A Proline-Rich Region with a Highly Periodic Sequence in Streptococcal beta Protein Adopts the Polyproline II Structure and Is Exposed on the Bacterial Surface.

Areschoug, Thomas; Linse, Sara; Stålhammar-Carlemalm, Margaretha; Hedén, Lars-Olof; Lindahl, Gunnar

Published in:
Journal of Bacteriology

DOI:

2002

Citation for published version (APA):

Total number of authors:
5

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 25. Jun. 2024
A Proline-Rich Region with a Highly Periodic Sequence in Streptococcal β Protein Adopts the Polyproline II Structure and Is Exposed on the Bacterial Surface

Thomas Areschoug,1 Sara Linse,2 Margaretha Stålhammar-Carlemalm,1 Lars-Olof Hedén,3 and Gunnar Lindahl1*

Department of Medical Microbiology, Dermatology and Infection,1 Department of Biophysical Chemistry,2 and Department of Cell and Organism Biology,3 Lund University, Lund, Sweden

Received 22 April 2002/Accepted 23 June 2002

Proline-rich regions have been identified in many surface proteins of pathogenic streptococci and staphylococci. These regions have been suggested to be located in cell wall-spanning domains and/or to be required for surface expression of the protein. Because little is known about these regions, which are found in extensively studied and biologically important surface proteins, we characterized the proline-rich region in one such protein, the β protein of group B streptococci. The proline-rich region in β, designated the XPZ region, has a proline at every third position, and the sequence is highly periodic in other respects. Immunochemical analysis showed that the XPZ region was not associated with the cell wall but was exposed on the bacterial surface. Moreover, characterization of a β mutant lacking the XPZ region demonstrated that this region was not required for surface expression of the β protein. Comparison of the XPZ region in different β proteins showed that it varied in size but always retained the typical sequence periodicity. Circular dichroism spectroscopy indicated that the XPZ region had the structure of a polyproline II helix, an extended and solvent-exposed structure with exactly three residues per turn. Because of the three-residue sequence periodicity in the XPZ region, it is expected to be amphipathic and to have distinct nonpolar and polar surfaces. This study identified a proline-rich structure with unique properties that is exposed on the surface of an important human pathogen.

Proline-rich regions play important roles in many protein-protein interactions, such as signaling events involving SH3 domains in eukaryotic cells (27, 59, 61). In other cases, proline-rich regions are important structural elements, e.g., in the hinge region of immunoglobulin A1 (IgA1) (11). However, for many proline-rich regions, the cellular localization and function are unclear.

Many surface proteins in pathogenic streptococci and staphylococci have been shown to include a proline-rich region. For example, such regions have been identified in the M6, ScIA, and ScIB proteins of Streptococcus pyogenes (20, 37, 38, 49, 50, 58), in protein A of Staphylococcus aureus (18), in protein G of group G streptococci (15), in PspA and PspC of Streptococcus pneumoniae (6, 13), in the P1 adhesin of Streptococcus mutans (12), in protein L of Peptostreptococcus magnus (26), and in FnBA of Streptococcus dysgalactiae (36). In all of these cases, the proline-rich region is located in the C-terminal, wall-proximal half of the protein. It has therefore been proposed that such proline-rich regions are associated with the bacterial cell wall (6, 16, 18, 45) or are required for cell surface expression (12), but their role remains unclear. Indeed, it is not known whether the proline-rich regions referred to above have similar or different functions. Because of the prevalence of proline-rich regions with unknown structure and function in extensively studied and biologically important bacterial surface proteins, we have characterized the proline-rich region in one such protein, the streptococcal β protein.

The ~125-kDa β protein (also known as β C and Bac) is expressed by many strains of group B streptococci (GBS), human pathogens that are the most common cause of life-threatening bacterial infections in the neonatal period (5, 52). The β protein elicits protective immunity (9) and has therefore been evaluated as a possible component in a vaccine against GBS disease (40). Immunochemical analysis has demonstrated that β has separate binding sites for human IgA-Fc (19, 24, 25, 47, 51) and the complement regulator factor H (3) (Fig. 1), properties that may allow β to interfere with IgA- and complement-mediated opsonization (3, 47).

