of the 30 archaean homologues are from species of the family *Halobacteriaceae*, which includes high salt tolerant species. The archaeans with completely sequenced genomes that possess an IPI homologue also possess at least one subtilisin homologue, which in halophilic archaeon DL31 and *Haloterrigena turkmenica* include proteins predicted to be intracellular. In none of the species with completely sequenced genomes is the gene for the IPI homologue in close proximity to a gene for a subtilisin homologue, and so they are not part of the same operon or regulon. In total, there are 33 species of bacteria and archaea that possess both an IPI homologue and a subtilisin homologue that is predicted to be intracellular (from absence of a signal peptide). There are 43 species from which an IPI homologue is known, but an intracellular subtilisin homologue is unknown, but of these 43 species only four have had their genomes completely sequenced (*Bacillus selenitireducens*, *Bacillus macauensis*, *Clostridium* sp. BNL1100 and *Selenomonas* sp. oral taxon 137).

It is relatively common for inhibitors to occur as multiple domains within the same protein. Figure 4 shows examples of domain architectures seen in proteins bearing an IPI-like domain. Most of the 109 proteins with IPI-like domains share the simple domain architecture of the *Bacillus subtilis* protein. The presence of a signal peptide shows that many members of the family are secreted rather than intracellular proteins. Even among the proteins that contain

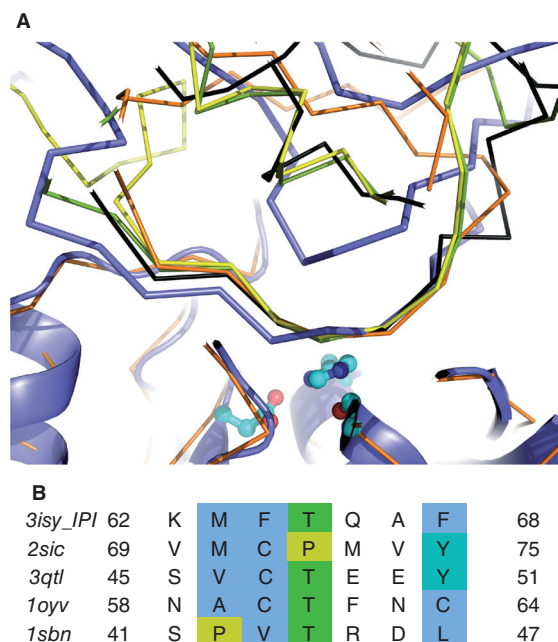


Figure 3. Comparison of Intracellular Protease Inhibitor (IPI) docked to a model of *B. subtilis* subtilisin with other subtilisin-inhibitor complexes. **A)** Inhibitors are drawn as ribbon and the *B. subtilis* subtilisin model as cartoon, with catalytic triad shown as ball-and-stick. The IPI complex is shown in mauve, others, after superposition on the enzymes structure, as green (2SIC; *Streptomyces* Subtilisin inhibitor), yellow (3QTL; Kazal inhibitor), black (1OYV; plant inhibitor class) or orange (1SBN; eglin inhibitor class). The inhibitory region lies immediately above the catalytic site in this view. **B)** Comparison of inhibitory region sequences from IPI and four distinct inhibitor classes. The sequences are derived from the structures shown in **A)** and coloured in Jalview (<http://www.jalview.org/>)⁴⁴ using the ClustalX scheme⁴⁵.

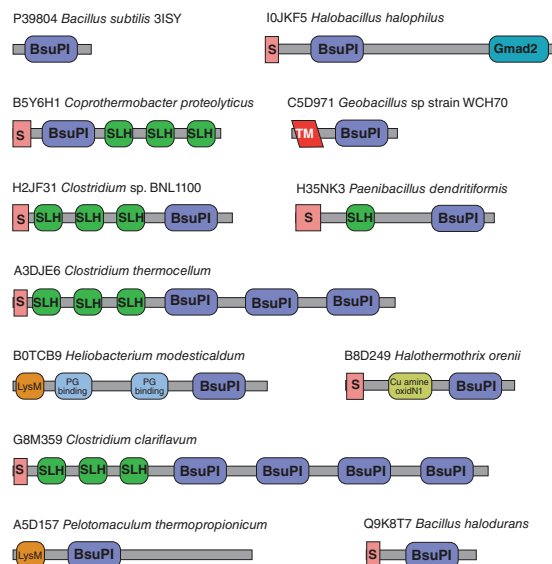


Figure 4. Domain architectures of proteins containing a domain homologous to Intracellular Protease Inhibitor (IPI). Key to domains: BsuPI, *Bacillus subtilis* protein inhibitor-like; S, signal peptide; Gmad2, immunoglobulin-like domain of bacterial spore germination; SLH, S-layer homology domain; TM, transmembrane helix; PG binding, peptidoglycan-binding domain; Cu amine oxidase N1, copper amine oxidase N-terminal domain; LysM, lysin motif domain.

