Structural and functional characterization of SplA, an exclusively specific protease of Staphylococcus aureus

Justyna Stec-Niemczyk, Katarzyna Pustelny, Magdalena Kisielewska, Michal Bista, Kevin T. Boulware, Henning R. Stennicke, Ida B. Thogersen, Patrick S. Daugherty, Jan J. Enghild, Krzysztof Baczynski, et al.

Structural and functional characterization of SplA, an exclusively specific protease of *Staphylococcus aureus*

Justyna Stec-Niemczyk[a], Katarzyna Pustelny[a], Magdalena Kisielewska[b], Michal Bista[a,g], Kevin T. Boulware[c], Henning R. Stennicke[d], Ida B. Thogersen[e], Patrick S. Daugherty[c], Jan J. Enghild[e], Krzysztof Baczynski[f], Grzegorz M. Popowicz[g], Adam Dubin[a,h], Jan Potempa[b] and Grzegorz Dubin[b]*

[a] Department of Analytical Biochemistry, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland
[b] Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland
[c] Department of Chemical Engineering, University of California, Santa Barbara, CA, USA
[d] Protein Engineering, Novo Nordisk A/S, Maaloev, Denmark
[e] Laboratory for Proteome Analysis and Protein Characterization, Department of Molecular Biology, University of Aarhus, Aarhus, Denmark
[f] Department of Computational Biophysics and Bioinformatics, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland
[g] Max-Planck Institute of Biochemistry, Martinsried, Germany
[h] BioCentrum Ltd., Krakow, Poland

[*]Corresponding author:

Jagiellonian University
Faculty of Biochemistry, Biophysics
and Biotechnology
Gronostajowa 7
30-387 Krakow, Poland
Phone: (+4812) 6646362
Fax: (+4812) 6646902
E-mail: grzegorzdubin@hotmail.com
gdubin@mol.uj.edu.pl

Keywords: serine protease; structure; substrate specificity; *Staphylococcus aureus*

Summary

Staphylococcus aureus is a dangerous human pathogen which antibiotic resistance is steadily increasing and no efficient vaccine is as yet available. This serious threat drives extensive studies on staphylococcal physiology and pathogenicity pathways, especially virulence factors. Serine protease like proteins (Spl) encoded by an operon containing up to six genes are a good example of poorly characterized secreted proteins likely involved in virulence. Here we describe an efficient heterologous expression system for SplA and detailed biochemical and structural characterization of the recombinant SplA protease. The enzyme shares a significant sequence homology to V8 protease and epidermolytic toxins which are well documented staphylococcal virulence factors. SplA has a very narrow substrate specificity apparently imposed by the precise recognition of three amino acid residues positioned N-terminal to the hydrolyzed peptide bond. To explain determinants of this extended specificity we resolve the crystal structure of SplA and define consensus model of substrate binding. Furthermore we demonstrate that artificial N-terminal elongation of mature SplA mimicking a naturally present signal peptide abolishes enzymatic activity. The likely physiological role of the process is discussed. Of interest, even though precise N-terminal trimming is a common regulatory mechanism among S1 family enzymes, the crystal structure of SplA reveals novel, significantly different mechanistic details.

Introduction

Species of *Staphylococcus* genus are commensal inhabitants of man and warm-blooded animal epithelium [1]. Nearly 30% of human population is persistently colonized by *Staphylococcus aureus* without visible adverse effects. At the same time the bacterium is a dangerous opportunistic pathogen. Microinjurys of skin or surgical wounds can lead to invasion of subepithelial tissues. *S. aureus* reveals a great ability to overcome host defense mechanisms and colonize diverse organs. The pathogen is equipped with a number of secreted factors facilitating superficial and systemic infections. Examples of soft tissue superficial infections include impetigo, folliculitis, abscesses, boils or infected lacerations [2]. Deep systemic infections are more dangerous and often life-threatening (e.g. endocarditis, meningitis, septic arthritis, pneumonia or osteomyelitis) [3]. The most serious outcome of *S. aureus* infection is septicemia. Moreover, the bacterium is responsible for several toxinoses, such as staphylococcal scalded skin syndrome (SSSS), toxic shock syndrome (TSS) and food poisoning. The wide variety and severity of infections caused by *S. aureus* together with a frightening increase of multi-antibiotic resistance and the lack of effective vaccines poses an urgent need for novel therapeutic strategies [4, 5]. This in turn necessitates deeper understanding of staphylococcal physiology and pathogenicity pathways.

Virulence factors responsible for staphylococcal pathogenicity encompass secreted toxins, immune-modulatory and adhesion molecules, signaling factors and extracellular enzymes, including a number of proteases [6]. The role of cysteine proteases staphopain A and B, metalloprotease aureolysin, serine protease V8 (glutamylendopeptidase) and epidermolytic toxins in staphylococcal virulence is documented in multiple reports [7-10]. Furthermore, screening of a library of staphylococcal proteins with antisera from patients with *S. aureus* endocarditis clearly showed that a new serine protease like protein, SplC, is expressed during infection [11]. Follow up investigations demonstrated that SplC is a product of a gene belonging to an operon encoding up to six homologous Spl proteases in various strains of *S. aureus* [12]. Despite their suggested involvement in pathogenesis, the Spl

proteases remain poorly characterized. Similar to other staphylococcal exo-proteins, *in vitro* the *spl* operon is transcribed during early stationary phase of bacteria growth [13]. It is postulated that similar regulation operates *in vivo* and is essential for dissemination of the pathogen from initial colonization sites. Therefore, Spl proteases together with multiple similarly regulated documented virulence factors may play a role in the second, invasive stage of the infection. The likely contribution of Spl proteases to *S. aureus* pathogenicity is further substantiated by structural similarities to established virulence factors, such as V8 protease and epidermolytic toxins, and clustering of the *spl* genes on a stable pathogenicity island (vSaβ) [14, 15].

To date out of six Spl proteins only SplB and SplC were studied in some details. The lack of proteolytic activity of SplC was explained on the atomic level by showing that the catalytic machinery is significantly distorted in the crystal structure of the protein. Moreover, crystal structure of highly specific protease SplB shown lack of an important element of serine-protease catalytic machinery, the oxyanion hole. By comparison to epidermolytic toxins authors speculated that binding of an appropriate substrate could induce activating changes in SplC, similarly to SplB which activity seems induced by binding of a substrate encompassing the sequence Trp-Glu-Leu-Gln [15, 16].

In this report, to cast more light on staphylococcal proteases of *spl* operon we have produced SplA using heterologous expression system in *B. subtilis* and performed biochemical and structural characterization of the recombinant protein. We demonstrate a crucial effect of proper N-terminal trimming for SplA activity. Of interest, we characterize the enzyme strict specificity imposed by recognition of amino acid residues at P1, P2 and P3 positions of the substrate (nomenclature according to Schechter and Berger [17]). The structural basis of substrate preference and significance of the N-terminal residue of the mature protein in SplA activity are visualized at the atomic level and contrasted with type proteases of chymotrypsin family.

Experimental

Bacterial strains, plasmids, protein expression and purification

Escherichia coli strains DH5α and BL21(DE3) (Invitrogen) were used respectively for genetic manipulations and protein expression. For production of GST-GS-SplA (GS indicates that following tag removal, a Gly-Ser artifact remains at the N-terminus of SplA compared to the native protein), the *splA* gene (without a signal peptide) was PCR amplified from the genomic DNA of *S. aureus* strain 8325-4 and cloned into BamHI/XhoI sites of pGEX-5T [18]. GST-YLYS-Staphostatin A fusion protein (GST-YLYS-ScpB; YLYS indicates a Tyr-Leu-Tyr-Ser sequence engineered between two globular fusion partners) was produced as previously described [19] with the exception that in the middle part of the construct the amino acid sequence recognized by thrombin (LVPRGS) was exchanged by site directed mutagenesis into consensus amino acid sequence recognized by SplA (LVYLYS). For SplA production *B. subtilis* strain WB800 carrying a pWB980 vector [20] with the *splA* gene (without a native signal peptide) inserted at HindIII/XhoI sites was used.

