A New Autocatalytic Activation Mechanism for Cysteine Proteases Revealed by Prevotella intermedia Interpain A*

Noemí Mallorqui-Fernández1, Surya P. Manandhar1,1, Goretti Mallorqui-Fernández1, Isabel Usón3, Katarzyna Wawrzonek4, Tomasz Kanyka1, Maria Solà2,2, Ida B. Thøgersen**, Jan J. Enghild**, Jan Potempa6,6, and F. Xavier Gomis-Ruth5,5

From the 1Department of Biologia Estructural and the 6Institució Catalana de Recerca i Studis Avançats, Institut de Biologia Molecular de Barcelona, Jordi Girona 18-26, 08034, Barcelona, Spain, the 2Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, the 3Department of Microbiology, Faculty of Biotechnology, Jagiellonian University, 30-387 Krakow, Poland, and the 5Department of Molecular Biology, Science Park, University of Århus, Gustav Wieds Vej 10C, 8000 Århus C, Denmark

Prevotella intermedia is a major periodontopathogen contributing to human gingivitis and periodontitis. Such pathogens release proteases as virulence factors that cause deterrence of host defenses and tissue destruction. A new cysteine protease from the cysteine-histidine-dyad class, interpain A, was studied in itszymogenogenic and self-processed mature forms. The latter consists of a bivalved moiety made up by two subdomains. In the structure of a catalytic cysteine-to-alaninezymogen variant, the right subdomain interacts with an unusual prodomain, thus contributing to latency. Unlike the catalytic cysteine residue, already in its competent conformation in thezymogen, the catalytic histidine is swung out from its active conformation and trapped in a cage shaped by a backing helix, azymogenichairpin, and a latency flap in thezymogen. Dramatic rearrangement ofto 20 Å of these elements triggered by a tryptophan switch occurs during activation and accounts for a new activation mechanism for proteolytic enzymes. These findings can be extrapolated to related potentially pathogenic cysteine proteases such as Streptococcus pyogenes SpeB and Porphyromonas gingivalis periodontain.

Periodontal disease (PD) affects the tissues that surround and support the teeth and may lead to loosening and eventual loss of teeth if untreated. It is caused by bacteria and affects mildly 90% and severely 10% of the population worldwide (1, 2). In addition, symptoms of PD appear in a series of systemic diseases due to its inflammatory and infective character (2, 3). Present day treatment of PD includes the mechanical cleansing of the affected area and is efficient in general. However, it is costly, time consuming, and painful and needs frequent repetition. In addition, it may entail the indiscriminate usage of antibiotics, which contributes to the spread of antibiotic-resistant strains (2, 4). Consequently, there is a need for innovative and specific therapeutic approaches against PD.

Prevotella intermedia is a major bacterial periodontal pathogen in humans together with Porphyromonas gingivalis, among others (5, 6). Such bacteria colonize the gingival crevice and produce virulence factors that cause disease. Bacterial infection leads to the bacterial secretion or induction of host overproduction of proteolytic enzymes such as bacterial collagenases, matrix metalloproteases, and serine and cysteine proteases (CPs) (2, 7, 8). These proteases destroy host tissue and compromise host defenses. In addition, proteases may give rise to fibrinolytic activity and inactive components of the blood-coagulation cascade such as the protease inhibitors, α1-proteinase inhibitor and α2-macroglobulin. Proteolysis further covers alimentary requirements, because most of bacterial nutrition is obtained from degraded periodontal tissue and tissue fluid (9).

Most studies on the bacterial proteolytic armamentarium in PD have been performed with P. gingivalis (9). In contrast, the factors governing P. intermedia infection, a black-pigmented Gram-negative obligate anaerobic non-motile rod bacterium, are poorly understood (7). In humans, Prevotella sp. have frequently been recovered from subgingival plaque in patients suffering from acute necrotizing gingivitis, pregnancy gingivitis, and adult periodontitis (10). In addition, Prevotella species easily acquire resistance toward antibiotics, which hampers their...
Activation of \textit{P. intermedia} InpA Cysteine Protease

elimination (11). A deep molecular knowledge of how infection and resistance occur is crucial for the development of alternative treatments. In \textit{P. intermedia}, several proteases have been described, among them trypsin-like serine proteases, a dipeptidyl peptidase IV and CPs (12–14), but no structural studies are available that could help in understanding their particular mode of action or facilitate the design of specific drugs. The structures of some clan-A papain-like CPs (according to the MEROPS data base (15)) from other infective bacteria are known, namely those of staphopain A and B from \textit{Staphylococcus aureus} (16, 17), the avirulence putative peptidase AvrPphB from \textit{Pseudomonas syringae} (18), and streptopain (alias streptococcal pyrogenic exotoxin B and SpeB) and IdeS endoproteinase, both from \textit{Streptococcus pyogenes} (19, 20). Together with other bacterial enzymes such as bleomycin hydrolase from \textit{Lactococcus lactis} and a calpain-like enzyme from \textit{P. gingivalis}, they may be among the ancestral enzymes that gave rise to the 20 families currently identified within this clan of proteases (15, 21). They display a relatively broad substrate specificity but are restricted to a small group of related bacterial species or are even limited to a single species, thus constituting attractive targets for the selective design of antibiotics (22). All these proteases have been identified as or proposed to be secreted virulence factors that elicit nutrient generation, evasion of the adaptive immune system response through inactivation of immunoglobulins, or release of bacterial proteins from the cell surface (23).

For more than 60 years, SpeB, a protein secreted by \textit{Streptococcus pyogenes} (24), was considered a unique CP, unrelated to plant papains or vertebrate cathepsins, and the founding member of family C10 within clan CA (15). A recent analysis of bacterial genomes identified genes encoding potential SpeB orthologues in several species, predominantly \textit{Bacteroidetes} (31). Interestingly, two forms of genes are common, either short orthologues encoding an SpeB-like protein with an N-terminal pro-domain and a catalytic CP domain or large orthologues with an additional large C-terminal extension, which shares no similarity with any other proteins sequenced. The latter orthologues are present in bacteria that are involved in pathogenicity of periodontal disease in humans. With this in mind, a genome search within \textit{P. intermedia} 17 was undertaken, and three open reading frames potentially encoding CPs were identified (22).

We studied the first of these potential proteases, interpain A (InpA), encoded by locus PIN0048. This gene encodes a long pro-domain (Ala1–Asn111, see Fig. 1), a catalytic domain (Val112–Pro359) and a further 465 C-terminal residues hereafter termed pro-cd-InpA and pro-cd-InpA C154A, respectively. We further analyzed the three-dimensional structures of a major fragment of pro-cd-InpA C154A and of the wt catalytic domain, cd-InpA. Unexpectedly, these studies have uncovered a hitherto undescribed activation mechanism for cysteine proteases and helped us to understand a family of virulence factors produced by human pathogens.

**EXPERIMENTAL PROCEDURES**

Expression, Mutant Construction, and Purification of Pro-interpain A—Genomic DNA of \textit{P. intermedia} was extracted from strain ATCC 25611. The structural gene region of InpA comprising the pro-domain and the catalytic domain, pro-cd-InpA, was amplified by PCR using forward primer 5’-ATGCCCATG-GCAAAACGCACAAAGGAACAG-3’ with an NcoI recognition site and reverse primer 5’-ATGCTGAGTGGTTT-TCCGTAACACCC-3’ with an XhoI recognition site. Because the NcoI site encompasses the ATG start codon, two bases (CA) were introduced into the forward primers immediately after the NcoI site for in-frame translation of the target protein. This genetic manipulation inserted a methionine before the N-terminal alanine residue of InpA. In addition, the reverse primer introduced two additional codons (CTC GAG) for a leucine and a glutamate following the C-terminal proline residue of pro-cd-InpA. The PCR product was purified and cloned into the NcoI/XhoI site of pET24d (+) expression vector (Novagen), which provides the coding sequence for a C-terminal hexahistidine tag (His6). The recombinant plasmid was transformed into \textit{Escherichia coli} strain BL21(DE3) pLysS under the control of the T7 promoter. The wt construct was used to produce mutation C154A using overlap extension PCR (26). The correctness of the constructs was verified by double-stranded DNA sequencing.

