Crystal structure of the zymogen form of the group A *Streptococcus* virulence factor SpeB: An integrin-binding cysteine protease

Todd F. Kagawa*, Jakki C. Cooney†, Heather M. Baker*, Sean McSweeney‡, Mengyao Liu§¶, Siddeswar Gubba§, James M. Musser§¶, and Edward N. Baker*‖**

*School of Biological Sciences and ‖Department of Chemistry, University of Auckland, Private Bag 92–019, Auckland, New Zealand; †Institute of Molecular Biosciences, Massey University, Palmerston North, New Zealand; ‡European Molecular Biology Laboratory Grenoble Outstation, 6 Av. J. Horowitz, Grenoble 38042, France; §Institute of Human Bacterial Pathogenesis, Department of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030; and ¶Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 903 South 4th Street, Hamilton, MT 59840

Communicated by Richard M. Krause, National Institutes of Health, Bethesda, MD, December 16, 1999 (received for review October 25, 1999)

Pathogenic bacteria secrete protein toxins that weaken or disable their host, and thereby act as virulence factors. We have determined the crystal structure of streptococcal pyrogenic exotoxin B (SpeB), a cysteine protease that is a major virulence factor of the human pathogen *Streptococcus pyogenes* and participates in invasive disease episodes, including necrotizing fasciitis. The structure, determined for the 40-kDa precursor form of SpeB at 1.6-Å resolution, reveals that the protein is a distant homologue of the papain superfamily that includes the mammalian cathepsins B, K, L, and S. Despite negligible sequence identity, the protease portion has the canonical papain fold, albeit with major loop insertions and deletions. The catalytic site differs from most other cysteine proteases in that it lacks the Asn residue of the Cys-His-Asn triad. The prosegment has a unique fold and inactivation mechanism that involves displacement of the catalytically essential His residue by a loop inserted into the active site. The structure also reveals the surface location of an integrin-binding Arg-Gly-Asp (RGD) motif that is a feature unique to SpeB among cysteine proteases and is linked to the pathogenesis of the most invasive strains of *S. pyogenes*.

Streptococcus pyogenes is a common but potentially deadly human bacterial pathogen. This Gram-positive organism generally causes only mild throat infections (strep throat), but it is also responsible for sudden, devastating invasive infections that can lead to death within hours (1). These invasive illnesses include streptococcal toxic shock syndrome and necrotizing fasciitis. *S. pyogenes* secretes a variety of proven or putative virulence factors that enable it to invade or otherwise detrimentally affect a human host. Among these protein toxins are a group of superantigens known as streptococcal pyrogenic exotoxins A, C, H, etc. (SpeA, SpeC, SpeH) (2) and a cysteine protease known as streptococcal cysteine protease or SpeB (reviewed in ref. 3). The superantigens are variably present among disease isolates of *S. pyogenes*, but SpeB is present in all isolates of *S. pyogenes* and can be the predominant extracellular protein, accounting for up to 95% of total secreted protein (3–5).

The importance of SpeB for *S. pyogenes* virulence has been proven in a mouse model by using isogenic strains with the cysteine protease inactivated by genetic manipulation (6). SpeB appears to contribute to *S. pyogenes* pathogenesis in several ways, including cleavage of host proteins and release of bacterial proteins from the cell surface. The streptococcal protease cleaves human fibronectin and vitronectin, two abundant extracellular matrix proteins involved in maintaining host tissue integrity (5). It also cleaves human IL-1β precursor to form bioactive IL-1β (7), processes monocytic cell urokinase receptor (8), and releases active kinins from kininogen (9). Its activation of a 66-kDa host matrix metalloproteinase may contribute to the extensive soft tissue destruction observed in many patients (10). The enzyme has fibrinolytic activity and causes myocardial necrosis when injected into rabbits (11). It also has been shown to trigger apoptosis (12), although the exact role of this process in pathogenesis is not known. Importantly, most strains of *S. pyogenes* that are associated with severe invasive disease express a variant of SpeB that binds to integrins, which may be linked to their pathogenesis (13).

Like many proteases, the streptococcal enzyme is secreted and folded as an inactive precursor, or zymogen (zSpeB) form (4, 14). The 40-kDa (371 amino acid residues) SpeB precursor is converted by autolytic cleavage of the N-terminal 118 amino acid residues (the prosegment), to generate the mature, active protease (mSpeB) of 253 residues (27.6 kDa). Many cysteine proteases from animal, plant, and microbial sources have been characterized (15). However, although mSpeB has broadly similar specificity to papain and has Cys, His, and Trp residues similarly positioned in its amino acid sequence (15, 16), its lack of any overall sequence identity puts it in a separate, unique class (15), apparently unrelated to any protease of known three-dimensional structure.

Here we present the three-dimensional structure of the precursor form of SpeB, determined at high resolution by x-ray crystallography. Surprisingly, despite negligible sequence identity, we find that the protease is structurally homologous with proteins of the papain superfamily (17) that includes plant enzymes such as papain and actinidin, and the mammalian cathepsins B, K, and L. However, the crystal structure reveals that the prosegment and its mode of inactivation of the protease are strikingly different when compared with the proforms of the cathepsins (18–20). The crystal structure of this proven *S. pyogenes* virulence factor provides information that will be useful for inhibitor design and immunoprophylaxis strategies.

