Characterization of the Bacillus anthracis S-Layer: Cloning and Sequencing of the Structural Gene

ISABELLE ETIENNE-TOUMELIN, JEAN-CLAUDE SIRARD, EDITH DUFLOT, MICHELE MOCK, AND AGNES FOUET*

Laboratoire de Genetique Moléculaire des Toxines, Unité de Recherche Associée 1858, Centre National de la Recherche Scientifique, Institut Pasteur, 75724 Paris Cedex 15, France

Received 3 August 1994/Accepted 11 November 1994

Bacillus anthracis, a gram-positive, spore-forming bacterium, is the etiological agent of anthrax. The gene coding for the S-layer protein (sap) was cloned on two contiguous fragments in Escherichia coli, and the complete sequence of the structural gene was determined. The protein, Sap, is composed of 814 residues, including a classical prokaryotic 29-amino-acid signal peptide. The mature form has a calculated molecular mass of 83.7 kDa and a molecular mass of 94 kDa on a sodium dodecyl sulfate-polyacrylamide gel. Sap possesses many charged residues, is weakly acidic, and contains only 0.9% methionine and no cysteine residues. The N-terminal region of Sap shares sequence similarities with the Acetogenium kivui S-layer protein, the Bacillus brevis middle wall protein, the Thermotoga maritima OmpC protein, and the Bacillus thuringiensis S-layer protein. Electron microscopy observations showed that this S-layer is not observed on B. anthracis cells in which sap has been deleted.

Bacillus anthracis, the causative agent of anthrax, is a gram-positive, spore-forming bacterium. Fully virulent bacilli are both capsule and sporeogenic. The two toxins (lethal and edema toxins) are encoded by the virulence plasmid pXO1, and capsule synthesis is dependent on the presence of a second virulence plasmid, pXO2. When the capsule is absent, the cell wall of B. anthracis appears layered and is composed of small fragments displaying a highly patterned ultrastructure (8). Holt and Leadbetter (9) previously described a hexagonal lattice on the surface of vegetative cells of B. anthracis. This cell surface structure most likely represents what is called an S-layer, with p6 symmetry and a center-to-center spacing of the particles of 7 to 10 nm.

S-layers, or surface arrays, have been found to be the outermost component of many archaea and eubacteria. In most cases, single proteins compose these structures. When present, they represent 5 to 10% of the total cell protein, implying that their synthesis is energy-consuming for the bacterium. The fact that S-layers are found ubiquitously suggests that they play vital roles in the interaction between the cell and its environment. It has been suggested that the S-layer is an important virulence factor for bacteria such as Aeromonas salmonicida, Campylobacter fetus, and Rickettsia spp., protecting against complement killing, facilitating binding of the bacterium to host molecules, or enhancing its ability to associate with macrophages (see reference 21).

Unlike most S-layers, that of B. anthracis is not the outermost component of the virulent bacilli since they are encapsulated. Azotobacter spp. are another of the rare examples of bacteria possessing both a capsule and an S-layer. The S-layer may have an important function in linking the capsule to the peptidoglycan wall or controlling the exchange of molecules with the environment. Bacillus thuringiensis, a closely related entomopathogenic bacillus, possesses an S-layer (12). The B. thuringiensis S-layer is composed of linear arrays of small par-ticles arranged with p2 symmetry, the constituent being a 91.4-kDa protein. Another bacillus, Bacillus brevis, has an unusual S-layer, since the bacterium is covered by two S-layer proteins, the outer wall protein (molecular mass, 103.7 kDa [25]) and the middle wall protein (MWP) (molecular mass, 114.8 kDa [24]). Transcription of the corresponding operon is initiated from multiple and tandemly arranged promoters (1).

In order to study the B. anthracis S-layer, we decided to analyze its composition. We hypothesized that a major cell protein with a molecular mass approximately equal to 94 kDa, which is often observed in high abundance in culture supernatant fluids of most strains of B. anthracis, was the component of the S-layer. Here, we describe the cloning and sequencing of the gene encoding the 94-kDa protein. The relationship between this protein and the surface array component described by Holt and Leadbetter (9) was determined by constructing a sap deletion mutant and analyzing it by electron microscopy.