Protein production and purification were essentially the same for the wt and the mutant protein. Cells freshly transfected with the expression plasmid were grown at 37 °C to an optical density (\(A_{600}\)) of 0.7–0.8 in 1 liter of Luria-Bertani medium supplemented with 2% glucose and kanamycin sulfate (50 \(\mu\)g/ml). The culture was induced with isopropyl-1-thio-\(\beta\)-d-galactopyranoside to a final concentration of 0.1 \(\text{mM}\) and further incubated at 26 °C for 2–3 h for protein production. Cells were harvested, washed with phosphate-buffered saline buffer, and resuspended in binding buffer A (20 \(\text{mM}\) sodium phosphate, 500 \(\text{mM}\) NaCl, 20 \(\text{mM}\) imidazole, pH 7.4) supplemented with 1.5 \(\text{mM}\) 4’,4’-dithiodipyridine (a reversible CP inhibitor), 6 \(\text{mM}\) 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate, 10 \(\mu\)M phenylmethylsulfonyl fluoride, 1 \(\text{mM}\) HgCl2, and 1 \(\text{mM}\) 1,4-dithio-\(\beta\)-d-threitol (DTT). The latter compounds were added to prevent protein aggregation and autolysis. Cells were lysed by ultrasonication on ice for ~5’, and cell lysates were cleared by centrifugation, filtered through 0.45-\(\mu\)m pore size filters and mixed with Fast Flow nickel-nitrotriacetic acid Sepharose resin slurry (2 ml) previously equilibrated with buffer B (buffer A supplemented with 1 \(\text{mM}\) HgCl2). After 1 h at room temperature (or overnight at 4 °C), the slurry was poured into a column and first washed in buffer B until the baseline (\(A_{280}\)) was stable and then in 2 ml of buffer B supplemented with 60 \(\text{mM}\) imidazole. The protein was eluted stepwise with buffer B further containing 100, 200, 300, 400, and 500 \(\text{mM}\) imidazole, respectively. The fractions collected were analyzed by SDS-PAGE, and those containing a single band attributable to the target protein were pooled, dialyzed at 4 °C overnight against 10 \(\text{mM}\) Tris-HCl, 1 \(\text{mM}\) HgCl2, pH 7.5, passed through 0.45-\(\mu\)m filters, and concen-
trated using a Centricon-10 device (Millipore). The last purification step comprised ion-exchange chromatography (Amerham Biosciences) with a Mono Q column equilibrated with 20 mM Tris-HCl, 1 mM HgCl₂, pH 7.5.

**Activity Assay**—Activity was determined with the fluorogenic substrate di-tertbutyl dicarbonate-Val-Leu-Lys-aminomethylcoumarin. Briefly, recombinant pro-cd-InpA protein was activated at 37 °C in buffer C (0.1 M Tris-HCl, 5 mM EDTA, pH 7.5, freshly supplemented with 2 mM DTT). The fluorogenic reaction was started by adding substrate (10 mM; final concentration in the reaction mixture, 250 μM) and the release of aminomethylcoumarin was recorded by measuring the increase in fluorescence using a micro-titer plate reader.

**Autocatalytic Assay**—A total of 100 μg of pro-cd-InpA protein, alone or with 0.7 μg of active cd-InpA protein, was preincubated at 37 °C in buffer C (0.1 M Tris-HCl, 1 mM HgCl₂, 2 mM DTT, pH 7.6). The autocatalytic reaction was initiated by diluting the sample with buffer D (buffer C but with 5 mM EDTA instead of 1 mM HgCl₂) at 37 °C (final pro-cd-InpA and cd-InpA concentrations were 10 and 0.1 μM, respectively). Aliquots were taken at distinct time intervals and mixed with E-64 inhibitor (N-[N-[[trans-carboxyoxiran-2-carbonyl]-L-leucyl]lamagatine] to stop the reaction. At the same time intervals, samples of the incubation mixture were assessed for activity against the above fluorogenic substrate, and the initial rate of substrate turnover was determined. As a negative control, the same experiments were carried out using buffer C. To ascertain whether pro-cd-InpA autoactivation was an intra- or an intermolecular process, thezymogen was incubated as described above at 10, 2, and 0.4 μM, respectively, with samples withdrawn from the same activation reaction mixture at the mentioned time intervals.

**Processing of Pro-cd-InpA C154A by wt cd-InpA**—Pro-cd-InpA C154A was tested as a substrate for wt cd-InpA in a reaction mixture containing 0.1 μM of the latter and 10 μM of the former protein in buffer D at 37 °C. Aliquots of 10 μl were withdrawn from the reaction mixture at distinct time intervals, and the reaction was quenched by addition of E-64. Results were analyzed by 12% SDS-PAGE.

**Generation of N-terminally Truncated Pro-cd-InpA C154A**—Pro-cd-InpA C154A (25 mg/ml) in 20 mM Tris-HCl, pH 7.6, was incubated with 1.7 μg of DTT-activated wt cd-InpA overnight at 21 °C. The reaction was terminated by addition of E-64 to 100 μM final concentration, and the protein was purified by ion-exchange chromatography employing a NaCl gradient. Fractions containing the N-terminally truncated 36-kDa form of pro-cd-InpA C154A were pooled, concentrated, and dialyzed against a buffer suitable for protein crystallization. N-terminal sequencing, mass spectrometry, and Western blot analyses revealed that this protein variant (ΔN1pro-cd-InpA C154A) encompassed residues Ala³⁹-Pro³⁵⁹ plus the C-terminal dipeptide asparagine-glutamate dipeptide but was lacking the His₉ tag.

**N-terminal Sequence Analysis**—Wild-type pro-cd-InpA, the C154A mutant protein, their truncated variants, as well as their cleavage products were analyzed by 12% SDS-PAGE and transferred to polyvinylidene difluoride membranes. Membranes were stained with 0.2% Amido Black and subjected to Edman degradation using a Procise 494-HT protein sequencer.

**Activation of P. intermedia InpA Cysteine Protease**

**Crystallization and Data Collection and Processing**—ΔN1pro-cd-InpA C154A was crystallized from sitting drops containing protein (22 mg/ml) and 10% polyethylene glycol 3000, 0.2 M magnesium chloride, 0.1 M sodium cacodylate, pH 6.5. Wt cd-InpA protein comprising residues Val¹¹²-Pro³⁵⁹ plus the C-terminal dipeptide was crystallized from drops comprising protein solution (16 mg/ml) and 28% polyethylene glycol 4000, 0.2 M magnesium chloride, 30% xylitol, 0.1 M Tris-HCl, pH 8.5. Diffraction data were collected at 110 K at the European Synchrotron Radiation Facility (Grenoble, France) beam line ID29 using an ADSC Q315 CCD area detector. ΔN1pro-cd-InpA C154A crystals diffracted beyond 1.5 Å resolution and belonged to space group C2 with one molecule per asymmetric unit. Wt cd-InpA crystals diffracted to 3.2 Å, belonged to the tetragonal space group P₄₁₂₁, and contained two molecules per asymmetric unit. Diffraction data were indexed and integrated with the program XDS (27) and scaled and reduced with SCALA within the CCP4 suite (28). Statistics on data collection and processing are presented in Table 1. Wt cd-InpA crystals diffracted very weakly. This led to a high value for the merge indicator Rmerge but to an acceptable Rmerge value due to the almost 8-fold average multiplicity of the data. In any case, these data were accurate enough to yield valid structural information. In the case of the ΔN1pro-cd-InpA C154A crystals, diffraction data were strong and of excellent quality, leading to low values for both Rmerge and Rmerge (see Table 1).