Unlike most surface proteins of GBS and other gram-positive bacteria (28, 44, 57), the β protein does not contain any long repeats, but its C-terminal part includes a proline-rich sequence, the XPZ region, which exhibits a unique sequence periodicity (19, 24). This region is composed of tandemly arranged three-residue XPZ motifs, in which the first residue (X) is invariably uncharged, the second residue (P) is proline, and the third residue (Z) is almost invariably charged. Moreover, the third residue alternates between a positive and a negative charge (Fig. 1). Thus, the distribution of amino acid residues in the XPZ region is highly periodic. The function and cellular location of this region have remained unknown, but several investigators have proposed that it is located within the cell wall-spanning part of the β protein (8, 24, 32).

The gene encoding β, the bac gene, has been sequenced in...
two unrelated GBS strains (19, 24). Comparison of the two deduced sequences shows that they have XPZ regions containing 30 and 40 XPZ motifs, respectively, but apart from that difference the amino acid sequences are identical except for a single residue. These limited data suggest that the protein varies in size due to size variability in the XPZ region.

We show here that the XPZ region is exposed on the bacterial surface and is not required for cell wall anchoring of the protein. Moreover, we present spectroscopic evidence that the XPZ region adopts the conformation of a left-handed polyproline II (PPII) helix, an extended and solvent-exposed structure (55, 59). The length and sequence of the XPZ region varied in different β proteins, but the sequence periodicity was strictly conserved. Together, these data indicate that the proline-rich XPZ region of the protein represents a novel type of surface structure in bacteria.

MATERIALS AND METHODS

Bacterial strains and media. The GBS strains used were of capsular serotypes 1a, 1b, and II. Strain SB35 was used to sequence the gene, which encodes β (19). Strains A909 and H36B (51) were obtained from C. Schalen (Lund University, Lund, Sweden), and strain 70339 (9) was from L. Bevanger (University of...
The peptides were determined by amino acid analysis after acid hydrolysis (Biomedical Center, Uppsala, Sweden).

**Other methods.** Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting were performed as described previously (54). Analysis of surface expression of β and of the XPZ region was performed by a two-step binding assay, with rabbit antiserum and radiolabeled protein G, as described previously (54); controls with preimmune rabbit serum were included in all tests and were negative. Analysis of the ability of bacteria to bind IgA-Fc or factor H was performed as described previously (3, 34). Reactivity of antibodies with β protein immobilized in microtiter wells was performed essentially as described above, with wells coated with pure β protein (1 μg/ml; 50 μl/well). Proteins were labeled with carrier-free 125I Na (Amersham International, Amersham, United Kingdom), with the chloramine-T method.

### Nucleotide sequence accession numbers

The nucleotide sequences corresponding to the XPZ regions reported here are as follows: BS22, AF522266; 70339, AF522264; BS26, AF522267; SB20, AF522272; H36B, AF522269; A909, AY126439; SB1, AF522270; BS4, AF522268; SB10, AF522271; and BS1, AF522265. For strain SB35, the sequence of the entire bac gene, including the part encoding the XPZ region, has been reported previously (19).

### RESULTS

#### Size variation and sequence periodicity in the XPZ region

The two published sequences of the β protein include 30 and 40 XPZ motifs (19, 24). However, the β protein encoded by one of these strains (24) is expressed at a very low level (31), which made it unclear whether size variation in the XPZ region is compatible with normal surface expression of the β protein. The first part of this study was therefore aimed at analyzing whether the XPZ region is present in all surface-expressed β proteins, whether it varies in length in these β proteins, and whether the characteristic sequence periodicity is present in all XPZ regions. For this purpose, we studied 11 GBS strains that expressed β on the surface, as judged by their ability to bind IgA-Fc (data not shown). These strains included SB35, which was used for sequencing of the structural gene for β, the bac gene (19).