MATERIALS AND METHODS

Bacterial strains, media, and growth conditions. Escherichia coli TGI (13) was used as a host for plasmid constructions. E. coli JM83(pRK24) (23) was used as a host for the mating experiments with B. anthracis RPI1 (18). The B. anthracis plasmidless strain, 9311, was obtained by curing RP51 of pXO1 (19). All these B. anthracis strains are derivatives of the Sterne strain, i.e., they are pXO1− pXO2−. Conjugal transfer from E. coli to B. anthracis was as described by Pezet et al. (18). Double crossover events were screened for as described by Cataldi et al. (3). E. coli was grown in L broth or on L-agar plates (14). B. anthracis was grown in brain heart infusion medium (Difco Laboratories) or in SPY medium [60 mM K2HPO4, 45 mM KH2PO4, 15 mM (NH4)2SO4, 10 mM MgSO4, 2.4 mM sodium citrate, 0.2% (wt/vol) glucose, 0.2% (wt/vol) yeast extract]. Spectinomycin was used at 60 μg/ml for B. anthracis and E. coli, and ampicillin was used at 100 μg/ml for E. coli.

DNA manipulations. Methods for the isolation and manipulation of recombinant DNA were as described by Maniatis et al. (13). Chromosomal DNA was prepared as described by Fouet and Sonenshein (6). DNA sequencing was carried out using Sequenase (U.S. Biochemicals) with either the primers pro-vvided or an oligomer hybridizing to a determined intragenic sequence.

Oligodeoxynucleotide probes, DNA libraries, and plasmids. The two oligodeoxynucleotide probes used were SEAG1 [5'-GGCTATGGAAACATTATTTCTCATAAGGTTCCTCAGGCT(G)ATG-3'] and SEAG2 [5'-TTCTATGTGTCTTATCCTCCGATCTACCATGACAT-3']. In contrast to SEAG1, the SEAG2 oligodeoxynucleotide probe was devised complementarily to that directly deduced from the amino acid sequence, the two probes being therefore convergent. DNA libraries were obtained by ligating the product of complete HindIII and EcoRV

* Corresponding author. Mailing address: Laboratoire de Ge-netique Moléculaire des Toxines, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France. Phone: (1) 45.68.86.54. Fax: (1) 45.68.89.54.
The N-terminal sequences of proteolytic fragments obtained from the 94-kDa protein.

<table>
<thead>
<tr>
<th>Peptide number(s)</th>
<th>Apparent molecular mass (kDa)</th>
<th>N-terminal sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94</td>
<td>Ala Gly Lys Thr Phe Pro Asp Val Pro Ala Asp</td>
</tr>
<tr>
<td>2, 4</td>
<td>80, 50</td>
<td>Thr Glu Ala Ala Lys Val Glu Ser Ala Lys</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>Thr Val Asp Val Asn Lys Val Gly Lys Thr Glu</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>Thr Val Val Leu Thr Ala Lys Ser Gly Glu</td>
</tr>
</tbody>
</table>

*The underlined residues are identical in the N-terminal sequence of the 94-kDa protein and that of the B. thuringiensis S-layer protein.

**The mature form of the protein.**
quence and the ATG initiation codon that is consistent with that observed in other gram-positive bacteria (26).

Two palindromic structures were observed downstream from the TAA stop codon, the first beginning just after the TAA (DG 52 11.6 kcal) and the second beginning 49 nucleotides farther downstream (DG 52 6 kcal) followed by a stretch of U’s. These two sites may correspond to a factor-independent transcription termination signal.

Translation of the ORF produced an 814-amino-acid protein with a calculated molecular mass of 86,567 Da. The sequences of the chymotryptic polypeptides were localized to various positions throughout the protein (Fig. 3). A comparison of the N terminus of the mature protein (Table 1) and the amino acid sequence deduced from the nucleotide sequence (Fig. 3) indicated that this protein is synthesized as a pre-polypeptide with a 29-amino-acid signal peptide. The signal

FIG. 3. DNA and deduced amino acid sequence of the region encompassing the sap gene. In the left margin, numbers in roman type indicate the first nucleotide of the line, and numbers in italic type refer to the amino acid residues, for which the one-letter code has been used (the asterisk represents the stop codon). The putative ribosome-binding site and promoter sequences as well as the four short internal repeats, have been underlined. The arrows indicate the palindromic structures. The HindIII and EcoRV restriction sites shown in Fig. 2 are indicated.

ETIENNE-TOUMELIN ET AL. J. BACTERIOL.
The greatest similarity was found between the N-terminal repeat (15) that exhibited 18.5% identity in a 346-amino-acid overlap. The amino acid composition of the Sap protein was compared with that of other S-layer proteins. It is noteworthy that all these proteins share a number of common features. For example, they are devoid of cysteine residues, contain very few methionines (0.9% in the mature form of the B. anthracis Sap protein), and have similar percentages of nonpolar, acidic, and basic residues [21]. Like the Acetogenium kivui S-layer protein [16], the B. anthracis Sap protein is not very acidic, with a calculated pI of 6.02. This is in contrast to other S-layer proteins that tend to have a more acidic pI.