E. coli cells containing pGEX5T-GS-SplA vector were cultured at 37˚C in LB containing ampicillin (100 μg/mL) until OD_{600nm} reached 0.8 absorbance units and latter at room temperature. Protein expression was induced with 1mM IPTG when the cultures reached OD_{600nm} of 1.2 absorbance units. After 3h cells were collected in PBS and lysed by sonication. A protein of interest was recovered using Glutathione-Sepharose (Amersham Biosciences). GST fusion tag was removed using thrombin (BioCentrum, Krakow). GS-SplA was further purified by ion exchange chromatography (SoruceS; Amersham Biosciences; 50mM Sodium Acetate pH 6.5) followed by gel filtration on Superdex 75pg (Amersham Biosciences) equilibrated with PBS.

Routinely recombinant SplA was produced in *B. subtilis* WB800 carrying pWB980-SplA plasmid driving a constitutive secretory expression. The cells were cultured at 37˚C overnight in TSB medium containing kanamycin (10 μg/mL). Cell free supernatants were collected by centrifugation and proteins were precipitated with ammonium sulfate at 80% saturation at 4˚C (561 g/L). Precipitated proteins were collected by centrifugation, resuspended in buffer A (50 mM amonium acetate, pH 5.5) and dialyzed overnight against the same buffer. SplA was recovered by ion exchange chromatography on SP-Sepharose Fast Flow (Amersham Biosciences) equilibrated with buffer A. Fractions with conductivity between 27 mS/cm and 37 mS/cm, containing proteolytic activity were pulled, and rechromatographed on SourceS (Amersham Biosciences) in buffer B (50 mM amonium acetate, pH 5.0). Proteolytically active fractions were pulled and further purified on Superdex 75pg (Amersham Biosciences) equilibrated with crystallization buffer (5 mM Tris, 50 mM NaCl, pH 8.0).

GST-YLYS-ScpB was purified as previously described for GST-ScpB [19].

Proteolytic activity assays

Activity of SplA was routinely tested by casein zymography [21]. Briefly, samples were incubated with SDS loading buffer (without reducing agent) in 37˚C for ½ hour and separated on 12% (w/v) polyacrylamide gel containing 0.1% β-casein. Following electrophoresis the gel was washed for 30 min in 2.5% (v/v) Triton X-100 and developed for 3 hours or overnight in 100 mM Tris (pH 7.8), 100 mM NaCl, 5 mM EDTA at 37˚C. Zones of proteolysis were visualized by staining with Amido Black.

Synthetic chromogenic substrates tested included: H-AF-*p*Na, H-GF-*p*Na, N-Suc-F-*p*Na, N-Suc-AAPF-*p*Na, N-Ac-LY-*p*Na, N-Suc-LF-*p*Na and N-Suc-AAA-*p*Na. SplA (20 μg) was incubated with 0.5 mM substrate in a total volume of 200 μL 100 mM Tris-HCl, pH 7.0 for 16h at 37°C in a microtitration plate. Following incubation OD at 405 nm was measured using a plate reader. pH optimum of SplA was determined using N-Suc-AAPF-*p*Na. This substrate was also used to assay the effect of synthetic serine protease inhibitors, including phenylmethanesulfonyl fluoride (PMSF), 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF) and 3,4-Dichloroisocumarin (DCI) on SplA activity.

A number of native proteins and peptides including: chicken egg white lysozyme and ovalbumin, soybean trypsin inhibitor, human serum transferrin, bovine serum albumin, RNAse, cytochrome c, β-casein, galanine, goat IgGs, bovine and human fibrinogen and whale mioglobin were tested for hydrolysis by SplA in Bis-Tris buffer (pH 6.5). Moreover, proteins denatured by chemical modification, including: carboxymethylated chicken egg white lysozyme and bovine pancreatic trypsin inhibitor BPTI, as well as apomioglobin and apocytochrome c were evaluated as substrates for SplA.

Human serum serine protease inhibitors including α-1-antitrypsin, α-1-antichymotrypsin, antithrombin and α_2-macroglobulin were tested for inhibition of SplA activity. The residual activity was determined using GST-YLYS-ScpB as a substrate after 30min preincubation of SplA with a tested inhibitor at 1:100 molar ratio. In addition, SplA-inhibitor reaction mixtures after 15 minutes of incubation at 37˚C were analyzed by SDS-PAGE for inhibitory complex formation or potential inhibitor cleavage.

Edman degradation and MALDI-TOF MS analysis

For N-terminal sequence determination, protein cleavage products were separated by SDS-PAGE, electro-blotted on PVDF and sequenced by automated Edman degradation sequencing at BioCentrum (Krakow, Poland).

For determination of molecular weights the protein cleavage products were separated by HPLC and subjected to MALDI-TOF MS (matrix-assisted laser desorption/ionization-time-of-flight) analysis.

Positional scanning substrate libraries

To assess the preference of SplA for amino acid residue at P1 substrate position positional scanning substrate libraries were used as described previously [22].

Cellular library of peptide substrates (CLiPS)

A complete consensus sequence recognized and cleaved by SplA protease was discerned using a CLiPS methodology as previously described [23]. In brief, a library of $6x10^7$ clones displaying a bait polypeptide containing an SH3 domain binding motif, five consecutive randomized amino acids linker and a streptavidin binding peptide ligand was screened for proper display and SplA hydrolysis using fluorescence activated cell sorting (FACS). Clones with intact baits show red and green fluorescence after incubation with phycoerythrin conjugated streptavidin (50 nM) and SH3 domain conjugated Green Fluorescent Protein GFP (250 nM). Clones displaying substrates that were specifically hydrolyzed after 16 hours of incubation with 0.5 μM SplA were obtained by sorting the green cells only. After three rounds of sorting individual clones were analyzed for hydrolysis by incubating with 0.5 μM SplA for 3 hours, and clones demonstrating hydrolysis were sequenced (Table I).

Protein crystallization and structure solution

Purified active SplA and SplA reacted with phosphonate inhibitor [24, 25] were concentrated by ultrafiltration to ~30 mg/ml in crystallization buffer and used for initial screening. After few weeks crystals appeared under several different conditions. For active SplA following further optimization a best single monocrystal of about 0.3 mm length was obtained using Hampton Research Index 23 (2.1 M DL-Malic acid, pH 7.0) and was used for measurements. The crystal was mounted in a glass capillary and the diffraction data were collected on a rotating anode laboratory source at room temperature. Best single monocrystal of inhibited SplA obtained from an in-house screen (0.1M Hepes pH 7.5, 0.2M $(NH_4)_2SO_4$, 30% PEG 6000) [26] was cryo-preserved in liquid nitrogen. The diffraction data were collected at SLS beamline X10SA.

Data were indexed and integrated using MOSFLM [27, 28]. Following computational steps were done using programs contained in CCP4 package [29]. The data were scaled in SCALA [30, 31]. Molecular replacement solution was found using Phaser [32] with SplC (PDB ID 2as9) as a search model. SplA model was manually built in resulting electron density maps. Restrained refinement of bond lengths and angles was imposed with Refmac 5.0 [33]. Throughout the refinement 5% of the reflections were used for cross-validation analysis [34], and the behavior of R_{free} was employed to monitor the refinement strategy. Water molecules were added using Arp/Warp [35] and subsequently manually inspected. All detailed features described in the text were validated using simulated annealing omit maps computed with CNS [36, 37]. The final models were deposited at Protein Data Bank under accession numbers: 2w7u and 2w7s. The data collection and refinement statistics are summarized in Table II.