Materials and Methods

Expression, Purification, and Crystallization of SpeB. The Cys-47–Ser (C47S) mutant zymogen form of SpeB (Cys-192–Ser when numbered from the start of the signal sequence; ref. 21) was expressed in *Escherichia coli* strain BL21 containing plasmid pGM5 and purified as described (21, 22). The selenomethionyl

Abbreviations: Spe, streptococcal pyrogenic exotoxin; zSpe, zymogen form of Spe; mSpe, mature form of Spe; NCS, noncrystallographic symmetry; PBL, prosegment binding loop.

Data deposition: The structure factors have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID code 1dki).

**To whom reprint requests should be addressed. E-mail: Ted.Baker@auckland.ac.nz.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.040549997. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.040549997

Table 1. Statistics for data collection and processing, structure determination, and refinement

Data collection, processing and phasing	λ_1 (0.9791 Å)	λ_2 (0.9793 Å)	λ_3 (0.9315 Å)	λ_4 (0.9788 Å)	C47S (0.9315 Å)
Resolution, Å	30.0–2.4	30.0–2.4	30.0–2.4	30.0–2.4	30.0–1.6
Unique reflections	56,207	56,253	52,939	55,210	168,512
Multiplicity	3.9	4.0	2.0	3.8	2.1
Completeness (final shell)* (%)	93.7 (79.8)	93.8 (80.0)	88.3 (65.9)	92.0 (79.3)	83.3 (48.1)
R_{sym}† (final shell)*	0.069 (0.197)	0.062 (0.175)	0.043 (0.125)	0.059 (0.201)	0.044 (0.279)
$<I/\sigma>$ (final shell)*	14.7 (3.4)	16.3 (4.3)	15.9 (4.6)	17.3 (4.0)	14.2 (2.1)
Phasing power‡ isom. (anom)	5.61 (2.43)	4.34 (1.88)	— (1.66)	3.30 (2.21)	
R_{cullis}‡ isom. (anom)	0.40 (0.57)	0.41 (0.65)	— (0.62)	0.49 (0.60)	

Refinement statistics (30.0–1.6 Å)			
Number of reflections	168,900/16,726	Protein atoms	Mol A 2306; Mol B 2429; Mol C 2591; Mol D 2575
R/R_{free} (%)	21.9/24.6	Solvent	639 water molecules, 4 sulfate ions
rms deviations (bonds/angles)	0.009/1.4	Mean *B* factors	Protease portion 33.7 Å^2; Pro-region 53.1 Å^2

*The final shell was 2.53–2.4 Å for the data at wavelengths λ1, λ2, λ3, and λ4, and 1.7–1.6 Å for the C47S data.
†$R_{sym} = \Sigma|I - <I>|/\Sigma<I>$ where *I* is the observed intensity and $<I>$ is the average intensity for equivalent reflections.
‡Phasing power and R_{cullis} as defined in SHARP (27).

derivative of recombinant C47S zymogen was made by using *E. coli* strain DL41 (Met−) containing plasmid pGM5 and grown in a chemically defined glucose medium supplemented with selenomethionine in place of methionine (23).

Crystals of both the C47S zymogen and its SeMet-substituted form were grown in hanging drops at room temperature by mixing 1-μl drops of protein solution [14.6 mg/ml protein, in 20 mM Tris·HCl (pH 8.0), 0.02% azide] and reservoir solution (16% PEG 4000, 4% PEG 8000, 0.2 M ammonium sulfate, 0.1 M formic acid, pH 4.2). The crystals were monoclinic, space group $P2_1$, with unit cell dimensions $a = 46.7, b = 116.6, c = 144.7$ Å, $\beta = 94.2°$, and four molecules per asymmetric unit (solvent content 50%). They could only be flash-frozen after titration of 2-methyl-2,4-pentanediol (MPD) into the mother liquor, in increments of 2% MPD per hr, to a final concentration of 12% (vol/vol) MPD.

Data Collection. Diffraction data were collected at 100 K, on beam line ID14–4 at the European Synchrotron Radiation Facility, Grenoble, France. Multiwavelength anomalous diffraction data were collected at four wavelengths on a single, randomly oriented crystal of the SeMet-incorporated C47S mutant zymogen. Wavelengths were chosen from the x-ray fluorescence spectrum of a test crystal: λ1 and λ2, selected to maximize the imaginary and real anomalous scattering components, respectively, and λ3 and λ4 from below the Se K-absorption edge. A high-resolution data set also was collected from a single crystal of the unmodified C47S zymogen at λ3. Data were integrated, scaled, and merged by using DENZO and SCALEPACK (24) and SCALA (25), giving a four-wavelength SeMet data set to 2.4-Å resolution and a C47S mutant data set to 1.6-Å resolution (Table 1).

Structure Determination and Refinement. The structure was solved by multiwavelength anomalous diffraction methods. Positions for 23 of the possible 24 selenium atoms in the asymmetric unit were determined with SOLVE (26) and further refined with SHARP (27). The noncrystallographic symmetry (NCS) operators used in the initial map averaging were calculated from the refined selenium positions. Initial phases from SHARP were solvent-flattened and averaged with DM (25). The new experimental phases, to 2.4-Å resolution, were extended to the limits of the higher-resolution C47S data and input into WARP (28) for automatic electron density tracing and model building. This procedure gave an initial model comprising 656 total residues, with 218 residues for the most complete molecule in the asymmetric unit. The other three molecules in the asymmetric unit then were generated from the most complete model, using new NCS operators from the WARP model. Further model building into the 2.4-Å experimental map was performed with O (29). The map quality (Fig. 1) readily allowed the majority of the protein sequence to be fitted.