S-layer proteins often harbor internal repeats, and two types were found in Sap. A short sequence was repeated four times throughout the second half of the protein (Fig. 3 and 4). It was lysine and glutamic acid rich and had a conserved motif (6 of 11 residues were identical or similar) (Fig. 4). A longer repeat was present twice and was localized in the N-terminal part of the protein, as has been observed for other S-layer proteins [21].

Several proteins sharing similarities with Sap were found by using the FASTA program of the Genetics Computer Group package to scan the SwissProt database. The proteins with the best initial scores were the S-layer protein of A. kivui [16] with 15.3% identity and the Omp protein of Thermotoga maritima [4] that exhibited 18.5% identity in a 346-amino-acid overlap. The greatest similarity was found between the N-terminal regions of these proteins (Fig. 5). The Pileup program was applied to Sap and five other sequences because of shared similarities, as suggested by the FASTA search, or as indicated by other authors. The proteins compared were Sap, the A. kivui S-layer protein [16], the MWP of B. brevis [24], the hexagonally packed intermediate layer protein of Deinococcus radiodurans [17], the Omp protein of T. maritima, and the outer wall protein of B. brevis [25]. Pileup indicated that the A. kivui S-layer protein and the B. brevis MWP are the most similar, immediately followed by the B. anthracis Sap and the T. maritima Ompa. Interestingly, while all similarities were not considered, the N-terminal regions aligned well. The greatest similarities were between regions which correspond to internal repeats in the S-layer proteins and overlap the longer Sap internal repeat (Fig. 5).

Characterization of the sap gene product. To study the role of the S-layer structure, a sap deletion mutant was constructed. The HindIII-EcoRV fragment between nucleotides 1005 and 1304 (pEAI 207 [Fig. 2]) was removed and replaced by a spectinomycin resistance cassette on a conjugative suicide vector. Subsequently, the wild-type sap gene was deleted by allelic exchange on the chromosome of RP10 which was subsequently cured of pXO1. The secreted proteins produced by the resulting sap-deleted strain (RBA2) were compared with those in the supernatant of strain 9131. On an SDS-polyacrylamide gel, there was an absence of the 94-kDa protein in RBA2 (Fig. 1).

We examined the correlation between the presence of Sap and an N-terminal region around B. anthracis. The electron micrographs of strains 9131 and RBA2 are shown in Fig. 6. The S-layer was visible around the 9131 bacterium, with a typical patterned ultrastructure, but was not observed around RBA2. This supports the notion that the sap gene encodes a protein associated with the B. anthracis S-layer. Colonies of RBA2 and 9131 were morphologically very different. The RBA2 colonies were much larger than the colonies of 9131, and they tended to connect when streaked. When the two strains were grown in SPY liquid medium, the asp strain flocculated during growth and sedimented as soon as shaking was arrested, which is different from the behavior of the wild type. Similar morphological changes were observed with other constructed mutants deficient in Sap production. Analysis with an optical microscope showed that cells from 9131 were classical rod-shaped bacilli (≈5 μm long), whereas RBA2 cells were long and filamentous (≈100 μm long). The presence of these long filaments could explain the flocculation observed in liquid culture and the formation of large colonies. It is tempting to correlate all these morphological differences with the variations in Sap protein quantity, i.e., with the presence or absence of an S-layer.

**DISCUSSION**

In a first step to study B. anthracis S-layer structure, the gene encoding the corresponding protein subunit was isolated on

![Fig. 5](http://jb.asm.org/) Comparison of the N-terminal region of the Sap protein with that of two other S-layer proteins and one related protein. The four sequences (B. brevis MWP, Bb; A. kivui S-layer protein, Ak; T. maritima Omp protein, Tm; and B. anthracis Sap protein, Ba) were aligned by introducing gaps (dots). The Sap protein internal repeat, starting at residue 170, was added. For the sake of clarity, the two B. anthracis sequences have been grouped at the bottom of the figure, whereas the Pileup program had inserted the N-terminal sequence of Sap between the A. kivui and the T. maritima sequences. The signal peptide cleavage sites (arrow) have been indicated. Identity and similarity (Fig. 4; also, A, G, D, E, N, F, W, Y; S, T, K, R) were established in comparison with the N-terminal portion of the Sap protein.
two contiguous DNA fragments. Oligonucleotide probes derived from established amino acid sequences of the 94-kDa major extracellular protein were successfully used to clone the corresponding gene (sap). Sequence data and deletion analysis support the notion that the S-layer comprises the 94-kDa protein.