The β proteins expressed by the 11 GBS strains were isolated in almost pure form by incubating bacterial suspensions at elevated pH, a procedure that causes selective release of β protein (35). Analysis of these β proteins by SDS-PAGE demonstrated a limited size variability (Fig. 1A). Western blot analysis showed that all proteins bound IgA-Fc and were recognized by anti-β antibodies, confirming that they were β proteins. Moreover, all of these β proteins were recognized by an antiserum specific for the XPZ region, implying that they all contained an XPZ region (data not shown). To analyze whether the size variation between different β proteins could be explained by a change in the length of the XPZ region, the corresponding region of the gene was amplified by PCR. The size of the PCR product correlated with the size of the purified β protein, suggesting that the limited size variability was indeed due to variation in the length of the XPZ region. In contrast, the PCR product corresponding to a region located upstream of the XPZ region did not show any variation in size (data not shown).

Sequence analysis of the XPZ region in the 11 β-expressing strains (Fig. 1B) showed that the size of this region varied from 30 to 40 residues (Fig. 1B, Table 1) with a peak at 35 residues (17 in the bacterial strain BS1). Amino acid sequence alignment of concentrated peptides stock at pH 7 were diluted in 5 mM potassium phosphate and/or 5 mM potassium phosphate with 0.7 M NaF. The concentration of peptides was determined by amino acid analysis after acid hydrolysis (Biomedical Center, Uppsala, Sweden).
in all XPZ regions. As indicated in Fig. 1B, the different protein sequences may all be derived from the shortest sequence by the addition of blocks containing multiples of six amino acid residues, i.e., by the addition of multiples of six codons at the DNA level.

To analyze whether the XPZ region also varies in size within a bacterial strain, strains A909 and H36B were grown in 10 ml of broth and diluted 100-fold for 10 consecutive days, corresponding to ~66 generations of growth. For each of the final cultures, 15 colonies were picked, and the length of the XPZ region was analyzed by PCR. No size variation was seen in this analysis, which would have detected a change in length corresponding to six codons. Thus, the XPZ region appears to be stable within a strain under laboratory conditions.

The XPZ region is not required for surface expression of β protein. For some surface proteins of streptococci and staphylococci, it has been suggested that the proline-rich region threads through the peptidoglycan layer (16, 18, 45), suggesting that this region is important for surface localization of the protein. Moreover, it has been reported that a proline-rich region in the P1 adhesin of Streptococcus mutans is required for surface expression (12). This situation made it of interest to analyze whether the XPZ region was needed for surface expression of β. For this purpose, we constructed a bac gene with an in-frame deletion covering the entire XPZ region, and a plasmid carrying this construct (plasmid pLZΔxpz) was used to transform a GBS strain (the Δbac strain) in which the bac gene had been deleted, generating the Δbac/pLZΔxpz strain. Surface expression of β was analyzed in this strain and in a strain expressing the wild-type β protein (the Δbac/pLZbac strain), with the Δbac strain as the negative control (Fig. 2A).

For this analysis, we used an antiserum directed against the N-terminal B6 fragment of β, which does not include the XPZ region (Fig. 1B). This antiserum reacted equally well with the strain expressing the wild-type β protein and that expressing the XPZ deletion variant of β (designated βΔXPZ) but did not react with the β-negative Δbac strain (Fig. 2A). Moreover, the binding of IgA-Fc was only slightly reduced for the strain expressing βΔXPZ compared to the strain expressing the wild-type protein (Fig. 2B). These data indicate that the XPZ region is not required for surface expression of the β protein and that it is not required for IgA-Fc binding to surface-expressed β.

The XPZ region is exposed on the bacterial cell surface. An antiserum directed against a synthetic peptide derived from the XPZ region (Fig. 1B) was used to analyze whether this region is exposed on the bacterial surface. The specificity of this antiserum was first studied by Western blot analysis of purified preparations of the β protein and the β variant lacking the XPZ region, the βΔXPZ protein (Fig. 3A). Both proteins reacted with antibodies to the N-terminal B6 fragment of β, but the anti-XPZ serum reacted only with the wild-type β protein. Thus, the anti-XPZ serum could be used to specifically detect the XPZ region.