Homology modeling

φ and ψ mainchain angles of a substrate were modeled as found in canonical inhibitors of serine proteases which bind in a manner resembling natural substrates [38]. Substrate binding surface of SplA was identified based on homology to S1 family proteases. The substrate model was manually docked into the substrate binding surface of the enzyme. Minor corrections of the mainchain disposition within the template derived from canonical inhibitors were performed to optimize recognized hydrogen bonds [39]. The side chains were added according to the identified consensus sequence recognized by SplA. Sidechain disposition was modeled based on structures of S1 family proteases complexed with synthetic and protein inhibitors. Following hydrogen atoms were added to the model using *AddH* module of UCSF Chimera package [40], potentials were assigned according to Amber 1994 force field [41] and

the energy of the complex was minimized using MMTK module of Chimera with default values [42]. Manual structure edition, structure analysis and molecular graphics were done using modules of Chimera and spdbv.

Bioinformatic analysis

Human (*Homo sapiens*, taxid: 9606) and staphylococcal (*S. aureus*, taxid:1280) proteomes were analyzed for potential cleavage sites with the deduced consensus sequence recognized by SplA (YLYS) using BLAST algroithm and Uniprot data base. Initial hits were manually examined to remove redundant sequences. The remaining proteins were analyzed for tissue and cellular localization (for a list of all nonredundant hits see supplementary materials table S1) to pinpoint those directly spacially colocalized with staphylococci during commensal colonization and/or infection.

Results

Production of recombinant SplA

Initially SplA was expressed as a GST fusion protein in *E. coli* and the tag was cleaved off with thrombin during purification procedure. Thus produced recombinant GS-SplA was inactive in zymographic assays using gelatin and β-casein as substrates. As was the case with related protease SplB [15, 16] the lack of the proteolytic activity could have been related to the presence of a Gly-Ser dipeptide, a remnant of the thrombin cleavage site, on the N-terminus of GS-SplA. To investigate this option, recombinant SplA was directed onto secretory pathway to ensure proper N-terminal trimming by signal peptidase. *Bacillus subtilis* expression plasmid pWB980 driving constitutive production and secretion of SplA was constructed and transformed into a knockout strain WB800 lacking majority of endogenous extracellular proteases. Indeed, obtained in this way, the recombinant SplA showed proteolytic activity in zymography indicating the importance of N-terminus for enzymatic activity. The system yielded around 10 mg purified protein per 1 liter of a starting culture. Thus produced protein was used in all further analyses.

Substrate specificity of SplA

To characterize SplA activity we have evaluated an array of synthetic *p*-nitroanilide substrates. Among tested compounds (see experimental procedures) only N-Suc-Ala-Ala-Pro-Phe-*p*Na was hydrolyzed, though very inefficiently. Nevertheless, using this substrate we estimated the pH optimum of the protease at pH 6.5 (Bis-Tris buffer). Moreover, we have demonstrated that amongst common low molecular weight serine protease inhibitors tested only 3,4-dichloroisocoumarin inhibited SplA.

The P1 substrate preference of SplA was determined by screening the enzyme activity against a panel of combinatorial libraries of synthetic fluorescent substrates with a general structure of Ac-Xaa-Xaa-Xaa-P1-aminomethylcoumarin. SplA exclusively hydrolyzed substrates with monocyclic aromatic residues in P1, namely tyrosine and to a lower extent phenylalanine (Figure 1.) The substrate specificity was further evaluated using MS and N-terminal microsequencing analysis of cleavage products of β-casein and carboxymethylated chicken lysozyme. The results confirmed SplA preference for phenylalanine and tyrosine in P1 (Table I). The cleavage kinetics of casein and lysozyme was however extremely slow requiring an excess of an enzyme and prolonged incubation times. Significantly, none of other tested native and denatured protein substrates or human serum serine protease inhibitors (for complete list see experimental procedures), all containing multiple Phe and Tyr residues were cleaved by SplA. The above results collectively suggest that SplA specificity is determined not only in P1, but also in further subsites.

To determine the preference of SplA for particular residues in P(n) as well as P(n)'

positions and to account for possible subsites cooperation we have applied a high-throughput CLiPS methodology [23]. A library presenting five consecutive randomized amino acids in a context of a cell surface protein was selected for SplA cleavage. The analysis demonstrated that SplA recognizes and cleaves a consensus sequence of four consecutive residues Trp/Tyr-Leu-Tyr-Thr/Ser (Table I).

Taking into account the P1 preference for monocyclic aromatic residues there are two possible cleavage sites in the consensus sequence. To determine which peptide bond is hydrolyzed and demonstrate that SplA can efficiently cleave proteins containing this sequential motif we have constructed a fusion protein of two globular partners with a linker containing the identified consensus amino acid sequence (YLYS). At 1:100 enzyme to substrate molar ratio, 160 μg of purified fusion protein (GST-YLYS-ScpB) was completely and specifically split into free GST and ScpB in 15 minutes (supplementary materials – Figure S2). Edman degradation sequencing of the released N-terminus revealed hydrolysis of the Tyr-Ser peptide bond within the consensus sequence.

Collectively, these data demonstrates that SplA is a highly selective protease which can cleave efficiently only substrates containing the Trp/Tyr-Leu-Tyr-Thr/Ser sequence motif.

Overall crystal structure of SplA protease

To explain narrow specificity of SplA at the atomic level we have crystallized and solved the enzyme structure. Two independently determined structures are discussed (Table II). A 2.43Å structure of an active SplA protease and 1.8Å structure of a protease reacted with phosphonate inhibitor. Both structures are essentially identical (the average RMSD of backbone atoms for all eight models equals 0.45) since the inhibitor has hydrolized during crystallization and is not accounted for by electron density. Therefore, the low resolution model is discussed mainly to demonstrate said identity, importance of which is stressed in the discussion. Unless otherwise indicated 1.8 Å structure is used to discuss the structural details.

The protein shows a chymotrypsin-like fold. SplA molecule consists of two domains (I and II), each of which is built of six anti-parallel β-strands folded into a β-barrel (Figure 2). The domains are connected via a stretched linker containing amino acids Asn95 through Val109. Domain I, although composed mostly of N-terminal residues, (Tyr15-Phe94) also contains a C-terminal fragment (Thr187-Lys200) encompassing an α-helix (Pro188-Asn197). Domain II is build of residues Lys110-Phe186 and the most N-terminal portion of the protein (Glu1-Ile7). The active site of the enzyme is located at the interface of the two barrels and consists of residues His39, Asp78 and Ser154 conserved in all enzymatically active chymotrypsin-like proteases.

The asymmetric unit contains four molecules. Superimposition of Cα atoms of those molecules yields average RMSD of 0.5 Å. Small differences are visible in positioning of surface loop Lys58-Gly63 which is characterized by some flexibility as evidenced by high B factors and week electron density, and in regions Thr87-Gly91 and Ala104-Ala107 which are affected by crystal packing interactions. However, most pronounced differences due to crystal packing are found in the region His39-Ser52. Residues Asn37 through Ile46 form a helix in models A and B which protrudes with its C-terminal part away the globular core of the molecule. In models C and D respectively residues 42-47 and 42-43 are not defined by electron density. The average B factors for residues 38-49 are higher than of surrounding residues. In all models the entire discussed region forms close contacts with surrounding molecules. Of note on top of the protrusion the hydrophobic side chain of Phe47 is buried in a shallow pocket of adjacent molecule.

Primary sequence analysis shows significant homology between SplA and other chymotrypsin-like serine proteases from *S. aureus* (38% with V8 protease and 27% with epidermolytic toxins ETA and ETB). Accordingly, the superimposition of the Cα trace of SplA with that of V8 protease yields RMSD of 1.0 Å (over 165 Cα atoms) and with that of epidermolytic toxins A and B of 1.1 Å (153 Cα atoms) and 1.1 Å (146 Cα atoms), respectively.