The model was refined by using the 1.6-Å C47S data set. Refinement was with CNS (30), with NCS restraints, using the maximum likelihood target with bulk solvent correction and periodic model rebuilding using O (29). NCS restraints were removed near the end of the refinement, as differences became apparent between the four molecules (A to D); both *R* and R_{free} then dropped a further 1%. For the final model *R* and R_{free} (ref 31; 10% of data, randomly chosen) were 21.9% and 24.6%, respectively, for data in the range 30 to 1.6 Å. The model conforms well with expected protein geometry; the rms deviation of bond lengths from standard values (32) is 0.009 Å, and 90.9% of residues are in the most favored regions of a Ramachandran plot, with no outliers. Refinement and model details are in Table 1. In the following discussion the prosegment residues are numbered 1p-118p and the protease residues 1–253.

Results and Discussion

The structure reported here, at 1.6-Å resolution, is for the proenzyme form of SpeB (zSpeB), after mutation of the catalytically essential Cys-47 to serine to prevent autolysis and activation. This structure is shown in Fig. 2*A*. Several regions of the polypeptide are undefined in the electron density and are assumed to be flexible. These regions are the same in the four independent molecules in the crystal asymmetric unit, although molecules C and D have been more completely modeled (Table 1). The proregion is markedly more flexible than the protease portion, as judged by *B* factors (Table 1) and has three undefined regions (residues 1p-3p, 22p-34p, and 114p-118p) compared with one in the protease portion (residues 233–243); the latter is a glycine-rich segment with six Gly in 11 residues.

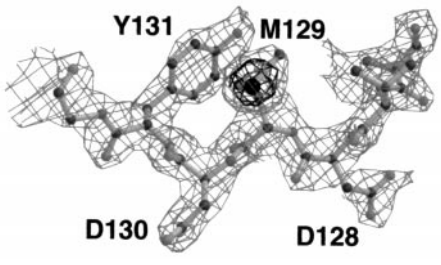

Fig. 1. Section of the 2.4-Å experimentally phased, NCS-averaged electron density map, contoured at 1.5 σ and 6 σ.