**Structure Solution and Analysis**—The structure of ΔN1pro-cd-InpA C154A was solved with program PHASER (29) using all diffraction data, and the coordinates of S. pyogenes pro-SpeB protein (Protein data bank (PDB) access codes 1pvj and 1dkf (19)) were used as a searching model. A refined final solution was found comprising 170.8, 60.0, and 294.0 for the Rmerge, Rfree, and Rmerge values, respectively. The coordinates of the latter and 10% polyethylene glycol 3000, 0.2 M magnesium chloride, 30% xylitol, 0.1 M Tris-HCl, pH 8.5. Diffraction data were collected at 110 K at the European Synchrotron Radiation Facility (Grenoble, France) beam line ID29 using an ADSC Q315 CCD area detector. ΔN1pro-cd-InpA C154A crystals diffracted beyond 1.5 Å resolution and belonged to space group C2 with one molecule per asymmetric unit. Wt cd-InpA crystals diffracted to 3.2 Å, belonged to the tetragonal space group P₄₁₂₁, and contained two molecules per asymmetric unit. Diffraction data were indexed and integrated with the program XDS (27) and scaled and reduced with SCALA within the CCP4 suite (28). Statistics on data collection and processing are presented in Table 1. Wt cd-InpA crystals diffracted very weakly. This led to a high value for the merge indicator Rmerge but to an acceptable Rmerge value due to the almost 8-fold average multiplicity of the data. In any case, these data were accurate enough to yield valid structural information. In the case of the ΔN1pro-cd-InpA C154A crystals, diffraction data were strong and of excellent quality, leading to low values for both Rmerge and Rmerge (see Table 1).

**Structure Solution and Analysis**—The structure of ΔN1pro-cd-InpA C154A was solved with program PHASER (29) using all diffraction data, and the coordinates of S. pyogenes pro-SpeB protein (Protein data bank (PDB) access codes 1pvj and 1dkf (19)) were used as a searching model. A refined final solution was found comprising 170.8, 60.0, and 294.0 for the Rmerge, Rfree, and Rmerge values, respectively. The coordinates of the latter and 10% polyethylene glycol 3000, 0.2 M magnesium chloride, 30% xylitol, 0.1 M Tris-HCl, pH 8.5. Diffraction data were collected at 110 K at the European Synchrotron Radiation Facility (Grenoble, France) beam line ID29 using an ADSC Q315 CCD area detector. ΔN1pro-cd-InpA C154A crystals diffracted beyond 1.5 Å resolution and belonged to space group C2 with one molecule per asymmetric unit. Wt cd-InpA crystals diffracted to 3.2 Å, belonged to the tetragonal space group P₄₁₂₁, and contained two molecules per asymmetric unit. Diffraction data were indexed and integrated with the program XDS (27) and scaled and reduced with SCALA within the CCP4 suite (28). Statistics on data collection and processing are presented in Table 1. Wt cd-InpA crystals diffracted very weakly. This led to a high value for the merge indicator Rmerge but to an acceptable Rmerge value due to the almost 8-fold average multiplicity of the data. In any case, these data were accurate enough to yield valid structural information. In the case of the ΔN1pro-cd-InpA C154A crystals, diffraction data were strong and of excellent quality, leading to low values for both Rmerge and Rmerge (see Table 1).

**Structure Solution and Analysis**—The structure of ΔN1pro-cd-InpA C154A was solved with program PHASER (29) using all diffraction data, and the coordinates of S. pyogenes pro-SpeB protein (Protein data bank (PDB) access codes 1pvj and 1dkf (19)) were used as a searching model. A refined final solution was found comprising 170.8, 60.0, and 294.0 for the Rmerge, Rfree, and Rmerge values, respectively. The coordinates of the latter and 10% polyethylene glycol 3000, 0.2 M magnesium chloride, 30% xylitol, 0.1 M Tris-HCl, pH 8.5. Diffraction data were collected at 110 K at the European Synchrotron Radiation Facility (Grenoble, France) beam line ID29 using an ADSC Q315 CCD area detector. ΔN1pro-cd-InpA C154A crystals diffracted beyond 1.5 Å resolution and belonged to space group C2 with one molecule per asymmetric unit. Wt cd-InpA crystals diffracted to 3.2 Å, belonged to the tetragonal space group P₄₁₂₁, and contained two molecules per asymmetric unit. Diffraction data were indexed and integrated with the program XDS (27) and scaled and reduced with SCALA within the CCP4 suite (28). Statistics on data collection and processing are presented in Table 1. Wt cd-InpA crystals diffracted very weakly. This led to a high value for the merge indicator Rmerge but to an acceptable Rmerge value due to the almost 8-fold average multiplicity of the data. In any case, these data were accurate enough to yield valid structural information. In the case of the ΔN1pro-cd-InpA C154A crystals, diffraction data were strong and of excellent quality, leading to low values for both Rmerge and Rmerge (see Table 1).
prised all residues from Ala 39 to Pro 359 (Fig. 1) except Ser295–Gln301.

The structure of wt cd-InpA was solved with program AMoRe (32) using the coordinates corresponding to residues Ala121–Pro359 from /H9004

N1pro-cd-InpA C154A and structure-factor amplitudes in the range 15–3.5 Å. These calculations unambiguously confirmed P41212 as the correct space group and a unique solution was found at 48.5, 88.4, 223.7, 0.1172, 0.5977, and 0.1152 (/H9251

/H9252

/H9253

x

y

z; refined values after rigid-body refinement; see Ref. 32) and 41.7, 88.6, 115.6, 0.1940, 0.0880, and 0.7122 for each of the two molecules A and B in the asymmetric unit, respectively, with a combined score CCF/crystallographic R factor, according to Ref. 32, of 54.4%/41.4%.

The appropriately rotated and translated coordinates were subjected to rigid-body and positional refinement applying strong non-crystallographic-symmetry restraints with CNS v. 1.2 (33). Despite the weakness and low resolution of the wt cd-InpA diffraction data, the resulting electron density maps clearly disclosed the entire polypeptide chain of the mature protease moiety. It contained an unambiguous trace for the first eight residues (Val112–Tyr120) of the polypeptide chain that had not been included in the search model, a proof of concept for data quality which ruled out model bias. Careful model building alternated with crystallographic refinement under application of strong non-crystallographic-symmetry restraints with programs CNS and REFMAC5 (at the final stages). The final wt cd-InpA model comprised residues Val112–Gly357 for molecule A and Val112–Pro359 plus two residues from the C-terminal tag (termed Leu360 and Glu361) for molecule B. These two molecules were almost equivalent in practice. Accordingly, under “Results and Discussion” we consider molecule A unless otherwise stated. Table 1 provides statistics on the final refinement steps and parameters of the quality of the resulting models.