Catalytic Triad

The conformational restraints of catalytic triad, an essential element of catalytic machinery of S1 family serine proteases, are not satisfied in the structure of SplA. Sequence alignment and superposition of V8 protease, SplB and SplA models demonstrates the presence of the three essential residues (His39, Asp78 and Ser154 in SplA), nevertheless the crucial hydrogen bonding network between the sidechains of those residues is not present in SplA structure (Figure 3). Although, when whole molecules are superposed, the positions of Asp78 and Ser154 of SplA are similar to those of correspondig residues in V8 and SplB, His39 sidechain is differently positioned. The distances between N^{ε} of His39 and O^{γ} of Ser154 and between N^{δ} of His39 and O^{δ} of Asp78 are significantly longer. Such distances indicate only week or no hydrogen bonding as opposed to strong hydrogen bonding observed in canonical structures [43]. In its extreme in model B of 2.43 angstrom structure the N^{ε} of His39 is hydrogen bonded to a carbonyl oxygen of A122 of adjacent molecule and in model A of the same structure a water molecule is hydrogen bonded between N^{δ} of His39 and O^{δ} of Asp78. The reason for such atypical His 39 conformation as compared to type proteases of family S1 is evident from what was discussed in previous section. The position of 37-46 helix is affected by crystal packing and so is the placement of H39 imidazole ring, though His39 is not directly involved in intermolecular contacts (besides His39_B of 2.43 angstrom structure). Described unusual arrangement of catalytic triad histidine is not compatible with the standard mechanism of catalysis. However, since model B of 1.8 angstrom structure shows almost canonical hydrogen bonding and we unambigously demonstrated here that SplA is catalytically active against peptide substrates we conclude that the discussed phenomenon is an effect of crystal packing interactions.

Oxyanion hole

In S1 family of serine proteases the oxyanion hole stabilizes a tetrahedral transition-state intermediate of the carbonyl carbon atom of a scissile peptide bond by compensating the negative charge formed on the carbonyl oxygen. In contrast to what was previously reported for closely related enzymes: SplB protease [16] and ETs [44] the conformation and Pro151-Gly152 peptide bond in SplA supports a functional oxyanion hole similarly to previously reported structure of V8 protease. In SplA Pro151 has average main chain angles of φ = -44º and ψ = 134º, which are similar to those of corresponding Gly166 in V8 protease (φ = -50º and ψ = 130º) and different from those of corresponding Pro192 in ETA where φ = -48º and ψ = -40º and corresponding Ser154 in SplB (φ = -62º and ψ = -34º). The Gly 152 has main chain angles of φ = 110º and ψ = -26º, which are again similar to those of corresponding Gly167 in V8 protease (φ = 121º and ψ = -30º) and different from those of corresponding Gly193 in ETA where φ = -60º and ψ = -22º and corresponding Gly155 in SplB (φ = -63º and ψ = -17º). In all mentioned enzymes φ and ψ angles of following and proceeding residues correspond to those found in type S1 family proteases. The described conformation of Pro151-Gly152 peptide bond in SplA is stabilized by hydrogen bonding of carbonyl oxygen of Pro151 with the amide of Pro119-Lys120 peptide bond. Since inhibitor treatment might have resulted in the observed placement of Pro151-Gly152 peptide bond the 2.43 Å structure was carefully examined in the region in question. No evidence was found to support any alternative placement of the carbonyl oxygen of the described bond (for details see supporting information). Therefore, apart from the described distortion of catalytic triad histidine residue the catalytic machinery of SplA corresponds to that found in type proteases of S1 family.

Specificity pockets

Substrate specificity of family S1 proteases is commonly determined by pronounced S1-P1 interactions. Accordingly, the S1 pocket of SplA is an easily distinguishable, spacious, bowl-shaped cavity. It is delineated by loops I (Ala149-Ser154; [39]) and II (Gly174-Glu179)

and strands Leu169-Ser173 and Lys180-Phe182. The cavity is lined with non-polar side chains of Ala149, Leu169 and Cβ atoms of Asn153 and Ser178 and mainchain atoms of mentioned secondary structures save for the strand Lys180-Phe182 which is buried and only stabilizes the pocket. As such it closely resembles the S1 pocket of chymotrypsin and should easily accommodate bulky side chains of Tyr and Phe as determined in specificity assays. Homology modeling of substrate docking based on multiple available structures of synthetic peptide and canonical protein inhibitors in complex with S1 family proteases demonstrates that the side chain oxygen of Asn181 which reaches the bottom of the S1 pocket in SplA can possibly interact through hydrogen bond with Tyr (P1) side chain hydroxyl of the substrate (Figure 4). The modeling also defines the S2 subsite. The mainchain of the substrate adopts an extended beta-sheet conformation and interacts via hydrogen bonds with antiparallel beta-strand of the enzyme formed by residues Leu169-Gly172. The branched side chain of leucine preferred in P2 docks into a shallow cleft formed by the side chains of two catalytic triad residues, His39 and Asp78, and further flanked by side chains of Ala171 and Tyr170. Contrary to observed strong preference for aromatic side chains at P3 substrate position the S3 subsite is not clearly defined. Comparisons with structures of serine proteases-inhibitor complexes and energy minimization of SplA – YLY peptide complex give unambiguous results. It therefore seems likely that an induced fit mechanism may be involved.

Hydrogen bond network at the N-terminal

Precise N-terminal trimming is a prerequisite for activity of type proteases of chymotrypsin family. Similarly, we demonstrate here that the addition of two aminoacids at the N-terminus of SplA renders the protein inactive. A likely explanation is provided by the crystal structure. The N-terminal glutamate it is firmly positioned on the surface of the molecule through a network of hydrogen bonds (Figure 5). The mainchain amide and carbonyl oxygen of Glu1 are hydrogen bonded respectively mainchain carbonyl oxygen of Asp146 and mainchain amide of Tyr148 as in a parallel β-sheet. The N-terminus amide is further hydrogen bonded to a sidechain of Glu179 while Glu 1 backbone carbonyl forms an additional hydrogen bond with a water molecule which is buried in the structure and further coordinated by the mainchain amide of Asn3. The side chain of Glu1 forms a salt bridge with the side chain of Arg112 and further hydrogen bonds with a water molecule and the sidechain of Thr132. Asn3 begins a terminal strand of an antiparallel β-sheet. The tightly positioned residues of Glu1 and Asn3 hold in place the Lys2 residue which by itself makes no pronounced contacts. Any N-terminal extention is not sterically compatible with the described descirbed Glu1 positioning and thus requires breaking of the hydrogen bond network described.

Putative *in vivo* SplA targets

A search of human and staphylococcal proteomes with SplA consensus cleavage sequence suggests likely *in vivo* targets of the enzyme. Out of 90 identified human and 3 staphylococcal proteins (supplementary materials, table S1), after analysis of tissue and cellular context together with the bacteria localization during commensal colonization as well as infection, two emerge as likely targets: mucin-16 (Q8WX52) and olfactory receptor 5I1 (Q13606).

Discussion

Similarly to SplB [15], the proper N-terminal processing is necessary for the enzymatic activity of SplA. In a closely related V8 protease of S1B family the N-terminal amino group released by cleavage of propeptide penetrates into the S1 pocket and it is directly responsible for charge neutralization of P1 glutamic acid of a scissile peptide bond. Such mechanism follows the classical pathway of S1A family serine protease zymogen activation, the most

thoroughly studied examples of which include digestive tract proteases [45]. There, as in a multitude of other serine proteases the new N-terminus liberated upon proteolytic processing of a precursor is directly involved in a formation of a functional primary specificity pocket (Figure 5) [46-48]. Sequence alignment of SplA and V8 protease demonstrates that the polypeptide chain of mature SplA is four amino-acids shorter at the N-terminus than that of V8 protease, making similar penetration in the former enzyme unlikely. Indeed, analysis of the crystal structure of SplA reveals a lack of direct involvement of the N-terminus in the formation of the S1 pocket. Nevertheless, the artificially, N-terminally elongated forms of Spl proteases, Gly-Ser-SplB [15] and Gly-Ser-SplA (this work) show no activity in zymography.