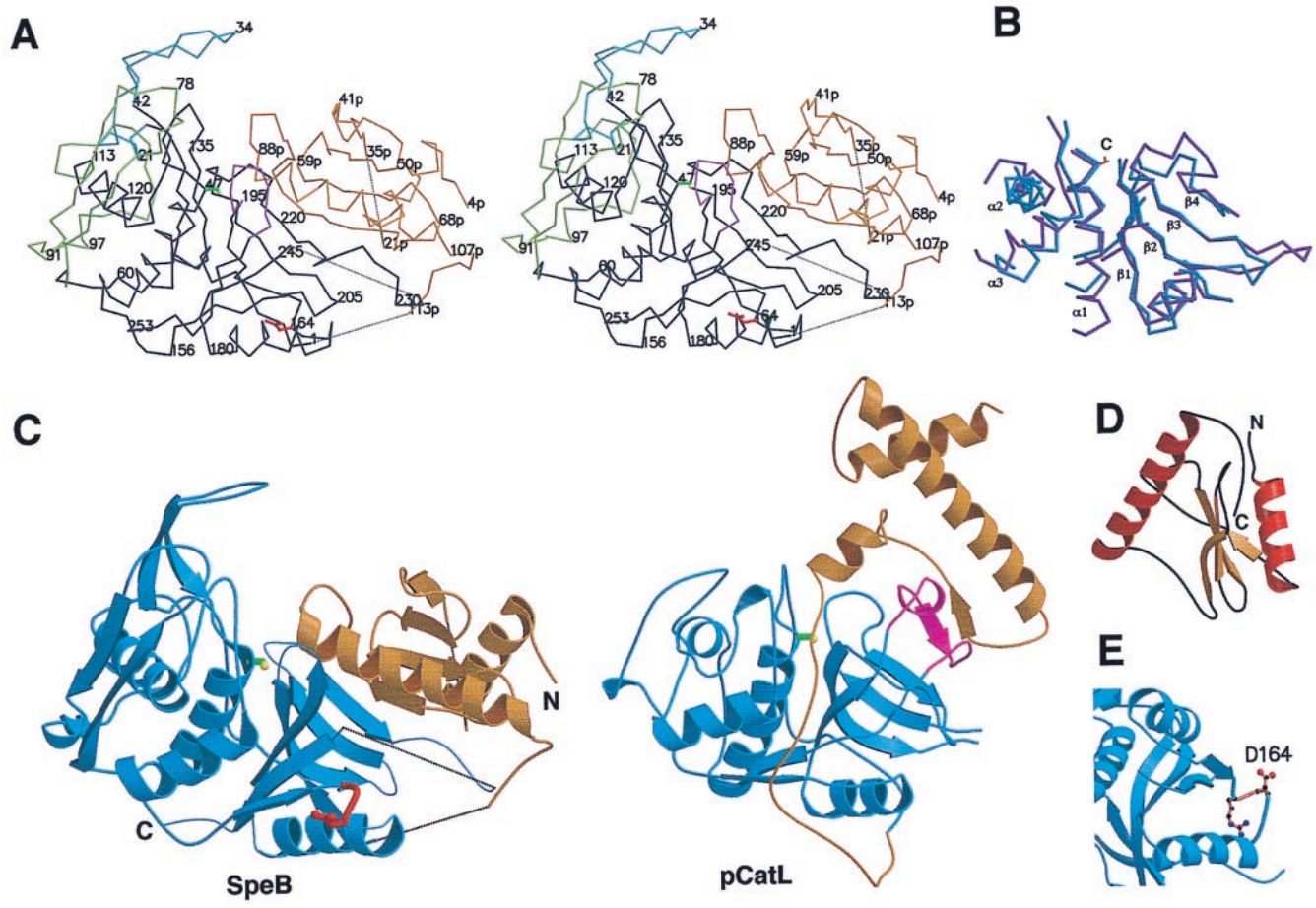

Fig. 2. Structure of SpeB and comparison with other papain-family proteins. (*A*) Stereo Cα plot of SpeB. The prosegment is colored gold and the protease domains black, except for the two N-domain insertions 19–42 (the finger loop, blue) and 66–113 (green) and the truncated β1-β2 C-domain loop 187–195 (magenta). Asp-164 of the RGD loop is red and the Ser-47 side chain yellow. Dotted lines indicate flexible regions. (*B*) Superposition of the core secondary structures of mSpeB (blue) and actinidin (magenta). The location of the catalytic cysteine (C) is shown. (*C*) Ribbon diagrams of SpeB (*Left*) and procathepsin L (*Right*), showing their similar protease domains (blue) and different prosegments (gold). The RGD loop of SpeB is red. The PBL, present in other papain-family enzymes but not SpeB, is colored magenta in procathepsin L. (*D*) Folding of the prosegment. (*E*). RGD loop (gold) of SpeB, showing the exposed Asp-164. These and other figures were produced with MOLSCRIPT (33) and RASTER3D (34).

Structure of the Protease Region. Despite a lack of significant amino acid sequence identity, the crystal structure clearly establishes that SpeB belongs to the papain superfamily of cysteine proteases (17). This family includes numerous plant enzymes, typified by papain and actinidin, the mammalian cathepsins B, K, and L, and others as diverse as the trypanosomal enzyme cruzain and the household dust mite allergen, derP1. SpeB is the most structurally divergent member of this superfamily and the first to be reported from a prokaryotic organism.

The protease portion of SpeB (mSpeB, residues 1–253), has the two-domain fold of other papain-family enzymes, with an N-terminal domain (residues 17–159) folded around three α-helices and a C-terminal domain (residues 1–16 + 160–253) based on a four-stranded antiparallel β-sheet (Fig. 2). Taking actinidin (35) as a representative family member, all of its major secondary structural elements are also present in mSpeB and have similar positions and orientations (Fig. 2*B*), although the N-terminal domain helices differ substantially in length. When the mSpeB and actinidin structures are superimposed based on this structurally conserved core, 93 Cα atoms (44% of the total for actinidin) match, with an rms difference in atomic positions of 2.0 Å.

Outside the conserved core, mSpeB has substantial insertions and deletions relative to actinidin, the cathepsins, and other papain family members. One significant deletion is in the surface loop that joins strands β1 and β2 on the C-terminal domain. This loop is severely truncated in SpeB, with important consequences for the association of the protease with its proregion (see below). Elsewhere, there are two large insertions in the N-terminal domain. The first is a 24-residue loop (residues 19–42) with an irregular antiparallel β-hairpin structure that is inserted between the catalytically important residues Gln-17 and Cys-47; this loop stretches out like a finger from the molecule (Fig. 2*A*). The second is a 48-residue insertion (residues 66–113) between helices α1 and α2, which forms a long β-hairpin that folds over the N-terminal domain surface and also substantially extends helix α2. Thus although the core of the protease portion conforms to the papain-family fold, the surface elaborations are very different. The structurally conserved regions of the molecule show 20% sequence identity between mSpeB and actinidin (20 of 97 residues identical) but elsewhere sequence identity is negligible, giving an overall identity of only 10%.

Structure of the Prosegment. In contrast to the protease portion of SpeB, the prosegment has a fold (Fig. 2*D*) that is unique among the prosegments of other proteases, as judged by a search of the current structural database. The zSpeB prosegment is based on a central four-stranded antiparallel β-sheet (residues 37p-80p)

with two helices (7p-21p and 87p-107p) that pack against either face of the sheet. A substantial hydrophobic core is enclosed between the β-sheet, the helix 87p-107p, and their associated loops, implying that the prosegment has a stable fold that would exist independently of its association with the protease portion.

The prosegments of other papain-like cysteine proteases fall into two groups, typified by cathepsins K and L (19, 20) whose prosegments of about 100 residues are mostly helical (Fig. 2*C*), and by cathepsin B with a mostly unstructured and smaller (60-residue) prodomain (18). Neither prosegment resembles that of SpeB. Our observation that the protease portion retains a strong structural similarity with other papain-family enzymes whereas the prosegment is quite different would be consistent with gene fusion of these components, as suggested (15), and identifies SpeB as a member of a third subfamily within the papain superfamily. The prosegment serves to inhibit enzyme activity until maturation, and at least in some cysteine proteases is essential for correct folding (36). These functions apparently can be achieved in several different ways.

Processing of zSpeB *in vitro* proceeds through a number of cleavages, occurring after Lys-26p, Asn-41p, Lys-101p, Ala-112p, and Lys-118p, with cleavage after Lys-26p giving the most prominent early processing intermediate (37). Except for Lys-101p, which belongs to the well-defined α-helix 87p-107p, all other processing sites correspond to regions of high flexibility in the prosegment structure that should provide favorable targets for intermolecular proteolysis.