Miscellaneous—The figures were prepared with programs TURBO-Frodo, SETOR (34), and MOLMOL (35). Structures were superimposed with TURBO-Frodo. Bioinformatic amino acid sequence similarity searches were undertaken within MEROPS data base and with the PSI-BLAST server (www.ncbi.nlm.nih.gov/blast). Structural similarity searches were performed with program DALI and secondary structure predictions with program JPRED. Close contacts and interaction surfaces (with a probe radius of 1.4 Å) were calculated with CNS taking the half of the total surface buried at the interface. The final coordinates of ΔN1pro-cd-InpA C154A and wt cd-InpA have been deposited with the Protein Data Bank at the Research Collaboratory for Structural Bioinformatics with access codes 3bb7 and 3bba.

RESULTS AND DISCUSSION

Protein Purification and Characterization—Pro-cd-InpA and pro-cd-InpA C154A were overexpressed as 40-kDa proteins and purified to homogeneity. The wt zymogen was readily converted into the fully processed mature 27-kDa catalytic domain during purification so that the zymogenic form could only be obtained if reversible CP inhibitors were included during homogenization of bacterial cells and purification (Fig. 2, A–C). Subsequent inhibition release resulted in time-dependent autocatalytic processing of the zymogen with the concurrent release of activity (Fig. 2, C and E). Processing and activity
Crystallographic data collection and refinement

| Dataset | ΔN1pro-cd-InpA (C154A) | cd-InpA |
|---------|-------------------------|---------|
|          |                         |         |
|          |                         |         |
| Space group | C2 | P4/2,2 |
| Cell constants (a, b, c, in Å; β in °) | 128.87, 38.81, 78.14, 127.3 | 129.24, 129.24, 81.97, 90.0 |
| Wavelength (Å) | 0.9763 | 0.9760 |
| No. of measurements/unique reflections | 344,943/49,377 | 90,923/13,017 |
| Resolution range (Å) (outermost shell) | 39.07–1.50 (1.58–1.50) | 47.25–3.22/3.40–3.22 |
| Completeness (%) | 99.5 (97.6) | 99.0 (98.2) |
| Rrel = Rmerge | 0.086 (0.448)/0.033 (0.180) | 0.352 (0.595)/0.120 (0.330) |
| Rmerge | 0.080 (0.409) | 0.329 (0.897) |
| Average intensity (I/(σ(I))) | 16.4 (5.3) | 6.2 (1.9) |
| B-factor (Wilson) (Å²) | 13.9/7.0 (6.1) | 46.8/7.9 (7.9) |
| Crystallographic reflections (free Rmerge) | 0.156 (0.193) | 0.207 (0.261) |
| Estimated overall co-ordinate error based on free Rfactor (Å) | 0.079 | 0.492 |
| No. of protein atoms/solvent molecules | 2,452/370 | 3,798/0 |
| Root mean square deviation from target values | 0.013/1.40 | 0.086 (0.448)/0.033 (0.180) 0.352 (0.595)/0.120 (0.330) |
| Bonds (Å/angles (°)) | 0.012/1.31 | 0.156/1.93 |
| Average B-factors for protein solvent molecule atoms (Å²) | 1.36/2.80 | 0.277/0.654 |
| Main-chain conformational angle analysis | 14.7/23.6 | 66.7/- |
| Residues in favored regions/outliers/all residues | 303/1/310 | 464/0/496 |

a Values in parentheses refer to the outermost resolution shell if not otherwise indicated.

b Rmerge = Σ|I(hkl)|/Σ|I(hkl)|, where |I(hkl)| is the i-th intensity measurement, Σ|I(hkl)| is the number of observations of reflection hkl, including symmetry-related reflections, and Σ|I(hkl)| is average intensity. 

c Rmerge = Σ|I(hkl)|/Σ|I(hkl)|, where |I(hkl)| is the i-th intensity measurement, Σ|I(hkl)| is the number of observations of reflection hkl, including symmetry-related reflections, and |I(hkl)| is average intensity. 

d Including atoms in alternate conformation.

e The last refinement step included anisotropic B-factor refinement of the protein atoms in the case of ΔN1 pro-cd-InpA (C154A).

f According to program MOLPROBITY (47).

Activation of P. intermedia InpA Cysteine Protease

Expression and activity of InpA. A, expression and purification of pro-cd-InpA C154A. Lanes 1 and 2, E. coli homogenate before and 3 h after protein expression induction, respectively. Lane 3, recombinant protein after affinity chromatography purification. Molecular masses of the distinct protein species (40 and 27 kDa) are shown on the left. B, same for wt pro-cd-InpA. C, time course analysis of autocalytic processing and activation of wt pro-cd-InpA (final concentration, 10 μM) incubated with 1 mM HgCl₂. The reaction was initiated by adding EDTA (5 μM final concentration) as an Hg₂⁺-chelator, i.e. by releasing metal-mediated inhibition. Samples were withdrawn at the time intervals specified (O/N, overnight incubation; lane C, pro-cd-InpA alone). D, same as C but after addition of active InpA (10 nM final concentration) to the reaction mixture. In this case, the reaction proceeded much faster. E, a subset of the volume withdrawn aliquots from C and D was used to quantify the activity released from wt pro-cd-InpA in the absence (C) and presence (D) of catalytic amounts of wt cd-InpA. As a control, pro-cd-InpA spiked with InpA but without EDTA was incubated in parallel (\( \square \)), concentration-dependent autoactivation of pro-cd-InpA. The reaction was initiated by releasing Hg₂⁺-mediated inhibition in mixtures containing 0.04 μM (\( \bigcirc \)), 2 μM (\( \bigcirc \)), and 10 μM (\( \triangle \))zymogen. At indicated time points, 50 μM (\( \square \)), 10 μM (\( \bigcirc \)), and 2 μM (\( \triangle \))were withdrawn from each reaction mixture and directly assayed for activity. G, SDS-PAGE of the digestion of pro-cd-InpA C154A by wt cd-InpA. The zymogen (final concentration of 10 μM) was incubated with cd-InpA (0.1 μM) for time intervals as specified (lane C, control pro-cd-InpA C154A incubated alone). N-terminal amino acid sequences of pro-cd-InpA derived fragments are indicated on the right. H, Western blot analysis of culture supernatant of P. intermedia using InpA-specific rabbit antiserum (lane 3). Wt cd-InpA and pro-cd-InpA (C154A) were loaded on lanes 1 and 2, respectively, for comparison.

release were accelerated by catalytic amounts of active cd-InpA (Fig. 2, D and E). This, together with the finding that the initial rate of activity generation was dependent on the zymogen concentration (Fig. 2F), suggested that the autocalytic maturation of pro-cd-InpA occurred in trans (intermolecularly). Pro-cd-InpA C154A was produced to elucidate the sequence of cleavage events during activation and for structural purposes. As in the case of other CPs, pro-cd-InpA C154A was enzymatically inert and did not undergo autoprocessing. Analysis of concentration- and time-dependent proteolysis of pro-cd-InpA C154A by the active protease revealed that the process occurred stepwise through a main 36-kDa intermediate (ΔN1-pro-cd-InpA C154A) generated by hydrolysis of peptide bond Thr28–Ala39. Accordingly, ΔN1-pro-cd-InpA C154A lacks the first 38 residues of the full-length zymogen (see Fig. 1). In addition, minor cleavages were mapped to Lys94–Ala95 and Ala95–Ile96 (Figs. 1 and 2G). Finally, limited proteolysis of the accumulating 36-kDa intermediate at the Asn111–Val112 peptide bond released the 27-kDa mature protein, which was resistant
to further degradation. The same activation pathway may operate in vivo, because a similar band pattern representing variably processed InpA species was detected in the P. intermedia culture medium (Fig. 2H). The maturing self processing of InpA resembles pro-SpeB with respect to formation of one major intermediate and several cleavages within the remaining part of the N-terminal pro-domain (36). Such a mechanism provides regulation of proteolytic activity independent of other secreted and host proteases, thus ensuring that the activity is developed when required.