The molecular mechanism by which the N-terminal extension controls activity of SplA protease is less obvious. The N-terminus is distant from either catalytic machinery or substrate recognition surface. Pronounced structural changes upon N-terminal elongation are unlikely since 1D NMR spectra of SplA and GS-SplA are almost identical (data not shown). Therefore, most probably only distinct conformational changes and possibly small rearangements in a network of hydrogen bonds regulate catalysis. Hydrogen bonding of Glu1 amide to a sidechain of Glu179, a residue close to P1 specificity pocket suggest influence on substrate recognition, however certainly the mechanism is distinct to that observed in type proteases of family S1. A structure of N-terminally elongated SplA (ex. GS-SplA) would probably help to clarify the mechanism though it may involve a substantial dynamic constituent which might be hard to sensibly model based on static crystallographic data.

The inactivity of N-terminally elongated forms of Spl proteases seems not only a cloning artifact but rather a physiologically relevant phenomenon. Most proteolytic enzymes are syntesized as inactive zymogens as a protection against unwanted proteolysis. Only at target sites the propeptides are removed to release activity. Spl proteases lack propeptides of any kind and are secreted directly in an active form being obviously harmless for the bacterium while outside the cell. The latter is similarly true for staphopains (*Staphylococcus arueus* cysteine proteases) which are found in a mature, proteolytically active form in the culture media. However, for efficient production of extracellular staphopains specific intracellular inhibitors are necessary. It was demonstrated that knocking out these inhibitors decreases the viability of the bacterium. This demonstrates that a small amount of enzyme is misdirected into cytoplasm [9]. One may suspect that similarly a small amount of every secretory protein and in particular of Spl proteases reamins inside the cell. This pool however contains a signal peptide at the N-terminus (which is cleaved off by a signal peptidase only upon secretion) and thus remains catalytically inactive. As such the signal peptide acts both as a secretion signal and a propeptide controlling the enzyme activity which likely explains the lack of demonstrated intracellular serine protease inhibitors in staphylococci.

The high substrate specificity of SplA protease, defined by three consecutive residues of a substrate, is uncommon among S1 family proteases. Homology modeling of substrate binding clearly defined the structural determinants of P1 and P2 preference. The strong propensity for aromatic P3 substrate residue could not have been satisfactory explained by modeling suggesting an induced fit mechanism might be involved. A crystal structure of SplA complexed with a substrate analogue is therfore desirable to clarify this issue.

Parallel to the classical subsite based substrate recognition, oxyanion hole distortion was suggested to take part in specificity determination of SplB [16]. In SplA the arrangement of the Ser154 and Gly152 amides is similar to that observed in type serine proteases and different from that in SplB therefore, similar mechanism is not involved here.

Many proteases produced by pathogenic bacteria play crucial role in virulence. The broad substrate specificity enzymes are generally involved in unspecific processes like acquisition of nutrients and dissemination (both associated with tissue damage). In turn, highly specific proteases often act as dedicated toxins exerting activity on a narrow subset of proteins or even on a single substrate. The narrow specificity of SplA suggests a similar role. Deciphering the role of specific proteases in virulence is most interesting, but also a highly

challenging undertaking. A search for physiological substrate of ETs took nearly 30 years to achieve [49]. Considering pronounced similarity of Spl proteases to staphylococcal virulence factors, location of the *spl* operon within the pathogenicity island [14], expression during staphylococcal endocarditis [11] and highly limited substrate specificity it is tempting to speculate that Spl proteases represent yet another group of *S. aureus* specific toxins. Analysis of human proteome with the deduced consensus sequence recognized by SplA points to two fascinating putative targets. Olfactory receptor (OR) 5I1 (Q13606), a member of a large family of transmembrane sensory proteins is clearly found in the primary colonization site of *S. aureus*. One may speculate that its cleavage may facilitate the colonization and/or persistence of *S. aureus* in the nares, or facilitate dissemination from this primary niche into the human body by yet unknown mechanisms. What is certain is that consensus cleavage site of SplB protease was found in another members of ORs family [16]. Another identified protein, mucin-16 (MUC16), possesses two putative cleavage sites for SplA. Mucin-16 is a heavily O-glycosylated transmembrane protein, which is expressed by the ocular surface epithelia in which it helps provide a disadhesive protection barrier [50]. *Staphylococcus aureus* is a leading cause of bacterial keratitis in adults that can result in irreversible corneal scarring and loss of visual acuity. To date mainly Alpha- and Beta-toxin of *S. aureus* were convincingly demonstrated as virulence factors in corneal disease. In the context of staphylococcal corneal infection, cleavage of MUC16 by SplA is worth detailed investigation. Although the above findings on possible physiological role of SplA protease are highly speculative, they open an exciting avenue for future research on related aspects of staphylococcal physiology.

Acknowledgements:

We acknowledge Sui-Lam Wong from the University of Calgary, Canada, and Jozef Oleksyszyn and Ewa Pietrusewicz from Wrocław University of Technology, Poland, respectively for a kind donation of *Bacillus subtilis* expression strain and plasmid, and phosphonate inhibitors.

This work was supported in part by funds of the Polish Ministry of Science and Higher Education (to JP), grants 13033 and N N301 032834 (to GD) and the European Social Fund and polish national budget in the framework of The Integrated Regional Operational Program (to GD). GD is also a recipient of The Foundation for Polish Science Scholarship for Young Scientists.

References

1 Lowy, F. D. (1998) Staphylococcus aureus infections. N Engl J Med **339**, 520-532
2 Noble, W. C. (1998) Skin bacteriology and the role of Staphylococcus aureus in infection. Br J Dermatol **139 Suppl 53**, 9-12
3 Archer, G. L. (1998) Staphylococcus aureus: a well-armed pathogen. Clin Infect Dis **26**, 1179-1181
4 Maskalyk, J. (2002) Antimicrobial resistance takes another step forward. Cmaj **167**, 375
5 Shorr, A. F. (2007) Epidemiology of staphylococcal resistance. Clin Infect Dis **45 Suppl 3**, S171-176
6 Lindsay, J. A. and Holden, M. T. (2004) Staphylococcus aureus: superbug, super genome? Trends Microbiol **12**, 378-385
7 Dubin, G. (2002) Extracellular proteases of Staphylococcus spp. Biol Chem **383**, 1075-1086

8 Redpath, M. B., Foster, T. J. and Bailey, C. J. (1991) The role of the serine protease active site in the mode of action of epidermolytic toxin of Staphylococcus aureus. FEMS Microbiol Lett **65**, 151-155

9 Shaw, L. N., Golonka, E., Szmyd, G., Foster, S. J., Travis, J. and Potempa, J. (2005) Cytoplasmic control of premature activation of a secreted protease zymogen: deletion of staphostatin B (SspC) in Staphylococcus aureus 8325-4 yields a profound pleiotropic phenotype. J Bacteriol **187**, 1751-1762

10 Bailey, C. J., Lockhart, B. P., Redpath, M. B. and Smith, T. P. (1995) The epidermolytic (exfoliative) toxins of Staphylococcus aureus. Med Microbiol Immunol **184**, 53-61

11 Rieneck, K., Renneberg, J., Diamant, M., Gutschik, E. and Bendtzen, K. (1997) Molecular cloning and expression of a novel Staphylococcus aureus antigen. Biochim Biophys Acta **1350**, 128-132

12 Reed, S. B., Wesson, C. A., Liou, L. E., Trumble, W. R., Schlievert, P. M., Bohach, G. A. and Bayles, K. W. (2001) Molecular characterization of a novel Staphylococcus aureus serine protease operon. Infect Immun **69**, 1521-1527