Association Between Prosegment and Protease. The zSpeB prosegment is intimately associated with the protease portion, with a large buried surface between the two components (1,150 Å^2 from each, calculated with AREAIMOL, ref. 25); this interface represents 10% of the protease surface and 20% of the prosegment surface. The association almost exclusively involves the protease C-terminal domain (Fig. 2 *A* and *C*). First, the four-stranded β-sheet of the prosegment hydrogen bonds to the four-stranded β-sheet of the protease domain, with residues 74p-80p antiparallel to residues 217–223. This creates a single eight-stranded β-sheet that forms an extended core of the protease C-terminal domain and the prosegment. Second, the long amphipathic α-helix 87p-107p of the prosegment packs against this β-sheet, contributing much of the buried surface between the prosegment and protease domains. Overall, the association is primarily hydrophobic, with just 13 hydrogen bonds, mostly involving the β-sheet, and no salt bridges.

Although the prosegment has a completely different fold from those reported for other cysteine proteases, there are some striking parallels. In procathepsins B, K, and L (18–20) the prosegment packs against a surface loop of the C-terminal domain (residues 138–155 in procathepsin L), referred to as the prosegment binding loop (PBL). Despite the truncation of this loop in SpeB (Fig. 2*C*), the prosegment binds in the same place, packing against the underlying β-sheet instead. A second feature of the other cysteine proteases is that many of the buried residues are aromatic. This is true also of zSpeB (Fig. 3), in which the association buries five aromatic residues from the proregion and seven from the protease.

A critical difference exists, however, in the way the prosegment inserts into the protease active site cleft and inactivates the enzyme. In other cysteine protease zymogens, an extended strand of the prosegment runs the full length of the cleft, in a direction opposite to that of a natural substrate, and inserts a side chain into the S2 specificity pocket of the enzyme (18–20). In contrast, in zSpeB residues 87p-90p (the first turn of the long prosegment α-helix) insert into the active site close to Trp-214. The rest of the prosegment polypeptide then follows a completely different course from that in the other proenzymes (Fig. 2*C*).

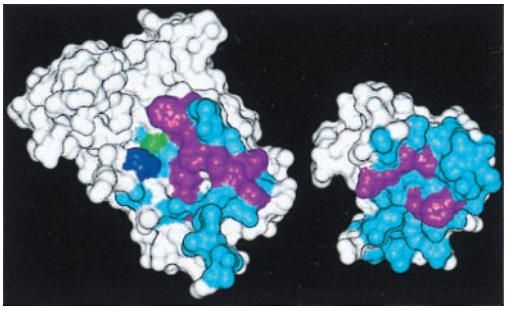

Fig. 3. Association of the SpeB prosegment and protease domains. The predominantly hydrophobic contact surfaces of the prosegment (*Right*) and protease (*Left*) are colored blue, with aromatic residues in purple. His-195 and Ser(Cys)-47 are blue and green, respectively. Hydrophobic residues involved include Phe-61p, Pro-71p, Tyr-76p, Phe-82p, Ile-90p, Phe-93p, Met-94p, and Tyr-97p from the prosegment and Tyr-186, Phe-197, Phe-207, Trp-212, Trp-214, Val-217, Gly-220, Phe-221, Phe-222, and Ala-226 from the protease.

Inhibition by the Prosegment and Implications for the Protease Active Site. The active site region of mSpeB has several features that confirm SpeB as a member of the papain superfamily. Ser-47 is situated at the N terminus of helix α1, at the interface between the two protease domains, with a conformation identical to that of Cys-25 in actinidin (Fig. 4). In the following discussion we assume that Cys-47 in wild-type mSpeB would be similarly oriented. Adjacent is Gln-17, analogous to Gln-19 in actinidin; this residue, with the amide NH of Cys-47, forms a potential oxyanion hole like that in actinidin. In the crystal structure of zSpeB a sulfate ion is bound at this site, hydrogen-bonded to Gln-17 Nε2 and Cys-47 NH. Also adjacent are Trp-214, analogous to a conserved Trp that is present in the active sites of all papain family members, and His-195, shown by mutagenesis to be catalytically essential (38).

Three striking differences exist, however, that relate to SpeB activity and inhibition. First, the active site histidine (His-195) is flipped out of the active site, away from Cys-47, by the insertion of the proregion, preventing any Cys. . . His interaction. The key residue in this disruption is Asn-89p, whose side chain inserts deep into the active site cleft, at a position that corresponds to the S1′ subsite of other papain-family enzymes, hydrogen-bonded to Trp-214 Nε1 and the peptide oxygens of Ala-196 and Trp-212 (Fig. 4). These hydrogen bonds are short and geometrically favorable, indicating a high degree of specificity. Asn-89p thereby blocks His-195 from assuming its catalytically competent position. Proteolytic removal of the proregion would allow His-195 to flip back into the active site. Most of this movement can be achieved by a simple torsional rotation about the His-195 Cα-Cβ bond, but some movement of the 187–195 loop to which it is attached also may be involved.