Structure Solution Employing a Novel Approach—Contrary to the intact mutant protein, the ΔN1pro-cd-InpA C154A variant crystallized. Its structure was solved by Patterson-search methods using maximum-likelihood criteria. This approach improves the definition of the target for the search by removing the contribution of unknown variables. This means that the errors attributable to lack of completeness of a search model are better estimated. In practice, this entails a larger radius of convergence (i.e. it yields a solution for structurally more distant searching models) than conventional search methods, which failed in the present case. Unfortunately, the current crystallographic refinement programs have a shorter radius of convergence. This restricted model refinement and led us to develop a novel approach based on a further development of the SHELX suite of programs (37). It consists of the application of the “free-lunch algorithm,” whose theoretical basis had been developed by Giacovazzo and coworkers (38), combined with autotracing, model refinement, and density modification. This process essentially envisaged that the initially (poorly) refined model, displaying a weighted mean-phase error of 64° with respect to the final refined model (as determined a posteriori), was used to calculate an electron density map that was subjected to density modification. With
this map, missing structure-factor amplitudes and phases were estimated within the resolution range of the experimental data. Further values were extrapolated to a nominal resolution of 1.0 Å. Subsequently, density modification (weighted mean-phase error = 33°), main-chain auto-tracing and phase-combination (weighted mean-phase error = 27°) eventually produced an accurate partial model for ~60% of the residues. In addition, the resulting electron density map was excellent, even in those parts where the original search model showed a different chain trace (Fig. 3). This permitted straightforward manual tracing of the entire molecule and successful refinement, enabling us to ascertain three differences in comparison to the sequence of the PIN0048 open reading frame in the Institute for Genomic Research data base that were subsequently confirmed by sequencing at the DNA level (see Fig. 1).

Structure of InpA Zymogen—The protein has an elongated shape with an N-terminal pro-domain (Ala39–Asn111) and a C-terminal papain-like CP domain (Val112–Pro359), which bifurcates into a right subdomain (RSD) and a left subdomain (LSD) (see Fig. 4A). RSD and LSD interact through a surface of 1,332 Å2 establishing 69 contacts (~4 Å), among them 11 hydrogen bonds (~3.4 Å) and 22 hydrophobic interactions (Table 2). The pro-domain contacts laterally the top of the CP moiety through a surface of 1,777 Å2, with 54 contacts (~4 Å), among them 12 hydrogen bonds and 19 hydrophobic interactions (Fig. 4A and Table 2). The pro-domain is stabilized by a central hydrophobic core and evinces an open-faced sandwich with a twisted antiparallel four-stranded β-sheet (sheet I; strands β1–β4) of simple up-and-down connectivity mediated by short loops. After β4, a segment in extended conformation (loop joining strands β4 and α1, Lβ4α1) leading to helix α1. The N-terminal part of the helix approaches the active-site cleft, thus contributing to latency, and is hereafter termed “backing helix.” The polypeptide reaches the molecular surface after α1 and undergoes a sharp turn, folding back along the surface and entering a connecting segment that links the pro-domain with the CP domain. This segment adopts an extended conformation from Asn107 to Pro117, i.e., optimal for binding to and cleavage by an active-site cleft of a protease (39). This

FIGURE 3. Experimental electron density maps. A, representative example of the Fcalc-type map obtained after multistep-density modification, contoured at 1a above threshold and superimposed with the model placed in accordance to the Patterson search calculations (black stick model; residues 145–148, see PDB 1pvj) and the final refined coordinates of ΔN1 pro-cd-InpA C154A (light-gray stick model; residues 115–123). B, (mFcalc − DFcalc)-type aweighted omit map contoured at +1.75 a above threshold illustrating the distinct chain trace around the zymogenic hairpin of molecule A (final model in light gray) as compared with the ΔN1pro-cd-InpA model employed for phasing (black stick model). Some residues of either structure are labeled in the respective gray-tone for reference. C, same as B but showing the latency flap.

FIGURE 4. Structures of ΔN1pro-cd-InpA C154A and wt cd-InpA. A, Richardson diagram of ΔN1pro-cd-InpA C154A in standard orientation. The pro-domain is displayed in blue/cyan and the mature protein moiety (subdivided into a right and a left subdomain) in yellow/brown. The subdomains, the regular secondary structure elements (see Fig. 1), the N- and the C terminus, the primary activation point (at Asn111-Val112), and the structure regions responsible for latency maintenance are marked and labeled. B, superimposition of the Ca-carbon traces of ΔN1pro-cd-InpA C154A (yellow) and wt mature cd-InpA (red) in standard orientation. Some residues of ΔN1pro-cd-InpA C154A are labeled for reference. C, close-up view of the active site of ΔN1 pro-cd-InpA C154A. Orientation as in B after a horizontal rotation of ~45°. D, same as in C but for wt active cd-InpA. E, Ca-trace of the structure of ΔN1pro-cd-InpA C154A (yellow) and wt mature cd-InpA (red) around the active site, including the catalytic cysteine residue (Cys154; mutated to alanine in ΔN1pro-cd-InpA C154A), imbedded in active-site helix α2 (circled 1), the zymogenic hairpin (circled 3) encompassing the catalytic histidine (His155) (undefined from Ser289 to Gln301 in ΔN1pro-cd-InpA C154A) (circled 4), the backing helix α1 (absent in cd-InpA) (circled 1), and the latency-flap, displayed from Tyr332 to Met351 for either structure (circled 2). The gray arrows indicate the displacements of the keynote structural elements upon zymogen activation as explained in the text.
Activation of P. intermedia InpA Cysteine Protease