only the IPI-like domain, a large number also possess signal peptides, including homologues from *Bacillus* species. One homologue from a *Geobacillus* species is predicted to be a type II membrane protein with an N-terminal transmembrane domain. The longest protein to include an IPI-like domain is the 585-residue Chte0880 protein from *Clostridium thermocellum*, which has three repeats each composed of an S-layer homology (SLH) domain followed by an IPI-like domain. The presence of SLH domains indicates that the protein binds to the proteoglycan of the cell wall.

IPI itself was initially characterized as intracellular. It was purified by lysing cells with lysozyme and its N-terminal sequence determined by Edman degradation⁵. Later, when the gene was sequenced⁷, no signal peptide was detected in the coding sequence. Now that the full genome for *Bacillus subtilis* strain 168 has been sequenced³⁹, it is clear that a possible alternative initiating methionine exists for the *ipi* gene and the coding sequence could be extended at the N-terminus, adding an extra in-frame 33 residues (MKRLVMLLPVLLIGCG-KDEQTEPDKEVSGG). The predicted protein sequences of IPI from *B. subtilis* strains BSn5⁴⁰ and QB928⁴¹ include this N-terminal extension. When the extended sequence was submitted to the SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) a signal peptide was predicted with cleavage at the Thr23-Glu bond. The N-terminal sequence determined by Nishino *et al.* (1986)⁵ was Glu34, with the assumption that Met33 was the initiating methionine. Signal peptidase 1 is unlikely to cleave the Met33-Glu bond, because of the hydrophilic nature of the region Glu24-Gly32. This may mean that alternative initiation exists for this protein; a secreted form starting at Met1 and an intracellular form starting at Met33. It is worth adding that even if a majority of IPI is in fact secreted, a protease inhibitor function is still physiologically plausible since most subtilisins are extracellular.

Is IPI truly a peptidase inhibitor?

The nearest structural neighbours of IPI, including the RbmA protein and coagulation factor XIII, are involved in protein-protein interactions. This may indicate that rather than functioning as an inhibitor, IPI is binding the peptidase in some non-specific way. Such an interaction might explain the peculiar results obtained with the substrate casein, which apparently “repressed” inhibition of the peptidase⁶. Binding to an intracellular subtilisin may therefore not represent the primary physiological function of IPI. However, there are similarities between the IPI protein and chagasin, and not only in terms of structure. Chagasin inhibition of cysteine peptidases is tight but reversible³³. If IPI inhibition of subtilisins were also reversible, then the repression seen with casein would be explained. IPI inhibition would not be via the classical Laskowski mechanism, in which a reactive site bond *permanently* occupies the peptidase active site, but access to the active site would instead be *reversibly* physically blocked.

Without further kinetics studies it is not possible to state categorically that IPI definitely is a peptidase inhibitor. However, the tertiary structure and docking predictions illustrate how it could feasibly inhibit its cognate enzyme. That hypothetical mode bears strong

similarities to those seen for other well-characterised, unrelated serine peptidase inhibitor families such as the Kazal and Kunitz groups in agreement with a substantial literature on convergent evolution in peptidase inhibition^{36,38}. It is intriguing that IPI may exemplify a second intracellular peptidase inhibitor family, after the chagasin and amoebiasin group, with an Ig-type fold. However, structural similarity between chagasin and IPI is weak and the two groups clearly share no recent common ancestor. We hope the resolution of the structure encourages others to further characterize IPI and proteins bearing IPI-like domains to further probe their functions.