Secondly, SpeB has a simple Cys. . His pair, instead of the usual Cys. . His. . Asn triad found in almost all enzymes of the papain superfamily (17). Mutagenesis of papain (39) has shown, however, that the Asn side chain is not essential for activity and probably has a primarily orientational role. Different functional groups fill this role in other cysteine proteases, for example, an Asp side chain in foot-and-mouth disease virus leader protease (40), a Glu side chain in pyroglutamyl peptidase I (41), and a peptide carbonyl in IL-1β converting enzyme (ICE) (42). Nevertheless, this is an unusual difference for a member of the papain superfamily. In SpeB, Trp-212 replaces the Asn of other papain-family enzymes. Its side chain is oriented away from the active site to participate in the hydrophobic interactions with the proregion; intriguingly, the 212–213 peptide is flipped relative to actinidin, such that the Trp-212 carbonyl group is oriented into the active site just like the Asn side-chain amide (Fig. 4). We

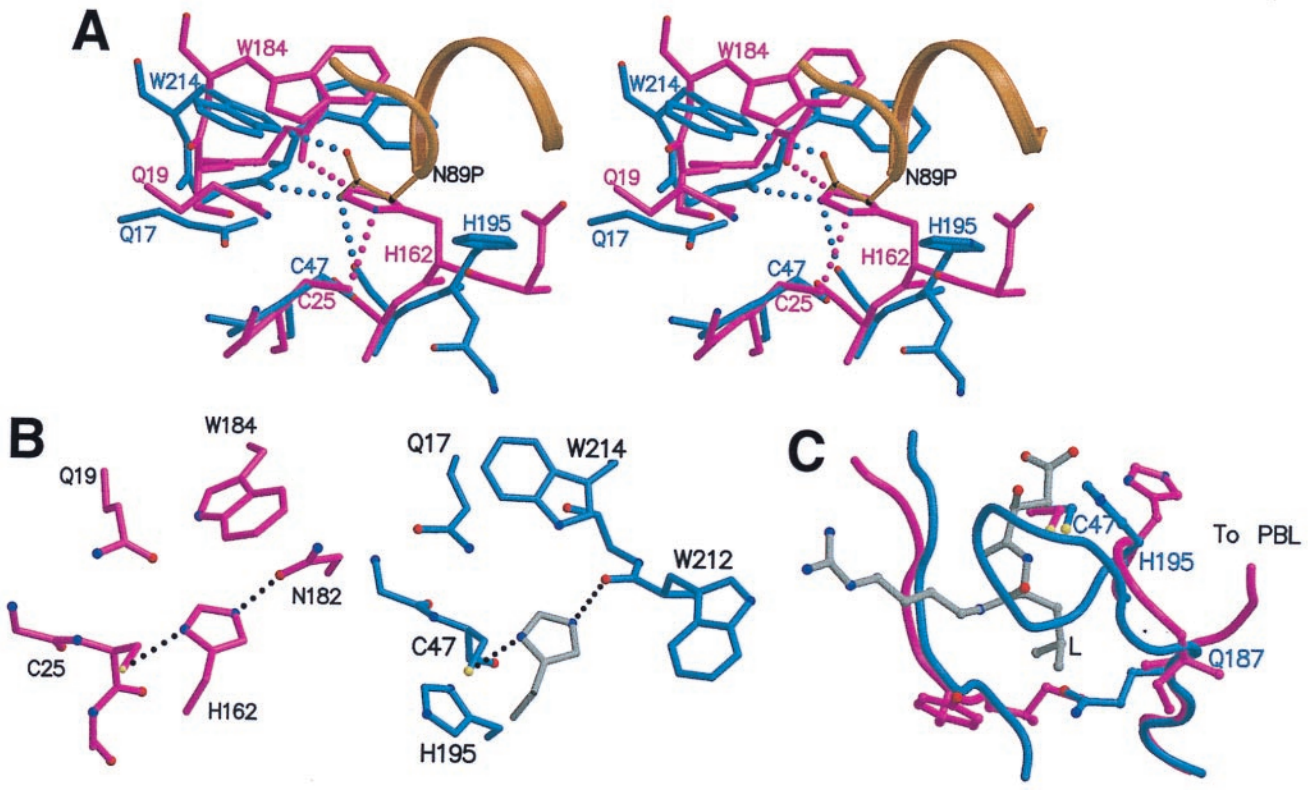

Fig. 4. Active site region of SpeB, compared with actinidin. (*A*) Stereo view of zSpeB (blue and gold) superimposed on actinidin (magenta). In zSpeB, the insertion of Asn-89p from the prosegment (gold) displaces His-195 from the catalytically competent position of His-162 in actinidin. Hydrogen bonds made by Asn-89p in SpeB and His-162 in actinidin are shown by dotted lines. (*B*) The Cys-His-Asn triad in actinidin (*Left*) compared with the putative Cys-His-(O=C) triad in mSpeB (*Right*). For mSpeB we assume that His-195 rotates from its position in the zymogen to a putative catalytic position shown in gray. (*C*) The S2 binding pocket. The leucyl side chain (L) of an inhibitor E64 (gray) is shown as it binds in actinidin (43). In zSpeB (blue) the loop 187–195 blocks the pocket; in actinidin (magenta) the chain goes in a different direction to form the PBL and the pocket is open.

speculate that in mSpeB, when His-195 flips back into the active site it may be oriented by a hydrogen bond to the main-chain oxygen of Trp-212, in a manner similar to that observed in ICE (42).

A third difference, which may have profound implications for the mode of substrate binding, concerns the S2 specificity pocket that is a conserved feature of all papain-family enzymes. Insofar as can be inferred from the zymogen structure, this pocket is not present in SpeB, as it is closed off by the loop 187–195 from the C-terminal domain (Fig. 4*C*). This short loop joins strands β1 and β2 in SpeB, substituting for the large PBL of the other papain family enzymes. It packs against residues 137–138 on the N-terminal domain, held in place by a hydrogen bond between the peptide nitrogen of Gly-194 and the peptide oxygen of Ser-137, thereby closing the pocket. This may be an additional mechanism by which the presence of the prosegment may inhibit enzyme activity.