| TABLE 2 | Inter-domain (pro-domain/protease domain) and inter-subdomain (RSD/LSD) interactions in ΔN1 pro-cd-InpA |
|----------|----------------------------------------------------------------------------------------------------------------------------------|
| Pro-domain | Protease moiety | Dist. | RSD | LSD | Dist. |
| van der Waals interactions | A | van der Waals interactions | A |
| Phe14 C | Val107 Cys1 | 3.59 | Leu152 Cys2 | Glu161 Cys | 3.97 |
| Pro14 Cy | Lys108 Cys | 3.72 | Leu152 Cys1 | Tyr165 Cys | 3.82 |
| Pro31 C | Leu236 Cys2 | 3.79 | Thr239 Cys2 | Thr239 Cys2 | 3.50 |
| Leu37 Cys2 | Leu47 Cys2 | 3.68 | Ala100 Cys2 | Leu153 Cys | 3.90 |
| Gly76 Cys | Tyr236 Cys | 3.81 | Thr239 Cys2 | Met246 Cys | 3.66 |
| Phe80 Cys | Val236 Cys1 | 3.90 | Tyr239 Cys1 | Thr239 Cys | 3.70 |
| Pro32 Cys3 | Pro42 Cys3 | 3.50 | Pro42 Cys3 | Cys270 Cys | 3.71 |
| Phe95 Cys | Val236 Cys2 | 3.63 | Pro42 Cys3 | Phe266 Cys | 3.73 |
| Trp43 Cys3 | Phe237 Cys2 | 3.90 | Ile250 Cys1 | Val252 Cys2 | 3.79 |
| Trp43 Cys2 | Leu250 Cys1 | 3.91 | Ile263 Cys2 | Met264 Cys | 3.93 |
| Trp43 Cys | Glu159 Cys | 3.78 | Ile263 Cys1 | Leu153 Cys | 3.94 |
| Trp43 Cys | Glu159 Cys | 3.78 | Ile263 Cys2 | Val252 Cys | 3.58 |
| Trp43 Cys2 | Tyr236 Cys1 | 3.58 | Ala100 Cys2 | Met246 Cys | 3.74 |
| Trp43 Cys1 | Leu236 Cys1 | 3.97 | Val152 Cys1 | Val252 Cys2 | 3.67 |
| Lys145 Cys | Phe160 Cys | 3.57 | Asp160 Cys2 | Met264 Cys | 3.73 |
| Ala145 Cys | Phe154 Cys | 3.97 | Val252 Cys2 | Met264 Cys | 3.91 |
| Ile146 Cys | Leu236 Cys2 | 3.73 | Val252 Cys1 | Leu252 Cys | 3.93 |
| Val236 Cys2 | Leu236 Cys2 | 3.95 | Val252 Cys1 | Val252 Cys2 | 3.86 |
| Val236 Cys1 | Thr239 Cys2 | 3.35 | Val252 Cys1 | Leu153 Cys | 3.88 |
| Hydrogen bonds | | | | | |
| Leu74 O | Lys332 N | 2.87 | Val252 Cys1 | Phe259 Cys | 3.75 |
| Val76 N | Tyr331 O | 2.79 | Tyr236 Cys | Phe266 Cys | 3.62 |
| Ser86 Oy | Asp339 O | 2.72 | | | |
| Thr127 Oy | Asp339 O | 2.75 | Thr127 Oy | Asp339 O | 2.85 |
| Asn96 Oy | Asp339 O | 3.19 | Thr127 Oy | Glu161 Oy | 2.75 |
| Ser96 Oy | Alas336 N | 3.01 | Thr127 N | Asp339 O | 2.97 |
| Trp97 N | Phe160 | 2.92 | Thr127 N | Gly298 | 2.99 |
| Trp97 N | Lys332 N | 2.87 | | | |
| Glu100 Oy | Phe266 Cys | 3.83 | Ala250 O | Arg494 Nyl | 2.78 |
| Glu100 Oy2 | Thr239 N | 3.04 | Ala250 O | Phe266 O | 3.03 |
| Asn110 N2 | Asp339 O2 | 2.48 | Asn110 O2 | Glu161 N2 | 2.97 |
| Asn111 O1 | Asp339 O | 3.17 | Lys332 N | Thr127 O | 3.33 |
| Asn111 O2 | Asp339 O2 | 3.25 | | | |
| Lys154 Nc | Tyr165 Oy | 2.80 |

stretch includes the activation cleavage point, Asn111–Val112 (Fig. 4A), which is superficial and accessible for processing.

At Val112, the polypeptide chain enters the RSD of the mature enzyme moiety, which is a split subdomain (Val112–Leu127 plus Thr260–Pro239) with an open-faced sandwich topology created by a six-stranded twisted antiparallel β-sheet (sheet II; strands β11–β16). The sheet extends from the bottom of the molecule (outermost strand β11) to the interface with the pro-domain at β15 (Fig. 4A). The twist gives rise to a concave and a convex face, and the latter mediates the main interaction with the LSD. The main contact between the pro-domain and the CP part is formed by the outermost strand of sheet I, β4, and the lateral strand of sheet II, β15. This gives rise to a continuous ten-stranded β-sheet that completely traverses the zymogen from its upper right to the bottom center (Fig. 4A).

After the inset of the LSD (see below), the polypeptide chain rejoins the RSD at strand β11 of sheet II, which runs outward approximately perpendicular to the view in Fig. 4A. After this strand, a short loop leads to helix α5, which nestles in the concave side of sheet II, followed by the next four strands of sheet II (β12–β15), inserted with simple up-and-down connectivity. These strands are connected by loops, which contribute to the substrate-binding cleft and the active site. The polypeptide chain is very well defined for the whole protein moiety except for the tip of a β-hairpin structure created by strands β12 and β13 and the enclosed loop, the “zymogenic hairpin” in the following. The hairpin is rigid at its trunk, because it is stabilized by six β-sheet interactions between β12 and β13, but flexible at its tip (between Ser295 and Gln301 (Fig. 4)). After β15, the polypeptide runs below the backing helix α1 and gives rise to what will now be referred to as the “latency-flap,” Lβ15β16, which spans the 16 residues from Ile334 to Gln349. This structure displays a unique conformation and is stabilized by a series of internal contacts. It consists of two sequential dextrohelical elements, Ile334–Asn338 and Ser344–Gln349, connected by two residues in extended conformation (Pro339–Gly340) and a tight 1,4-turn of type I (Asn341 O–Ser344 N, 3.13 Å), which protrudes from the molecular surface. The bottom of the first dextrohelical segment is anchored to Lβ11α5 through a bidentate interaction of its main chain with the completely buried side chain of Arg267 and includes another tight 1,4-turn of type I (Ile334 O–Leu337 N; 2.94 Å). In addition to this arginine anchor, the structure of the latency flap is galvanized by a total of nine internal hydrogen bonds that confer an extraordinary rigidity to this structural element. After this flap, the protein chain enters the second strand of sheet II, β16, and leads to the surface C terminus of the molecule at Pro352, whose position permits additional downstream domains in the full-length InpA protein (Fig. 4A).

The LSD (Leu128–Phe259) is inserted into the RSD and is characterized by a central three-helical bundle made up by helices α2, called the “active-site helix,” as well as α3 and α4, which traverse the subdomain from the back to the front (Fig. 4A). In addition, three β-hairpins are found on the front side of the LSD, β5β6, β7β8, and β9β10. After α4, the polypeptide chain rejoins the RSD at Thr260 and leads to β11. As observed for the pro-domain, the LSD is held together by a large central hydrophobic cluster that reaches the subdomain surface at the bottom and at the left of the molecule and accommodates active-sites helix α2.

Substrate-binding Crevices and Active Site—The active-site cleft of InpA is in a crevice formed by loops connecting strands of sheet II at its carboxyl end. The walls of the crevice are provided by RSD and LSD (see Fig. 4). Classic CPs like papain, cathepsin B, and staphopain have a short, four-residue segment connecting the two residues that are topologically equivalent to Gln134 and Gly135 of InpA, respectively, as contributors to the left rim of the cleft on its primed side. In contrast, InpA displays between the latter two residues an 18-residue insertion that forms a unique upper-left region of the molecule. This entails that the zone ascribable to substrate binding would be reduced in InpA to Gln133–Gly135 and Thr152–Gly153, immediately preceding the catalytic cysteine, Cys154. The former stretch includes Gln133, whose position is absolutely conserved among CPs and which, by analogy, would be involved in the formation of an oxanion hole together with the amide nitrogen of Cys154, which would bind the scissile carbonyl (21, 40). On the non-primed side of the cleft, Ser242–Met246 and Tyr264 would also contribute to the left rim. Again in contrast to classic CPs, InpA possesses a much longer connection between helices, which shapes part of the front surface and gives rise to a unique β-hairpin, novel for CPs (β9β10). This entails that the residues from Pro238 to Gly241 should further assist Ser242–Met246 in shaping the cleft rim. In even greater contrast to clas-
sic CPs, the segments shaping the right-hand rim of the cleft on its primed side may be restricted to the side chains of the strongly conserved Trp^{284} from Lβ14β15, which becomes rearranged upon activation, and the previously mentioned Gln^{154} (21). Regarding the right rim on the non-primed side of the cleft, binding may be provided by the main chain of the rearranged zymogenic hairpin, in particular His^{305}.Ala^{306} and Tyr^{293}.Gly^{293}, as well as Asp^{350}