13 Ziebandt, A. K., Becher, D., Ohlsen, K., Hacker, J., Hecker, M. and Engelmann, S. (2004) The influence of agr and sigmaB in growth phase dependent regulation of virulence factors in Staphylococcus aureus. Proteomics **4**, 3034-3047

14 Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K. (2001) Whole genome sequencing of meticillin-resistant Staphylococcus aureus. Lancet **357**, 1225-1240

15 Popowicz, G. M., Dubin, G., Stec-Niemczyk, J., Czarny, A., Dubin, A., Potempa, J. and Holak, T. A. (2006) Functional and structural characterization of Spl proteases from Staphylococcus aureus. J Mol Biol **358**, 270-279

16 Dubin, G., Stec-Niemczyk, J., Kisielewska, M., Pustelny, K., Popowicz, G. M., Bista, M., Kantyka, T., Boulware, K. T., Stennicke, H. R., Czarna, A., Phopaisarn, M., Daugherty, P. S., Thogersen, I. B., Enghild, J. J., Thornberry, N., Dubin, A. and Potempa, J. (2008) Enzymatic activity of the Staphylococcus aureus SplB serine protease is induced by substrates containing the sequence Trp-Glu-Leu-Gln. J Mol Biol **379**, 343-356

17 Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun **27**, 157-162

18 Berthold, H., Frorath, B., Scanarini, M., Abnev, C.C., Ernst, B. and Northemann, W. (1992) Plasmid pGEX-5T: An alternative system for expression and purification of recombinant proteins Biotechnology Letters **14**, 245-250

19 Rzychon, M., Sabat, A., Kosowska, K., Potempa, J. and Dubin, A. (2003) Staphostatins: an expanding new group of proteinase inhibitors with a unique specificity for the regulation of staphopains, Staphylococcus spp. cysteine proteinases. Mol Microbiol **49**, 1051-1066

20 Wu, S. C. and Wong, S. L. (1999) Development of improved pUB110-based vectors for expression and secretion studies in Bacillus subtilis. J Biotechnol **72**, 185-195

21 Arvidson, S., Holme, T. and Lindholm, B. (1973) Studies on extracellular proteolytic enzymes from Staphylococcus aureus. I. Purification and characterization of one neutral and one alkaline protease. Biochim Biophys Acta **302**, 135-148

22 Snipas, S. J., Wildfang, E., Nazif, T., Christensen, L., Boatright, K. M., Bogyo, M., Stennicke, H. R. and Salvesen, G. S. (2004) Characteristics of the caspase-like catalytic domain of human paracaspase. Biol Chem **385**, 1093-1098

23 Boulware, K. T. and Daugherty, P. S. (2006) Protease specificity determination by

using cellular libraries of peptide substrates (CLiPS). Proc Natl Acad Sci U S A **103**, 7583-7588

24 Oleksyszyn, J. and Powers, J. C. (1989) Irreversible inhibition of serine proteases by peptidyl derivatives of alpha-aminoalkylphosphonate diphenyl esters. Biochem Biophys Res Commun **161**, 143-149

25 Oleksyszyn, J. and Powers, J. C. (1994) Amino acid and peptide phosphonate derivatives as specific inhibitors of serine peptidases. Methods Enzymol **244**, 423-441

26 Jancarik, J. and Kim, S. H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. J Appl Cryst **24**, 409-411

27 Leslie, A. G. (2006) The integration of macromolecular diffraction data. Acta Crystallogr D Biol Crystallogr **62**, 48-57

28 Leslie, A. G. W. (1992) Recent changes to MOSFILM package for processing film and image plate data. Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography

29 Collaborative Computational Project (1994). The CCP4suite:programs for protein crystallography. Acta Ctystallogr D Biol Crystallogr **50**, 760-763

30 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr **62**, 72-82

31 Evans, P. R. (1997) Scala. Joint CCP4 and ESF-EACBM Newsletter **33**, 22-24

32 McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr **63**, 32-41

33 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr **53**, 240-255

34 Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature **355**, 472-475

35 Lamzin, V. S. and Wilson, K. S. (1993) Automated refinement of protein models. Acta Crystallogr D Biol Crystallogr **49**, 129-147

36 Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system. Nat Protoc **2**, 2728-2733

37 Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr **54**, 905-921

38 Apostoluk, W. and Otlewski, J. (1998) Variability of the canonical loop conformations in serine proteinases inhibitors and other proteins. Proteins **32**, 459-474

39 Perona, J. J. and Craik, C. S. (1997) Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. J Biol Chem **272**, 29987-29990

40 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem **25**, 1605-1612

41 Cornell, W. D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules J. Am. Chem. Soc. **117**, 5179-5197

42 Hinsen, K. (2000) The molecular modeling toolkit: A new approach to molecular simulations. Journal of Computational Chemistry **21**, 79-85

43 Leskovac, V., Trivic, S., Pericin, D., Popowic, M. and Kandrac, J. (2008) Short hydrogen bonds in the catalytic mechanism of serine proteases. J Serb Chem Soc **73**, 393-403

44 Vath, G. M., Earhart, C. A., Monie, D. D., Iandolo, J. J., Schlievert, P. M. and Ohlendorf, D. H. (1999) The crystal structure of exfoliative toxin B: a superantigen

with enzymatic activity. Biochemistry **38**, 10239-10246

45 Khan, A. R. and James, M. N. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Sci **7**, 815-836

46 Fehlhammer, H., Bode, W. and Huber, R. (1977) Crystal structure of bovine trypsinogen at 1-8 A resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. J Mol Biol **111**, 415-438

47 Wang, D., Bode, W. and Huber, R. (1985) Bovine chymotrypsinogen A X-ray crystal structure analysis and refinement of a new crystal form at 1.8 A resolution. J Mol Biol **185**, 595-624

48 Bode, W. and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. Eur J Biochem **204**, 433-451

49 Amagai, M., Yamaguchi, T., Hanakawa, Y., Nishifuji, K., Sugai, M. and Stanley, J. R. (2002) Staphylococcal exfoliative toxin B specifically cleaves desmoglein 1. J Invest Dermatol **118**, 845-850

50 Blalock, T. D., Spurr-Michaud, S. J., Tisdale, A. S., Heimer, S. R., Gilmore, M. S., Ramesh, V. and Gipson, I. K. (2007) Functions of MUC16 in corneal epithelial cells. Invest Ophthalmol Vis Sci **48**, 4509-4518

Biochemical Journal Immediate Publication. Published on 28 Jan 2009 as manuscript BJ20081351

Table I. Experimentally determined SplA cleavage sites.

Residues corresponding to the consensus sequence are highlighted bold. Cleavage products of β-casein and carboxymethylated chicken lysozyme were identified by MS and Edman degradation, respectively.

P4	P3	P2	P1	P1'
Bovine β-casein				
I	H	P	F	A
V	E	P	F	**T**
F	L	**L**	**Y**	Q
Chicken carboxymethylated lysozyme				
S	T	D	**Y**	G
Synthetic pNa substrate				
A	A	P	F	pNa
CLiPS (Cellular Library of Peptide Substrates)[*]				
R	**W**	**L**	**Y**	(S)
(S)	**W**	**L**	**Y**	**S**
G	**W**	**L**	**Y**	(S)
I	**Y**	E	**Y**	A
L	**Y**	E	**Y**	(S)
V	**Y**	E	**Y**	(S)
V	**Y**	M	**Y**	(S)
V	**Y**	A	**Y**	**S**
F	**Y**	T	**Y**	**S**
L	**Y**	**L**	**Y**	G
A	F	**L**	**Y**	**S**
T	F	**L**	**Y**	(S)
I	V	**L**	**Y**	**T**
V	V	**L**	**Y**	**T**
Y	**W**	**L**	S	**T**
Y	**W**	M	N	**T**
Y	**W**	W	Y	**T**
A	**W**	**L**	Y	(S)
Consensus sequence				
	W/Y	**L**	**Y**	**T/S**
Fusion protein[*]				
(V)	**Y**	L	**Y**	**S**

[*] Residues in parentheses constitute the template (invariant region) of the CLiPS library and tested fusion proteins.