Sequences resembling the recognition sequences of Bacillus subtilis $\sigma^A$ factor were looked for in the region upstream from this ORF, and none which were really convincing were found. This absence is not entirely surprising. Such sequences have not been found for B. anthracis promoters where transcription is initiated at a high level, for example, upstream from a toxin component gene (10). Moreover, the promoter, which is probably responsible for the constitutive high-level synthesis of the B. brevis cell wall proteins, lacks the $-35$ sequence (1). The absence of a $\sigma^A$ consensus sequence in the vicinity of the known B. anthracis genes could be due to altogether different $\sigma$ factor recognition sequences. This would be unusual, however, since phylogenetically more distant organisms, such as E. coli and B. subtilis, share the same consensus sequences. Another explanation is that these B. anthracis genes are not transcribed by a major $\sigma$ factor, although less utilized $\sigma$ factor recognition sequences were not found. The hypothesis that we favor is that these genes are positively regulated and that, as with such genes from B. subtilis, an activator compensates for the poor recognition sequence. The regulatory region of the B. anthracis sap gene awaits further analysis.

The amino acid sequence of Sap was compared with those of other S-layer proteins. Sequence similarities were found for the N-terminal regions of Sap, two other S-layer proteins (MWP from B. brevis and the S-layer protein from A. kivui), and the T. maritima OmpA protein. These sequences contain internal repeats that could play an important role in the assembly of the layer. Interestingly, proteins implicated in the attachment of Clostridium thermocellum cellulosome to the membrane also have similarities with the N-terminal repeats.

![FIG. 6. Electron micrographs of the cell walls of B. anthracis cells. Cells were negatively stained with ammonium molybdate. (a) Sap+ strain 9131; (b) Sap− mutant strain RBA2; (c) ultrathin section of intact cell of Sap+ strain 9131; (d) ultra thin section of intact cell of Sap− mutant strain RBA2. S, S-layer; pg, peptidoglycan; cm, cytoplasmic membrane. Bar, 0.1 μm.](http://jb.asm.org/)

618 ETIENNE-TOUMELIN ET AL. J. BACTERIOL.
from the S-layer proteins (7). Therefore, these repeat regions may be important in the interaction between the peptidoglycan wall and the S-layer structure. Moreover, a highly charged sequence was found to be repeated four times in Sap. It could have an important structural role. It has been suggested that the T. maritima OmpA protein connects the outer membrane to the inner cell body, with its carboxy-terminal hydrophobic tail probably being buried in the membrane. Therefore, the B. anthracis Sap protein may anchor the capsule to the peptidoglycan wall, utilizing both its repeated sequences and its hydrophobic C-terminal tail.

The function and the in vivo expression of the B. anthracis S-layer remain unknown. In the absence of the S-layer, dramatic morphological changes are observed, under both liquid and solid culture conditions. Since synthesizing the S-layer is energy-consuming, it would be interesting to determine where the pressure to maintain it occurs: in the soil or in the host. A major cell-associated protein with a molecular mass of 91 kDa (EA1) has been reported to be a predominant antigen following vaccination with the live spore vaccine (5). It should now be of interest to determine if EA1 corresponds to the Sap protein.

If the host is where the S-layer is maintained, its role cannot be simply to protect against complement killing or to enhance the binding of the bacterium to host molecules or macrophages, since in vivo the bacteria are encapsulated. Therefore, although B. anthracis probably goes through few, if any, life cycles in the soil, an eventual protective role against osmotic pressure could exist in this environment where the S-layer could be the outermost wall structure. In the future, an extensive study of the synthesis of the S-layer in vivo as well as that of an eventual interaction between the S-layer and the capsule will have to be undertaken.

ACKNOWLEDGMENTS

We are indebted to P. Gounon and M. C. Prévost, (Station Centrale de Microscopie Electronique) for the electron microscopy stereotyping plates and to J. P. Péron for the N-terminal peptide sequences. We thank P. Trieu-Cuot for the gift of the spectinomycin resistance cassette and S. Kochi for critical reading of the manuscript and helpful discussion.

REFERENCES

22a. Trieu-Cuot, P. Personal communication.