Definitive information on substrate binding to mSpeB must await determination of the structure of the mature enzyme, given the possibility of loop movements and the presence of one undefined loop on the protease portion of zSpeB. Nevertheless, two alternative scenarios for substrate binding to SpeB are suggested by the zSpeB structure. In the first, removal of the prosegment would result in the movement of His-195 into a position equivalent to that of His-162 in actinidin, coupled with movement of the 187–195 loop so as to open up the S2 binding pocket. The loop movement could be facilitated by Gly residues at positions 188, 190, 193, and 194, and substrates then would bind as in other papain-family enzymes.

Alternatively, despite its similar specificity profile to other papain family enzymes (37), we speculate that mSpeB could have a quite different mode of substrate binding. In the zSpeB structure, not only is the S2 pocket blocked, but the truncation of the PBL removes one wall of the active site cleft. A substrate modeled into mSpeB can be reoriented such that its interactions with Cys-47 and the oxyanion hole are preserved, but it is swung around so that its N-terminal residues follow the direction taken by the prosegment helix 87p-107p, and its C-terminal residues extend toward the inserted "finger" loop 19–42. The substrate P2 side chain then could fit into a hydrophobic pocket formed by Phe-197, Trp-212, and Phe-222, exposed by the removal of the proregion. This possibility would mirror the relationship seen in the other cysteine proteases, whereby the prosegment mimics substrate binding (albeit in an opposite direction). It also would give a potential functional role to the finger loop, which has a rather similar spatial position to the occluding loop that modifies the activity of cathepsin B (18).

Structural Variation and Integrin Binding. Sequence analysis of the SpeB gene from more than 200 streptococcal isolates collected worldwide has revealed limited structural variation in SpeB (13). Most variations in the prosegment and protease domains involve fully exposed amino acid side chains on the protein surface although there are some conservative substitutions in the hydrophobic cores of the prosegment (V53pI, A84pV, and A91pV) or protease domains (I160V and V249L). None of these replacements should disrupt the three-dimensional structure. Notably, several regions completely lack amino acid variation. These areas encompass the entire active site, the putative substrate binding region, and the surfaces that mediate association of the proseg-

ment and protease. The prominent finger loop (residues 19–42) that extends from the N-terminal domain, near the active site, is also invariant, suggesting that antibodies directed against this loop could be effective therapeutic agents.

The most notable site of natural amino acid variation is at residues 162–164 (307–309, numbering from the signal peptide start). All serotype M1 organisms, which are the most common cause of invasive disease, and some 20% of other isolates, have an RGD sequence motif at this site, compared with RSD for other organisms (13). RGD motifs mediate ligand binding to many integrins (44) and the RGD-containing variant of SpeB (but not those with RSD) binds to the human integrins $\alpha_v\beta_3$ and $\alpha_{IIb}\beta_3$ through this motif (13). In the SpeB structure, Gly-163 and Asp-164 are fully solvent-exposed (Fig. 2*E*). The exposure of the Asp-164 side chain is particularly significant because integrin binding is believed to occur through coordination of the Asp residue of the RGD motif to a divalent metal on the integrin receptor (45).

The RGD motif in SpeB is located on a surface loop inserted between a short β-strand and the α-helix that leads into the C-terminal protease domain (Fig. 2*C*). Remarkably, this β-loop-α structure is similar to that of the RGD motif of the capsid protein VP1 of foot-and-mouth disease virus (FMDV) (46). This region of the VP1 protein is involved in recognition of cellular receptors for FMDV, such as $\alpha_v\beta_3$.

The RGD-containing loop of SpeB represents a unique insertion relative to other papain family members, in which the β-strand instead leads directly into the α-helix. The loop is remote from both the active site of mSpeB and the site where the prosegment associates with the protease, implying that integrin binding through this RGD motif can occur for both the zymogen and mature protease forms of SpeB, in accord with functional studies (13). Hence, when the protease is bound through its RGD motif, its active site should be fully exposed. Integrin binding thereby could concentrate mSpeB at the host cell surface, enhancing the efficiency of local degradation of host molecules such as fibronectin and vitronectin, as proposed (13).

We thank Lars Bjorck for encouragement and help, Michael Cooke and Christina Baker for technical assistance, W. A. Hendrickson for the gift of *E. coli* DL41, C. D. Lima for technical advice, and Vivian Ward for help with the cover illustration. The work has been supported by the Health Research Council of New Zealand, the New Zealand Lottery Grants Board (Grants MR17841 to J.C.C. and E.N.B. and MR36247 to J.C.C. and Paul O'Toole), the United States Public Health Service (Grant AI-33119 to J.M.M.), and the Texas Technology Development and Transfer Program (Grant 004949–036 to J.M.M.). E.N.B. also acknowledges research support as an International Research Scholar of the Howard Hughes Medical Institute.