Author contributions

DJ Rigden, ND Rawlings and Y Chang wrote the text. DJ Rigden performed the molecular modelling. Q Xu refined the crystal structure. DJ Rigden, RY Eberhardt and RD Finn prepared figures. All authors revised and approved the final version of the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

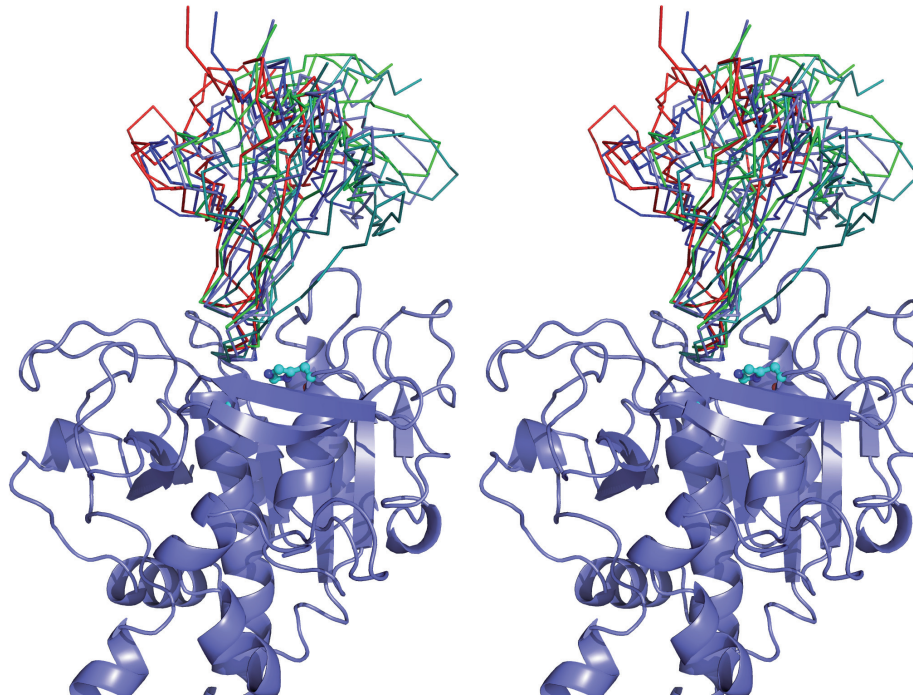
This work was supported by the Wellcome Trust (grant number WT077044/Z/05/Z: RYE and NDR); Howard Hughes Medical Institute (RDF); National Institutes of Health (NIH) Protein Structure Initiative (U54 GM094586-03: QX and R01GM101457: YC); National Science Foundation (IIS-0646708 and IIS-1153617: NDR, RYE, QX, YC). Portions of this research were conducted under the auspices of the Joint Center for Structural Genomics (JCSG: NIH U54 GM09458) and structural determination was carried out SSRL Molecular Biology Program at the Stanford Synchrotron Radiation Lightsource (SSRL) supported by the Department of Energy and NIH (P41 GM103393: QX and YC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of DOE, NSF, NIGMS, NCRR or NIH.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We are grateful to the Sanford Burnham Medical Research Institute and UC San Diego for hosting the DUF annotation jamboree in June 2013 which allowed the authors to collaborate on this work. We would like to thank all the participants of this workshop for their intellectual contributions to this work: L. Aravind, Herbert L. Axelrod, Alex Bateman, Penny Coghill, Debanu Das, Adam Godzik, William Hwang, Lukasz Jaroszewski, Alexey Murzin, Padmaja Natarajan, Marco Punta, Mayya Sedova, Anna Sheydina, John Wooley. We thank the members of the JCSG high-throughput structural biology pipeline for their contribution to this work.

Supplementary materials



Supplementary figure 1. Stereo view of the superposition of docking poses from ZDOCK (numbers 1, 4 and 5, coloured shades of blue), GRAMM-X (number 2, coloured red) and ClusPro (number 3, coloured green) showing broadly similar modes of inhibitor interaction with the enzyme, shown as cartoon below with catalytic triad drawn in ball-and-stick.

Supplementary table 1. Crystallographic data and refinement statistics for IPI, BSU11130 (PDB code 3ISY). Values in parentheses are for the highest resolution shell. Space group: I 4₁ 2 2. Unit cell parameters: a = 73.6, c = 132.9 Å.