Table II. Data collection and refinement statistics.

1. Data collection		PDB ID 2w7u	PDB ID 2w7s
Space group		$P2_12_12_1$	$P2_12_12_1$
Cell constants:	a (Å)	73.99	72.62
	b (Å)	82.22	81.43
	c (Å)	134.19	132.02
Wavelenght (Å)		1.5418	1.0000
B-Factor (Wilson) (Å^2)		29.2	23.1
Resolution range (Å)		20-2.43 (2.56-2.43)	20-1.80 (1.90-1.80)
Completeness (%)		98.2 (97.2)	99.6 (100.0)
R_{merge} (%)		7.5 (36.5)	5.5 (21.6)
R_{meas} (%)		8.5 (41.8)	6.1 (24.0)
Observed reflections		146,329 (17,924)	387,591 (56,106)
Unique reflections		30,882 (4403)	72,964 (10,566)
I/σ(I)		16.8 (4.1)	19.6 (6.5)
Average multiplicity		4.7 (4.1)	5.3 (5.3)
2. Refinement			
Resolution (Å)		20 - 2.43	20-1.80
No. of reflection used		29,282	69,255
R-factor (%)		20.9	21.3
R_{free} (%)		24.8	24.4
Average B (Å^2)		30.7	24.6
RMS from ideal values:			
Bond lenghts (Å)		0.006	0.007
Bond angles (°)		0.632	1.015
Ramachandran statistics (%)			
Most favored regions		85.2	88.7
Additionaly allowed regions		14.1	10.7
Generously allowed regions		0.7	0.6
3. Content of asymmetric unit			
Average RMSD (C^α) of monomers (Å)		0.54	0.47
No. of protein molecules		4	4
No. of protein residues / atoms		797 / 6158	777 / 5861
No. of solvent molecules		36	247

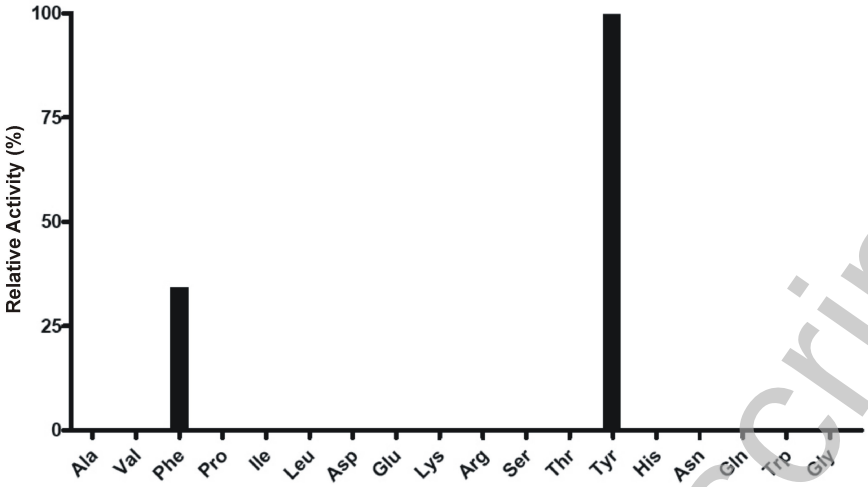

Figure 1 P1 substrate preference of SplA.

Specificity of recombinant SplA was determined using a set of positional scanning libraries of general structure Ac-Xaa-Xaa-Xaa-P1-AMC as described in the Experimental section. Vertical axis represents the rate of AMC production as a percentage of the maximum rate observed in each experiment. The *x* axis shows the positionally defined amino acids (P1) as represented by the three-letter code.

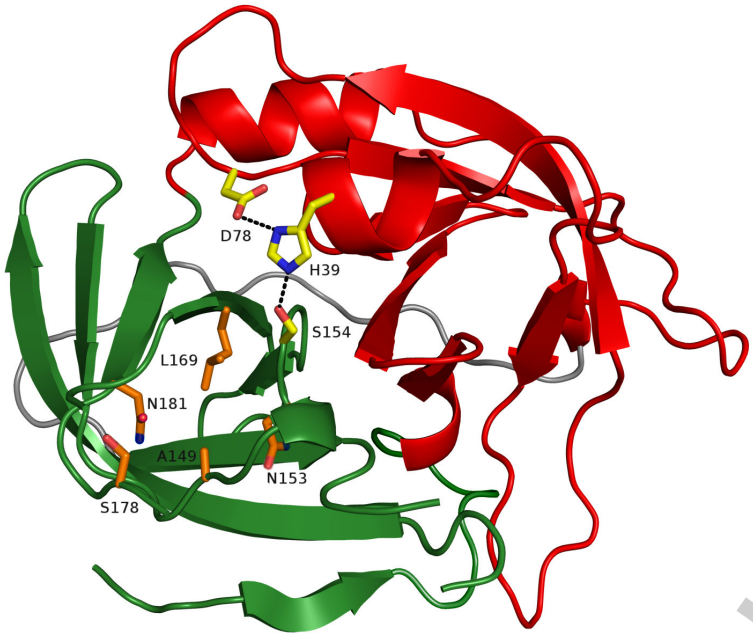

Figure 2 Overall fold of SplA.

SplA shows a typical chymotrypsin like fold. Domains I and II are shown respectively in red and green. A connecting linker of 15 residues (Asn95 through Val109) is depicted in grey. The catalytic triad residues His39, Asp78, and Ser154 are shown in yellow. Side chains building the S1 pocket (Ala149, Asn153, Leu169, Ser178 and Asn181) are coloured orange. Secondary structures were assigned using DSSP (Kabsch and Sander, 1983, Biopolymers 22:2577-2637)

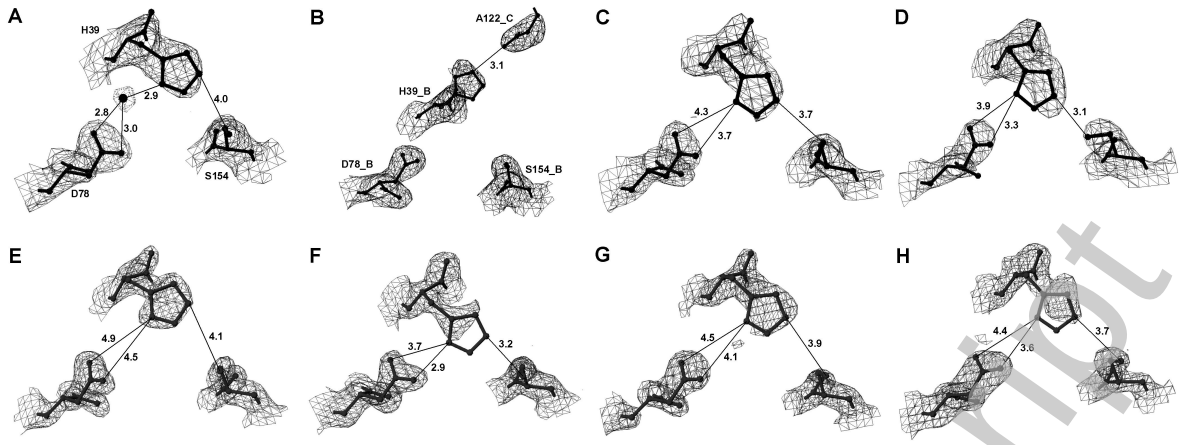

Figure 3. Atypical conformation of catalytic triad histidine.