The anti-XPZ serum reacted with bacteria expressing the β protein but not with the isogenic β-negative Δbac mutant (Fig. 3B), indicating that the XPZ region is at least partially exposed on the bacterial surface and is not buried in the cell wall. As expected, the anti-XPZ serum did not react with a GBS strain expressing the βΔXPZ protein. Although these data indicated that the XPZ region is expressed on the surface of GBS, it seemed possible that the result was due to β protein that had been released from the bacteria, followed by (nonspecific) readsorption to the surface. A control experiment was therefore performed in which a suspension of the β-negative Δbac mutant (10^9 CFU/ml) was incubated for 30 min at 23°C in the

![Fig. 2](image-url)
with that of antiserum against the XPZ region. These two antisera were first analyzed for reactivity with pure β immobilized in microtiter wells, i.e., under conditions in which all parts of the β protein should be equally accessible to antibodies. Under these conditions, the titer of the anti-XPZ serum was ~2-fold lower than the titer of the anti-B6 serum (Fig. 4A). When the same two antisera were analyzed for reactivity with β present on the bacterial surface, the titer of the anti-XPZ serum was also ~2-fold lower (Fig. 4B). This result shows that the N-terminal region and the XPZ region are equally accessible to antibodies whether β is present in pure form or is exposed on the surface of GBS. Finally, electron microscopic analysis with immunogold techniques showed that β is surface exposed on virtually all bacteria (data not shown). Together, these data indicate that the XPZ region is fully surface exposed on a GBS strain expressing the β protein.

**CD spectroscopy indicates that the XPZ region has PPII helix structure.** A sequence of four or more prolines in a row adopts the conformation of a left-handed PPII helix, an extended structure with exactly three residues per turn (29, 55, 59). This conformation is also known to be adopted by many regions that are rich in proline but contain other amino acids, including some regions in which every third residue is proline (59).

To analyze whether the XPZ region adopts a PPII structure, we characterized a 24-residue synthetic peptide derived from the XPZ region. This peptide, designated (XPZ)6, was studied by circular dichroism (CD) spectroscopy (Fig. 5). At 5°C, the CD spectrum of (XPZ)6 showed a deep negative peak at 200 nm (~33 deg mol⁻¹ cm⁻²). The CD spectrum of polyproline was characterized by a large negative band (~59 deg mol⁻¹ cm⁻²) at 205 nm and a small positive band (~4.6 deg mol⁻¹ cm⁻²) at 228 nm (23). However, many other peptides with the PPII helical structure display spectra very similar to the one we observed here for (XPZ)6, with a minimum at around 200 and the positive peak at 228 nm missing (2, 39, 48).

The spectral shifts relative to polyproline are often attributed to slight deviation from the ideal PPII helical structure, the presence of cis-proline residues, or formation of higher-order helical forms (4, 43). Raising the temperature to 25, 45, 65, and 85°C causes a gradual decrease in intensity of the negative band, a shift towards a higher wavelength, and the appearance of a negative shoulder at 220 to 230 nm. Similar temperature effects are seen for many other PPII structures and are due to unfolding to an unordered state. In summary, we may conclude that (XPZ)6 adopts a left-handed PPII helix that is disrupted at elevated temperature.

Because of the intriguing sequence of alternating positive and negative charges in the Z position, CD spectra were recorded as a function of salt and pH to evaluate the electrostatic contribution to the structure. Spectra recorded at 0.005, 0.15, and 0.7 M ionic strength superimposed completely (data not shown), indicating that screening of the electrostatic interactions does not affect the structure of (XPZ)6. Another way to perturb potentially stabilizing contributions from attractive electrostatic interactions between opposite charges is to go to extreme pHs. One may expect that the Z position carries alternating positive and uncharged residues at pH 1.5 and alternating negative and uncharged side chains at pH 12. The CD spectra obtained at pH 1.5 and pH 12 were, however, very
similar to that at pH 7, but with a small decrease in the intensity at 200 nm (by 8% at pH 1.5 and by 12% at pH 12) (data not shown). The lack of substantial salt and pH effects shows that ionic interactions are only minor determinants of the PPII structure of \((XPZ)_8\), which is hence governed by other forces.

**DISCUSSION**

The data reported here indicate that the proline-rich XPZ region of the streptococcal \(\beta\) protein is exposed on the bacterial surface and adopts the conformation of a left-handed PPII helix. Although the exact function of the XPZ region is not yet known, these findings shed new light on the properties of proline-rich regions in surface proteins of streptococci and staphylococci. In particular, our data indicate that the XPZ region is not required for cell wall anchoring and/or surface expression of the \(\beta\) protein. Rather, the surface exposure of this region and the knowledge that proline-rich regions commonly participate in protein-protein interactions (59) suggest that the XPZ region could promote interactions with host proteins or with other bacterial surface proteins. The XPZ region could also have a structural role, e.g., it might act as a scaffold that allows surface exposure of other regions of the \(\beta\) protein. However, our data do not support this hypothesis, because deletion of the XPZ region had little or no effect on surface exposure of the N-terminal IgA-binding part of \(\beta\).