1. Quinn, R. W. (1989) *Rev. Infect. Dis.* **11,** 928–953.
2. Proft, T., Moffatt, S. L., Berkhan, C. J. & Fraser, J. D. (1999) *J. Exp. Med.* **189,** 89–101.
3. Musser, J. M. (1997) in *Superantigens: Molecular Biology, Immunology, and Relevance to Human Disease*, eds. Leung, D. Y. M., Huber, B. T. & Schlievert, P. M. (Dekker, New York), pp. 281–310.
4. Musser, J. M., Hauser, A. R., Kim, M., Schlievert, P. M., Nelson, K. & Selander, R. K (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2668–2672.
5. Kapur, V., Topouzis, S., Majesky, M. W., Li, L.-L., Hamrick, M. R., Hamill, R. J., Patti, J. M. & Musser, J. M. (1993) *Microb. Pathog.* **15,** 327–346.
6. Lukomski, S., Sreevatsan, S., Amberg, C., Reichardt, W., Woischnik, M., Podbielski, A. & Musser, J. M. (1997) *J. Clin. Invest.* **99,** 2574–2580.
7. Kapur, V., Majesky, M. W., Li, L.-L., Black, R. A. & Musser, J. M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 7676–7680.
8. Wolf, B. B., Gibson, C. A., Kapur, V., Hussaini, I. M., Musser, J. M. & Gonias, S. L. (1994) *J. Biol. Chem.* **269,** 30682–30687.
9. Herwald, H., Collin, M., Muller-Esterl, W. & Bjorck, L. (1996) *J. Exp. Med.* **184,** 1–9.
10. Burns, E. H., Jr., Marciel, A. M. & Musser, J. M. (1996) *Infect. Immun.* **64,** 4744–4750.
11. Kellner, A. & Robertson, T. (1954) *J. Exp. Med.* **99,** 495–504.
12. Kuo, C.-F., Wu, J.-J., Tsai, P.-J., Kao, F.-J., Lei, H.-Y., Lin, M. T. & Lin, Y.-S. (1999) *Infect. Immun.* **67,** 126–130.
13. Stockbauer, K. E., Magoun, L., Liu, M., Burns, E. H., Jr., Gubba, S., Renish, S., Pan, X., Bodary, S. C., Baker, E., Coburn, J., *et al*. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 242–247.
14. Elliott, S. D. (1945) *J. Exp. Med.* **81,** 573–592.
15. Rawlings, N. D. & Barrett, A. J. (1994) *Methods Enzymol.* **244,** 461–486.
16. Tai, J. Y., Kortt, A. A., Liu, T.-Y. & Elliott, S. D. (1976) *J. Biol. Chem.* **251,** 1955–1959.
17. Berti, P. J. & Storer, A. C. (1995) *J. Mol. Biol.* **246,** 273–283.
18. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C. & Mort, J. S. (1996) *Structure (London)* **4,** 405–416.
19. Sivaraman, J., Lalumiere, M., Menard, R. & Cygler, M. (1999) *Protein Sci.* **8,** 283–290.
20. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S. & Cygler, M. (1996) *EMBO J.* **15,** 5492–5503.
21. Musser, J. M., Stockbauer, K., Kapur, V. & Rudgers, G. W. (1996) *Infect. Immun.* **64,** 1913–1917.
22. Gubba, S., Low, D. E. & Musser, J. M. (1998) *Infect. Immun.* **66,** 765–770.
23. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990) *EMBO J.* **9,** 1665–1672.
24. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
25. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D* **50,** 760–763.
26. Terwilliger, T. C. & Berendzen, J. (1999) *Acta Crystallogr. D* **55,** 849–861.
27. De La Fortelle, E. & Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494.
28. Petrakis, A., Morris, R. & Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463.
29. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
30. 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., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
31. Brunger, A. T. (1992) *Nature (London)* **355,** 472–475.
32. Engh, R. A. & Huber, R. (1991) *Acta Crystallogr. A* **47,** 392–400.
33. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
34. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.
35. Baker, E. N. (1980) *J. Mol. Biol.* **141,** 441–484.
36. Tao, K., Stearns, N. A., Dong, J. & Sahagian, G. G. (1994) *Arch. Biochem. Biophys.* **311,** 19–27.
37. Doran, J. D., Nomizu, M., Takebe, S., Menard, R., Griffith, D. & Ziomek, E. (1999) *Eur. J. Biochem.* **263,** 145–151.
38. Gubba, S., Cipriano, V. & Musser, J. M. (2000) *Infect. Immun.*, in press.
39. Vernet, T., Tessier, D. C., Chatellier, J., Plouffe, C., Lee, T. S., Thomas, D. Y., Storer, A. C. & Menard, R. (1995) *J. Biol. Chem.* **270,** 16645–16652.
40. Guarné, A., Tormo, J., Kirchweger, R., Pfistermueller, D., Fita, I. & Skern, T. (1998) *EMBO J.* **17,** 7469–7479.
41. Odagaki, Y., Hayashi, A., Okada, K., Hirotsu, K., Kabashima, T., Ito, K., Yoshimoto, T., Tsuru, D., Sato, M. & Clardy, J. (1999) *Structure (London)* **7,** 399–411.
42. Wilson, K. P., Black, J.-A. F., Thomson, J. A., Kim, E. E., Griffith, J. P., Navia, M. A., Murcko, M. A., Chambers, S. P., Aldape, R. A., Raybuck, S. A. & Livingston, D. J. (1994) *Nature (London)* **370,** 270–275.
43. Varughese, K. I., Su, Y., Cromwell, D., Hasnain, S. & Xuong, N-H. (1992) *Biochemistry* **31,** 5172–5176.
44. Wickham, T. J., Mathias, P., Cheresh, D. A. & Nemerow, G. R. (1993) *Cell* **73,** 309–319.
45. Bergelson, J. M. & Hemler, M. E. (1995) *Curr. Biol.* **5,** 615–617.
46. Logan, D., Abu-Ghazaleh, R., Blakemore, W., Curry, S., Jackson, T., King, A., Lea, S., Lewis, R., Newman, J., Parry, N., *et al*. (1993) *Nature (London)* **362,** 566–568.