**Structures Related to InpA**—As might have been expected, a search for structural relatives of ΔN1pro-cd-InpA identified pro-SpeB as the closest homologue, with an rms deviation of 2.0 Å over 275 topological equivalent residues (PDB 1dki and 1pvj (19)). This protein is secreted as a zymogen, and no structural information on the mature protein is currently available. InpA and SpeB are the only members of the catalytic-dyad enzymes, i.e., those lacking a catalytic asparagine, structurally studied to date (19). Because *P. intermedia* has been shown to degrade connective-tissue constituents and to interfere with the tightly regulated defense mechanism of the host (9), like SpeB in *S. pyogenes* (22), it is tempting to speculate that InpA is a virulence factor equivalent to SpeB in *P. intermedia*. In addition, *S. pyogenes* was shown to harbor a further CP, periodontain (41), which is closely related to SpeB and InpA. Accordingly, we conclude that *P. intermedia* and *S. pyogenes* may have inherited these homologous proteins from a common ancestor and that they may undergo a similar activation mechanism (42).

Overall, the core of the protease and the pro-domain of InpA conform to the pro-SpeB fold (Fig. 5A). However, the difficulties encountered during ΔN1pro-cd-InpA structure solution employing pro-SpeB as a search model for phasing and a sequence identity of just 28%, *i.e.* in the twilight zone of protein sequence alignments (43), already pointed to significant differences in structure. The pro-SpeB crystal structure displays unconnected electron density for a helix nesting on the concave side of sheet I of the pro-domain. Several secondary structure prediction algorithms consistently predicted an α-helix to similarly run from Lys^6 to Asn^{18} in pro-cd-InpA (Fig. 1). However, differences in length in the loop connecting this (putative) first helix with strand β1 (nomenclature of ΔN1pro-cd-InpA, see Fig. 1 for equivalences), as well as in Lβ1β2, lead sheet I to have a bulge on its left-hand side in the streptococcal enzyme, which is compensated in

**FIGURE 5. Comparison of ΔN1pro-cd-InpA C154A with pro-SpeB.** A, superimposition of the Cα-carbon traces of ΔN1pro-cd-InpA C154A (blue) and pro-SpeB (yellow) in standard orientation revealing the large-scale structural deviations of the activation segments. Some residues of ΔN1pro-cd-InpA C154A are labeled for reference. B, detail of the region around the active site, including segments shown under Fig. 4E, *i.e.* the catalytic cysteine residue (Cys^{154}, alanine in ΔN1pro-cd-InpA C154A), within active-site helix α2, depicted for its residues Gly^{153}.Ala^{158} (pro-cd-InpA numbering, see Fig. 1) (circled 1), the zymogenic hairpin including the catalytic histidine (His^{305}), shown for Tyr^{291}.Phe^{307} (circled 2), the backing helix α1 from the pro-domain, shown for Pro^{187}.Ala^{195} (circled 3), and the segment containing the latency-flap, displayed from Gly^{340} to Asp^{350} (circled 4).
Activation of P. intermedia InpA Cysteine Protease

Lβ9β10, possibly involved in the left rim of the crevice in InpA (see above), also diverges due to the three extra residues in the latter protease. The greatest differences, however, affect regions surrounding the active site, in particular the zymogenic hairpin and the latency flap. The former, comprising the catalytic histidine in both proteins, is flexible and five residues longer in ΔN1 pro-cd-InpA, where it adopts a different orientation. In turn, the segment equivalent to the latency flap is completely disordered between Ser\(^{230}\)\(_{\text{SPE}}\) and Glu\(^{239}\)\(_{\text{SPE}}\) and has six additional residues in pro-SpeB (Figs. 1 and 5), following a completely different path. Hence, it does not contribute significantly to interactions with the zymogenic hairpin or the backing helix to maintain the zymogenic structure. Furthermore, in pro-SpeB the polypeptide preceding Ser\(^{230}\)\(_{\text{SPE}}\) invades the space occupied by the segment connecting the pro-domain with the mature moiety in pro-cd-InpA, thus pointing to differences in the segment flanking the primary activation cleavage point.

In contrast to these differences, there are also similarities in detail. As in ΔN1 pro-cd-InpA, latency is achieved in pro-SpeB through a catalytically incompetent conformation of the catalytic histidine, His\(^{195}\)\(_{\text{SPE}}\) while the catalytic cysteine is probably in a functional position. It is conceivable, extrapolating from our structures, that the zymogenic hindrance is exerted in the streptococcal enzyme by a simple ~90° rotation around the \(\chi_1\) angle of His\(^{195}\)\(_{\text{SPE}}\). This movement swings the imidazole side chain away from its cysteine-binding position and establishes a van-der-Waals interaction with Val\(^{192}\)\(_{\text{SPE}}\) C-γ2 within the hairpin segment equivalent to the zymogenic hairpin in ΔN1 pro-cd-InpA. The competent imidazole position is occupied in pro-SpeB by a unique asparagine, Asn\(^{89}\)\(_{\text{SPE}}\) from the pro-domain, which establishes a highly-specific key hydrogen bonding network with Trp\(^{214}\)\(_{\text{SPE}}\) Ne1, Ala\(^{106}\)\(_{\text{SPE}}\) O, and Trp\(^{212}\)\(_{\text{SPE}}\) O (19). The position equivalent to Asn\(^{89}\)\(_{\text{SPE}}\) is occupied in ΔN1 pro-cd-InpA by Ser\(^{88}\), which establishes one of these three interactions (with Ala\(^{306}\) O) in the InpA zymogen as one of the elements likewise leading to an incompetent histidine conformation. Accordingly, the major differences in the structure of thezymogens do not preclude that the novel activation mechanism described below may also be valid with variations for pro-SpeB and, possibly, periodontal activation.

A Novel Mechanism For Latency Maintenance and Activation—The mature enzyme structure confirms that the pro-domain, including the backing helix, is removed upon activation and that it does not sterically block access to the substrate-binding cleft in ΔN1 pro-cd-InpA. There are two parallels between thiszymogen and other CPzymogens such as human and rat pro-cathepsin B (44), K (45), and pro-staphopain B (16). In the latter, the pro-segment packs against a surface loop of the C-terminal domain termed the pro-segment binding loop. This loop is absent in ΔN1 pro-cd-InpA but its pro-domain binds in the same place. A further common feature is that the association between the pro-domain and the protease domain is based on hydrophobic residues. However, in the mentioned CPzymogens the pro-domain segments run the full length of the cleft in the opposite direction to a peptidyl substrate, and block the crevice. In the InpAzymogen, in contrast, backing helix α1 and the preceding loop Lβ4α1 are inserted laterally like a wedge (Fig. 4A). Trp\(^{324}\), from Lβ14β15 within the CP domain, stops the wedge with its side chain (Fig. 4C). The relative antipodal disposition on the molecular surface of the active site and Val\(^{122}\) (Fig. 4), which are ~26 Å apart, supports kinetic data (see above) suggesting that autolytic activation of InpA is likely to occur in trans. The CP domain is similar in both the mature enzyme and thezymogen (239 out of 248 common Ca atoms show a root mean square deviation of 0.82 Å; see Fig. 4B). Interestingly, activation is not correlated with significant displacement of the newly formed N terminus at Val\(^{122}\) (Fig. 4B). Despite this similarity, detailed comparison of the two structures reveals that selected structure elements display completely different chain traces (Fig. 3).