The conclusion that the XPZ region adopts the conformation of a PPII helix was based on CD analysis of the \((XPZ)_8\) peptide, which exhibits the characteristic sequence periodicity. Because this periodicity is conserved throughout the XPZ region, it seems likely that the intact XPZ region is a PPII helix. This extended and solvent-exposed conformation, which has been studied extensively due to its importance for protein-protein interactions (1, 29), may even occur in regions lacking proline (1). The importance of this conformation is underlined by its presence in each of the three strands of collagen (7, 59) and by the finding that class II major histocompatibility complex molecules bind peptides in a PPII-like conformation (22, 42).

Because there are exactly three residues per turn in a PPII helix (29, 59) and because the XPZ region exhibits three-
residue periodicity, the helix corresponding to the XPZ region is expected to be amphipathic and to have three surfaces with very different properties, corresponding to the X, P, and Z residues. The X surface is mainly hydrophobic, although with a fair number of hydroxyl groups. The proline rings will line up to form an extended hydrophobic surface, while the Z residues will form a hydrophilic stretch of alternating charges. This separation of surface properties may allow the XPZ region to interact simultaneously with molecules having binding sites of very different characters. Possibly, the alternating charges of the third residue could also contribute to dimerization of two staggered XPZ regions by a zipper-like mechanism (46).

The XPZ region varies in size between different β proteins and is present in all β proteins analyzed, and the sequence periodicity is strictly maintained in XPZ regions of different lengths. These data suggest that the XPZ region is important for the function of the β protein, and they are in good agreement with recent studies that employed PCR to demonstrate size variation in the XPZ region (8, 30). The mechanisms that cause size variation in the XPZ region are not known, and it is not clear whether the mechanisms commonly used to explain size variability in repetitive sequences (10, 14) can also explain size variation in the XPZ region, because this region has a periodic rather than a repetitive sequence. Possibly, the size variability arises through an error-prone process (41), causing duplications and deletions of multiples of six codons, followed by a selection that eliminates mutants that do not retain the typical periodicity.

In this context, it is of interest that three of the four proline codons are used for the central residue in the XPZ motif, suggesting that the XPZ region is indeed subject to a high mutation rate and to a selection process responsible for conservation of the periodicity (33). The selection of XPZ variants that retain the sequence periodicity may reflect a structural requirement, while the sequence variability could reflect immunological selection of antigenically different variants, a hypothesis implying that sequence variability in the XPZ region allows GBS to escape host immunity.

In summary, we have identified a PPII helix with a highly periodic sequence that is exposed on the surface of GBS. To our knowledge, a surface structure of this type has not previously been identified in bacteria. Finally, we note that “natural” antibodies directed against self-antigens preferentially recognize epitopes rich in proline and were suggested to provide a first line of defense against pathogens expressing proline-rich surface proteins (56). Conversely, it seems possible that surface-exposed proline-rich regions in pathogens may contribute to the emergence of natural antibodies or even to the development of autoimmunity. This argument is particularly relevant with regard to the XPZ region studied here, because the PPII helix structure of this region may favor presentation by major histocompatibility complex class II molecules (22, 42).

ACKNOWLEDGMENTS

We are grateful to Lars Stenberg and Charlotte Bergquist for contributions in the early stages of this study and to Erik Lindahl for advice. We also thank the students in the Prokaryotic Molecular Genetics course at Lund University for help with sequence analysis.

This work was supported by grants from the Swedish Research Council (grant 09490 to G.L. and grant 11552 to S.L.), the Medical Faculty of Lund University, the Royal Physiographic Society in Lund, the Swedish Society for Medical Research, and the Golje Groschinsky, Hedberg, Hierta, Jerring, Kock, Lundström, and Österlund foundations.

REFERENCES