Data collection	λ1 MAD-Se	λ2 MAD-Se
Wavelength (Å)	0.91162	0.97934
Resolution range (Å)	38.0–2.61 (2.75–2.61)	41.0–2.71 (2.86–2.71)
No. of observations	41,044	36,556
No. of unique reflections	5864	5264
Completeness (%)	100 (100)	100 (99.9)
Mean I/σ (I)	18.7 (2.6)	19.0 (3.5)
R _{merge} on I† (%)	9.4 (87.4)	10.6 (63.0)
R _{meas} on I‡ (%)	10.1 (94.2)	11.4 (67.9)
Model and refinement statistics		
Resolution range (Å)	37.96–2.61	
No. of reflections (total)	5851§	
No. of reflections (test)	266	
Completeness (%)	99.9	
Data set used in refinement	λ1	
Cutoff criteria	F >0	
R _{cryst} ¶	0.203	
R _{free} ¶	0.243	
Stereochemical parameters		
Restraints (r.m.s.d. observed)		
Bond angles (°)	1.521	
Bond lengths (Å)	0.014	
Average isotropic B value†† (Å ²)	45.520	
ESU‡‡ based on R _{free} (Å)	0.273	
Protein residues/atoms	117/967	
Waters/solvent molecules	22	

† $R_{\text{merge}} = \frac{\sum hkl \sum i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum i I_i(hkl)}$. Highest resolution shell in parentheses

‡ $R_{\text{meas}} = \frac{\sum hkl [N(N-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum i I_i(hkl)}$ ⁴⁶

§ Typically, the number of unique reflections used in refinement is slightly less than the total number that were integrated and scaled. Reflections are excluded owing to negative intensities and rounding errors in the resolution limits and unit-cell parameters

¶ $R_{\text{cryst}} = \frac{\sum hkl | |F_{\text{obs}}| - |F_{\text{calc}}| |}{\sum hkl |F_{\text{obs}}|}$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. R_{free} is the same as R_{cryst} but for 4.9% of the total reflections chosen at random and omitted from refinement

†† This value represents the total B that includes TLS and residual B components

‡‡ Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994;⁴⁷)

Supplementary table 2. Comparison of the secondary structural elements assigned by Li et al.¹⁶ by NMR and those observed in the crystal structure (Figure 2A).

Secondary structure element (Li et al. ¹⁶)	Residue range (NMR)	Residue range (crystal structure)
β-1	6–12	5–15
β-2	18–27	18–28
β-3	32–37	33–37
β-4	43–48	42–49
β-5	55–57	53–58
β-6	62–65	not seen
β-7	68–73	69–73
β-8	78–86	77–85
β-9	93–100	92–101
β-10	103–104	103–105
β-11	109–118	110–118