Conformation of histidine 39 in SplA does not support the classical hydrogen bonding of the Asp-His-Ser catalytic triad of S1 serine proteases. The atypical H39 conformation results from dense crystal packaging in the vicinity o the active site. Distances between respecitve atoms are to long for hydrogen bond formation. In the extremes in panel A a water molecule is held between D78 sidechain and imidazole ring of H39 and in panel B H39 is hydrogen bonded with A122 carbonyl of adjacent molecule. Closest to the canonical conformation are molecules shown in panels D and F.

Panels A through D respectively models A to D of 2w7u. Panels E through H respectively models A to D of 2w7s. All panels: Thick solid lines – refined model interpretations of experimental data. Thin solid lines and numerals - distances in angstroms between respective atoms. Solid line meshes: F_o-F_c maps at the level of 4σ (panels A-D) and 3.5σ (panels E-H). H39, D78, S154 of the analysed model and all residues in the vicinity of 3.5Å were omitted from the structure prior to simulated annealing and structure factor calculation. Additionally in panel A: Dotted line mesh: F_o-F_c map at the level of 3.5σ. All water residues were omitted from the structure prior to simulated annealing and structure factor calculation.

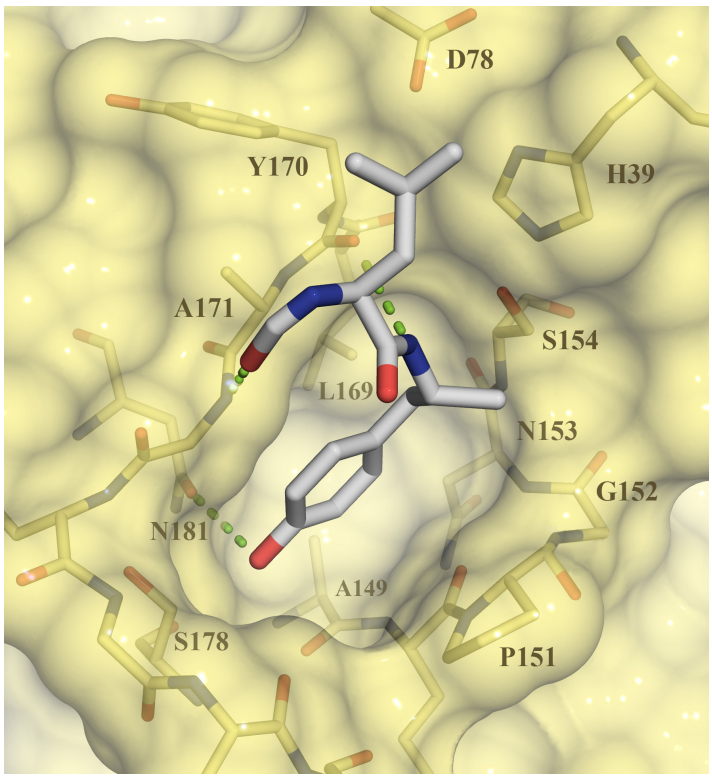

Figure 4 Putative binding mode of a consensus substrate to SplA.

Residues P1 (Tyr) and P2 (Leu) of a consensus substrate (foreground, stick model, Corey, Pauling, Koltun (CPK) coloring) docked to SplA (surface representation - yellow). Residues of SplA involved in substrate binding and catalysis are highlighted in CPK (background). Hydrogen bonds between the substrate and the enzyme are shown as green dotted lines. The likely binding mode of P3 residue was not evident from modeling studies and therefore is not depicted here.

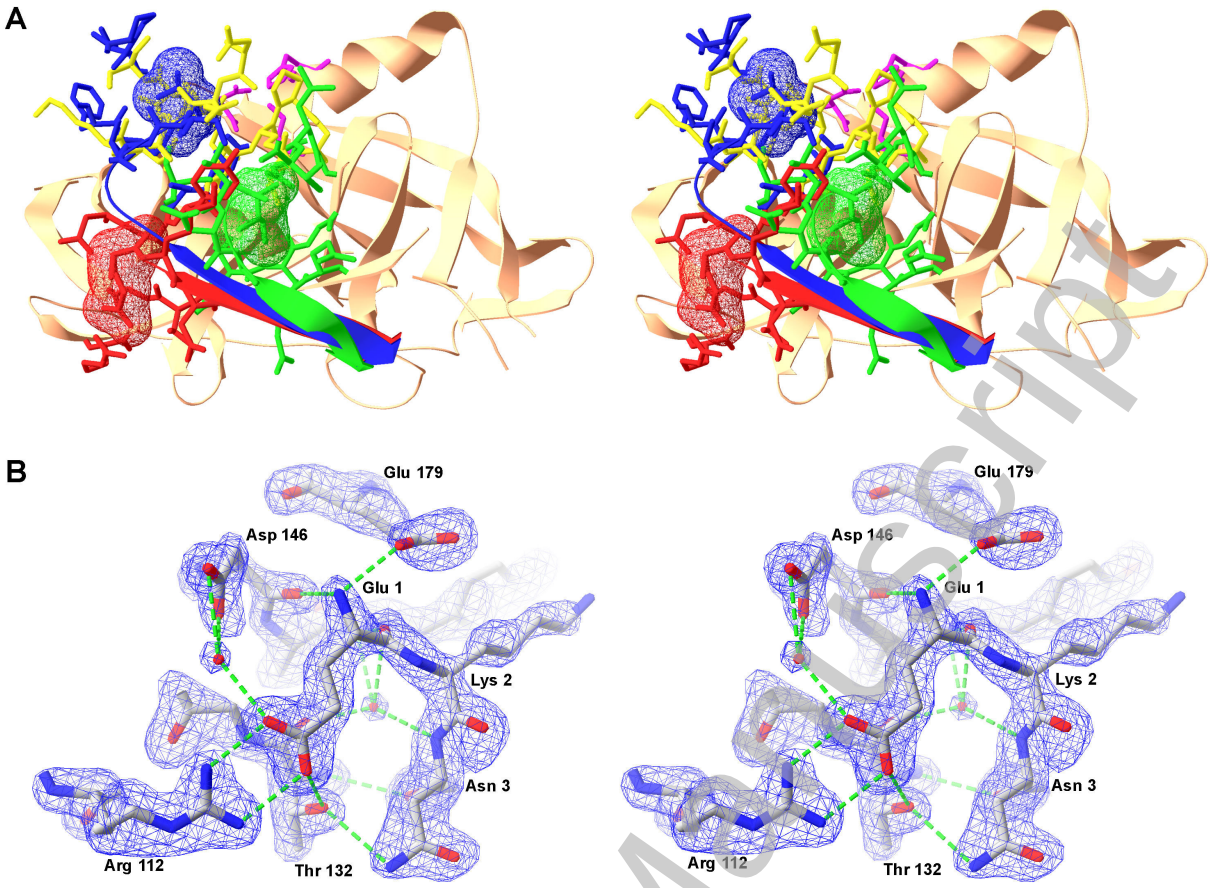

Figure 5 The role of N-terminal residue in the structures of mature SplA, V8 and chymotrypsin.

(A) Direct involvement of the N-terminal residue in P1 specificity pocket formation in V8 protease (1qy6; blue) and chymotrypsin (1afq; green). In SplA (2w7s; red) the N-terminal residue does not directly interact with P1 specificity pocket. Shown are the N-terminal residue (mesh) and all residues in vicinity of 4Å (stick model) overlaid over SplA structure (light brown, ribbon model; P1 specificity pocket forming residues – yellow, stick model, catalytic triad residues – magenta, stick model) (B) Hydrogen bond network at the N-terminus of SplA. All capable atoms of the N-terminal Glu1 of mature SplA are involved in hydrogen bonding interactions. Shown is the F_o-F_c map at the level of 3.5σ. Glu1 (model B, 2w7s) and all residues in the vicinity of 4Å were omitted from the structure prior to simulated annealing and structure factor calculation. The interpretation of shown map region is depicted as a stick model (CPK colouring). Hydrogen bonds are labeled as green dotted lines.