In CPs, function requires a correct spatial arrangement of the catalytic cysteine provided by the active-site helix within LSD and the catalytic histidine of the RSD to render a functional thiolate-imidazolium ion pair (46) (Fig. 4, C–E). Unlike InpA, most other CPs also have an asparagine with a supportive role (46). The position and conformation of the active-site helix and the cysteine, Cys\(^{154}\), are maintained in both InpAstructures. In contrast, the catalytic histidine, His\(^{305}\), undergoes major rearrangement. In the mature enzyme it is oriented to favor the interaction with Cys\(^{154}\) and through a hydrophobic bond, His\(^{305}\) Ne2–Trp\(^{322}\) O, and is further stabilized in this position by a hydrophobic environment created by Trp\(^{324}\), Phe\(^{407}\), Phe\(^{445}\), and Trp\(^{322}\) (Fig. 4D).

In thezymogen, however, the histidine is swung out from its active position. It requires a rotation of ~45° around bond Ala\(^{306}\) Ca-N and of ~180° around its \(\chi_1\) angle to adopt an active conformation (Fig. 4, C–E). The position of His\(^{305}\) in thezymogen is stabilized by three of elements that contribute to a compact “histidine cage” structure (Fig. 4C): the backing helix, thezymogenic hairpin, and the latency flap. The first one is completely removed and the further two undergo major rearrangement upon activation (Fig. 4D).

As mentioned, thezymogenic hairpin is only defined until residue Gly\(^{294}\) and from Asp\(^{302}\) onwards in the ΔN1 pro-cd-InpA structure so that the enclosed region Lβ12β13 is disordered. The position of the hairpin base is kept by interactions with surrounding elements. Strand β12 establishes three inter-main-chain contacts with β16. In turn, β13 interacts with the backing helix through a hydrophobic bond (Ala\(^{306}\) N–Ser\(^{88}\) O, 3.01 Å) and face-to-face ring stacking between His\(^{305}\) and Trp\(^{91}\). Removal of the backing helix leads to a rearrangement of thezymogenic hairpin due to a rotation of ~45° producing a maximal displacement of ~7 Å (measured at Ala\(^{303}\) Ca). In addition, thehairpin becomes rigid and fully defined by electron density (Fig. 4, C–E). The hairpin-constituting β-strands are extended to Gly\(^{297}\) (β12) and from Gln\(^{301}\) onwards (β13) and give rise to a new intra-main-chain interaction (Ser\(^{295}\) N–Ala\(^{303}\) O). These changes carry along the activatory reorientation of His\(^{305}\) (Fig. 4E).

Another important element shaping the histidine cage in thezymogen is the latency flap, which anchors the catalytic histidine in the non-competent position through a hydrophobic bond (Glu\(^{348}\) Oe1–His\(^{305}\) Ne2). In addition, the latency flap interacts with the backing helix through three hydrophobic bonds, a hydrophobic interaction, and a small hydrophobic cluster made up by

2880 JOURNAL OF BIOLOGICAL CHEMISTRY
the side chains of Ala$^{395}$, Val$^{399}$, Leu$^{337}$, and Thr$^{346}$. This cluster extends below the side chains of His$^{305}$ and Trp$^{31}$ and further incorporates Tyr$^{281}$, Phe$^{307}$, Trp$^{322}$, Tyr$^{333}$, Ile$^{334}$, and Trp$^{345}$. Upon removal of the backing helix, the latency flap undergoes a large rearrangement and displacement caused by a $\sim$110° rotation that causes maximal displacement of $\sim$22 Å. Simultaneously, the latency flap adopts a $\beta$-hairpin structure with two extended segments, Leu$^{338}$–Pro$^{339}$ and Tyr$^{343}$–Phe$^{345}$, parallel to each other and establishing hydrogen bonds (see Fig. 4, C–E). These two extended segments are joined at the top and form a small hairpin. In its new position, the latency flap resides on a hydrophobic pillow created by the region preceding the flap, $\beta$14–$\beta$14B15–$\beta$15. This region accommodates the backing helix in thezymogen and remains unchanged after activation, only the interacting partners change. In addition to these interactions, the latency flap provides a physical support to the zymogenic hairpin in the active enzyme through four main-chain and a side-chain/main-chain interactions. A last finding concerning concomitant with the removal of the backing helix is the reorientation of the side chain of stopper Trp$^{324}$ through two rotations of $\sim$100° and $\sim$60° around its angles $\chi_1$ and $\chi_2$, respectively (see Fig. 4, C and D). In this way, the side chain of this residue joins the previously mentioned hydrophobic pillow acting as a “tryptophan switch” and participating in the activating rearrangement.

In summary, we have described a new cysteine protease from a highly-active pathogenic bacterium, InpA, which undergoes autolytic activation in vitro and, possibly, in vivo. The structural features reported reveal a new mechanism of activation/latency maintenance within CPs, distinct from cathepsin and plant CPs, which may also be valid for related proteins such as S. pyogenes SpeB and P. gingivalis periodontin. This mechanism starts when the backing helix is removed after proteolytic cleavage at the Asn$^{311}$-Val$^{312}$ scissile peptide bond (step 1 in Fig. 4E). This liberates a space that enables stopper Trp$^{324}$, actually a tryptophan switch, to reorient (see Fig. 4, C and D) and contribute to a hydrophobic pillow created by the apolar side chains of segment $\beta$14–$\beta$14B15–$\beta$15 of the CP moiety. This segment participates in a large hydrophobic core with the backing helix in thezymogen and remains unchanged in the active enzyme. The movement of Trp$^{324}$ correlates with the large displacement and internal rearrangement observed for the latency flap which, by pivoting around Met$^{351}$ and Tyr$^{332}$, causes this segment to adopt a $\beta$-hairpin-like structure and to occupy the space released by the backing helix (step 2 in Fig. 4E). Consequently, the zymogenic hairpin becomes rigid at its top and folds back to a $\sim$60° rotation pivoting around the anchor points Gly$^{292}$ and Ala$^{306}$, thus liberating the substrate-binding cleft of the enzyme (step 3 in Fig. 4E). This rearrangement correlates with a $\sim$180° rotation of the His$^{305}$ side chain to a competent position to interact with the catalytic cysteine (step 4 in Fig. 4E).

Acknowledgments—We thank Robin Rycroft and Mary Kopecki for helpful contributions to the manuscript and George M. Sheldrick for providing an $\alpha$-version of program SHELXE.
Activation of P. intermedia InpA Cysteine Protease

38. Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., and Siliqi, D. (2005) Acta Crystallogr. Sect. D Biol. Crystallogr. 61, 556–565
39. Tyndall, J. D. A., Nall, T., and Fairlie, D. P. (2005) Chem. Rev. 105, 973–999
40. Drenth, J., Kalk, K. H., and Swen, H. M. (1976) Biochemistry 15, 3731–3738
41. Nelson, D., Potempa, J., Kordula, T., and Travis, J. (1999) J. Biol. Chem. 274, 12245–12251
42. Madden, T. E., Clark, V. L., and Kuramitsu, H. K. (1995) Infect. Immun. 63, 238–247
43. Rost, B. (1999) Protein Eng. 12, 85–94
44. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) Structure 4, 405–416
45. Sivaraman, J., Lalumière, M., Ménard, R., and Cygler, M. (1999) Prot. Sci. 8, 283–290
46. Polgár, L. (2004) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D., and Woessner Jr., J. F., eds) Vol. 2, 2nd Ed., pp. 1072–1079, 2 vols., Elsevier Academic Press, London
47. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Bryan Arendall, W., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) Nucleic Acids Res. 35, W375–W383