References

1. Burnett TJ, Shankweiler GW, Hageman JH: **Activation of intracellular serine proteinase in *Bacillus subtilis* cells during sporulation.** *J Bacteriol.* 1986; **165**(1): 139–45.
[PubMed Abstract](#) | [Free Full Text](#)
2. Band L, Henner DJ, Ruppen M: **Construction and properties of an intracellular serine protease mutant of *Bacillus subtilis*.** *J Bacteriol.* 1987; **169**(1): 444–6.
[PubMed Abstract](#) | [Free Full Text](#)
3. Lee AY, Goo Park S, Kho CW, *et al.*: **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.* 2004; **4**(11): 3437–45.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Gamble M, Künze G, Dodson EJ, *et al.*: **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.* 2011; **108**(9): 3536–41.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Nishino T, Shimizu Y, Fukuhara K, *et al.*: **Isolation and characterization of a proteinaceous protease inhibitor from *Bacillus subtilis*.** *Agric Biol Chem.* 1986; **50**(12): 3059–3064.
[Publisher Full Text](#)
6. Nishino T, Murao S: **Interaction of proteinaceous protease inhibitor of *Bacillus subtilis* with intracellular proteases from the same strain.** *Agric Biol Chem.* 1986; **50**(12): 3065–3070.
[Publisher Full Text](#)
7. Shiga Y, Yamagata H, Udaka S: **Characterization of the gene encoding an intracellular proteinase inhibitor of *Bacillus subtilis* and its role in regulation of the major intracellular proteinase.** *J Bacteriol.* 1993; **175**(22): 7130–7.
[PubMed Abstract](#) | [Free Full Text](#)
8. Rawlings ND, Tolle DP, Barrett AJ: **Evolutionary families of peptidase inhibitors.** *Biochem J.* 2004; **378**(Pt 3): 705–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Lenarcic B, Krizaj I, Zunec P, *et al.*: **Differences in specificity for the interactions of stefins A, B and D with cysteine proteinases.** *FEBS Lett.* 1996; **395**(2–3): 113–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Todd B, Moore D, Deivanayagam CC, *et al.*: **A structural model for the inhibition of calpain by calpastatin: crystal structures of the native domain VI of calpain and its complexes with calpastatin peptide and a small molecule inhibitor.** *J Mol Biol.* 2003; **328**(1): 131–46.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Wang SX, Pandey KC, Scharfstein J, *et al.*: **The structure of chagasin in complex with a cysteine protease clarifies the binding mode and evolution of an inhibitor family.** *Structure.* 2007; **15**(5): 535–43.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Miura Y, Kawabata S, Iwanaga S: **A *Limulus* intracellular coagulation inhibitor with characteristics of the serpin superfamily. Purification, characterization, and cDNA cloning.** *J Biol Chem.* 1994; **269**(1): 542–7.
[PubMed Abstract](#)
13. Skorupski K, Tomaschewski J, Rügter W, *et al.*: **A bacteriophage T4 gene which functions to inhibit *Escherichia coli* Lon protease.** *J Bacteriol.* 1988; **170**(7): 3016–24.
[PubMed Abstract](#) | [Free Full Text](#)
14. Schechter I, Berger A: **On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain.** *Biochem Biophys Res Commun.* 1968; **32**(5): 898–902.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Laskowski M Jr: **Protein inhibitors of serine proteinases—mechanism and classification.** *Adv Exp Med Biol.* 1986; **199**: 1–17.
[PubMed Abstract](#)
16. Li Q, Chen AS, Gayen S, *et al.*: **¹H, ¹³C and ¹⁵N chemical shift assignments for an intracellular proteinase inhibitor of *Bacillus subtilis*.** *Biomol NMR Assign.* 2013; **7**(2): 129–32.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Elsliger MA, Deacon AM, Godzik A, *et al.*: **The JCSG high-throughput structural biology pipeline.** *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2010; **66**(Pt 10): 1137–42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Kabsch W: **XDS.** *Acta Crystallogr Sect D Biol Crystallogr.* 2010; **66**(Pt 2): 125–32.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Sheldrick GM: **A short history of SHELX.** *Acta Crystallogr A.* 2008; **64**(Pt 1): 112–122.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Smart OS, Womack TO, Flensburg C: **Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER.** *Acta Crystallogr D Biol Crystallogr.* 2012; **68**(Pt 4): 368–80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Winn MD, Isupov MN, Murshudov GN: **Use of TLS parameters to model anisotropic displacements in macromolecular refinement.** *Acta Crystallogr D Biol Crystallogr.* 2001; **57**(Pt 1): 122–33.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Rose PW, Bi C, Bluhm WF, *et al.*: **The RCSB Protein Data Bank: new resources for research and education.** *Nucleic Acids Res.* 2013; **41**(Database issue): D475–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Ashkenazy H, Erez E, Martz E, *et al.*: **ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.** *Nucleic Acids Res.* 2010; **38**(Web Server issue): W529–W533.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Bordoli L, Schwede T: **Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal.** *Methods Mol Biol.* 2012; **857**: 107–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
25. Tovchigrechko A, Vakser IA: **GRAMM-X public web server for protein-protein docking.** *Nucleic Acids Res.* 2006; **34**(Web Server issue): W310–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Pierce BG, Hourai Y, Weng Z: **Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library.** *PLoS One.* 2011; **6**(9): e24657.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Kozakov D, Hall DR, Beglov D, *et al.*: **Achieving reliability and high accuracy in automated protein docking: Cluspro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19.** *Proteins. Structure, Function, and Bioinformatics.* 2010; **78**(15): 3124–3130.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Matthews BW: **Solvent content of protein crystals.** *J Mol Biol.* 1968; **33**(2): 491–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Chen VB, Arendall WB 3rd, Headd JJ, *et al.*: **MolProbity: all-atom structure validation for macromolecular crystallography.** *Acta Crystallogr D Biol Crystallogr.* 2010; **66**(Pt 1): 12–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Holm L, Sander C: **Mapping the protein universe.** *Science.* 1996; **273**(5275): 595–603.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Giglio KM, Fong JC, Yildiz FH, *et al.*: **Structural basis for biofilm formation via the *Vibrio cholerae* matrix protein RbmA.** *J Bacteriol.* 2013; **195**(14): 3277–86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Yee VC, Pedersen LC, Le Trong I, *et al.*: **Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII.** *Proc Natl Acad Sci U S A.* 1994; **91**(15): 7296–300.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Smith BO, Picken NC, Westrop GD, *et al.*: **The structure of *Leishmania mexicana* ICP provides evidence for convergent evolution of cysteine peptidase inhibitors.** *J Biol Chem.* 2006; **281**(9): 5821–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Casados-Vázquez LE, Lara-González S, Bribea LG: **Crystal structure of the cysteine protease inhibitor 2 from *Entamoeba histolytica*: functional convergence of a common protein fold.** *Gene.* 2011; **471**(1–2): 45–52.
[PubMed Abstract](#) | [Publisher Full Text](#)
35. Chothia C, Lesk AM: **The relation between the divergence of sequence and structure in proteins.** *EMBO J.* 1986; **5**(4): 823–6.
[PubMed Abstract](#) | [Free Full Text](#)
36. Tyndall JD, Fairlie DP: **Conformational homogeneity in molecular recognition by proteolytic enzymes.** *J Mol Recognit.* 1999; **12**(6): 363–70.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Bode W, Huber R: **Structural basis of the endoproteinase-protein inhibitor interaction.** *Biochim Biophys Acta.* 2000; **1477**(1–2): 241–52.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Otlewski J, Jelen F, Zakrzewska M, *et al.*: **The many faces of protease-protein inhibitor interaction.** *EMBO J.* 2005; **24**(7): 1303–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Kunst F, Ogasawara N, Moszer I, *et al.*: **The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.** *Nature.* 1997; **390**(6657): 249–56.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Deng Y, Zhu Y, Wang P, *et al.*: **Complete genome sequence of *Bacillus subtilis* BSn5, an endophytic bacterium of *Amorphophallus konjac* with antimicrobial activity for the plant pathogen *Erwinia carotovora* subsp. *carotovora*.** *J Bacteriol.* 2011; **193**(8): 2070–2071.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Yu CS, Yim KY, Tsui SK, *et al.*: **Complete genome sequence of *Bacillus subtilis* strain QB928, a strain widely used in *B. subtilis* genetic studies.** *J Bacteriol.* 2012; **194**(22): 6308–6309.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Monteiro AC, Abrahamson M, Lima AP, *et al.*: **Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in *Trypanosoma cruzi*.** *J Cell Sci.* 2001; **114**(Pt 21): 3933–42.
[PubMed Abstract](#)

43. Laskowski RA: **PDBsum new things**. *Nucleic Acids Res.* 2009; **37**(Database issue): D355–D359.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Waterhouse AM, Procter JB, Martin DM, *et al.*: **Jalview Version 2—a multiple sequence alignment editor and analysis workbench**. *Bioinformatics.* 2009; **25**(9): 1189–1191.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Larkin MA, Blackshields G, Brown NP, *et al.*: **Clustal W and Clustal X version 2.0**. *Bioinformatics.* 2007; **23**(21): 2947–2948.
[PubMed Abstract](#) | [Publisher Full Text](#)
46. Diederichs K, Karplus PA: **Improved R-factors for diffraction data analysis in macromolecular crystallography**. *Nature Struct Biol.* 1997; **4**(4): 269–75.
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Cruickshank DW: **Remarks about protein structure precision**. *Acta Crystallogr D Biol Crystallogr.* 1999; **55**(Pt 3): 583–601.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Referee Status:



Version 2

Referee Report 22 January 2014

doi:10.5256/f1000research.2157.r3278



Patricia C Weber

Imiplex LLC, Bristol, PA, USA

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 09 September 2013

doi:10.5256/f1000research.2157.r1733



Xiayang Qiu

Department of Structural Biology, Pfizer Inc, Groton, CT, USA

Looks good.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 29 July 2013

doi:10.5256/f1000research.1790.r1142



Xiayang Qiu

Department of Structural Biology, Pfizer Inc, Groton, CT, USA

This article reports the first crystal structure of the IPI protein from *B. subtilis*, a family of protease inhibitor proteins with homologues across many different species. Although being assigned to the general IG-like fold, the structure shows a distinct topology, hence could also be regarded as a novel fold and a

significant contribution to the protease inhibitor field. Additional studies are needed to understand the functional aspects of this new protein and structural scaffold.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 29 July 2013

doi:10.5256/f1000research.1790.r1144



Patricia C Weber

Imiplex LLC, Bristol, PA, USA

Thorough discussion of the inhibition mechanism and the protein's possible biological role(s) make this manuscript of interest to a wide audience.

Minor changes to improve clarity include delaying (or removing) discussions of ISP in the Abstract because ISP is not the central subject of the manuscript, expanding comments on the NMR structure and possibly citing similarities beyond the common identification of beta strands, and fixing Figure 1A where greater separation between the stereo images is needed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.
