MicroReview
Expression, secretion and antigenic variation of bacterial S-layer proteins

Hein J. Boot1 and Peter H. Pouwels1,2

1BioCentrum Amsterdam, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands.
2TNO Nutrition and Food Research Institute, PO Box 5815, 2280 HV Rijswijk, The Netherlands.

Summary
The function of the S-layer, a regularly arranged structure on the outside of numerous bacteria, appears to be different for bacteria living in different environments. Almost no similarity exists between the primary sequences of S-proteins, although their amino acid composition is comparable. S-protein production is directed by single or multiple promoters in front of the S-protein gene, yielding stable mRNAs. Most bacteria secrete S-proteins via the general secretory pathway (GSP). Translocation of S-protein across the outer membrane of Gram-negative bacteria sometimes occurs by S-protein-specific branches of the GSP. O-polysaccharide side-chains of the lipopolysaccharide component of the cell wall of Gram-negative bacteria appear to function as receptors for attachment of the S-layer. Silent S-protein genes have been found in Campylo-bacter fetus and Lactobacillus acidophilus. These silent genes are placed in the expression site in a fraction of the bacterial population via inversion of a chromosomal segment.

Introduction
More than 300 different species of Eubacteria and Archaeabacteria possess a regularly arranged structure which covers the cell envelope, named the S-layer (short for Surface-layer). The mono-molecular S-layer is the result of the secretion and subsequent crystallization of a single protein (S-protein), which is in some cases glycosylated or phosphorylated (Messner and Sleytr, 1992). S-proteins contain all the information necessary for their crystallization into a regular structure as demonstrated by crystallization in vitro of isolated S-proteins into authentic S-layers.

S-layers are found on the surfaces of bacterial species living in very different environments, signifying the importance of S-layers. Although some particular functions could be assigned to specific S-proteins, such as barriers between the cell and the environment (Paula et al., 1988) and attachment structure for extracellular enzymes (Matuschek et al., 1994), no general function has been described which is found for all S-layers. For Archaeabacteria, it has been reported that the S-layer plays a role in maintenance of the cell shape (Pum et al., 1991). The S-layer found on the outside of pathogenic species belonging to the genera Aeromonas and Campylobacter is involved in pathogenesis; loss of the S-layer results in severe reduction of the virulence of Aeromonas salmonicida. S-layers of pathogenic bacteria but also of non-pathogenic bacteria such as Lactobacillus subsp. have been shown to interact with receptors on epithelial cells of the host (Borinski and Holt, 1990; Toba et al., 1995).

For more details regarding the structure and function of S-layers, the reader is referred to other reviews (Messner and Sleytr, 1992; Beveridge, 1994; Sleytr et al., 1995). In this MicroReview, we will focus on the genetic information for the synthesis of S-proteins, the secretion of S-proteins and the interaction of secreted S-proteins with cell-wall components. We will also discuss the presence of silent S-protein genes in relation to S-layer variation which has recently been described for pathogenic and non-pathogenic bacterial species.

Primary structure of S-proteins
The DNA sequence of S-protein genes has been determined for several species (see Table 1). No overall alignment of the S-protein genes can be given because sequence similarity only exists between genes of related species. The deduced molecular mass of the different S-proteins ranges from 46 kDa (Lactobacillus spp.) to 169 kDa (Rickettsia spp.). Several post-translational modifications such as glycosylation, phosphorylation and cleavage of N- or C-terminal fragments, have been reported to influence the final mass of the mature S-protein.
Although the S-proteins of non-related species do not show overall sequence homology, the deduced amino acid composition is quite similar. The amino acid composition is characterized by the relative abundance of threonine, serine, and hydrophobic amino acids, while cysteine and methionine are, in most cases, absent. Most S-protein genes show a strongly biased codon usage, typical for genes that are efficiently translated (our unpublished observations). Except for S-proteins of lactobacilli (pi > 9.5) and Methanothermus (pi = 8.4), all other S-proteins of which the primary structure is known have an acidic isoelectric point. Multiple copies of an S-layer homology (SLH (Lupas et al., 1994)) domain have been found in either the N- or C-terminal part of several S-proteins. This SLH domain is thought to be involved in interaction with either the cell wall or extracellular proteins (see below).

Expression of S-protein genes

Electron-microscopic analysis of bacteria which possess an S-layer reveals that the bacterial surface is completely covered with an S-layer during the entire growth cycle. During division of Clostridium thermosaccharolyticum, an excess of S-protein is found at the site of separation of the two daughter cells, which prevents the exposure of the newly synthesized parts of the cell wall to the environment (Sleytr and Glauert, 1976). Several bacteria produce an excess of S-protein to ensure complete coverage of the cell wall during all growth phases, and either store excess S-protein in the peptidoglycan layer or secrete it into the environment. An additional S-layer is formed on the inner face of cell walls after disruption of Bacillus stearothermophilus cells and isolation of cell-wall fragments (Breitwieser et al., 1992). S-protein overproduction also explains the fast regeneration of the S-layer after it has been extracted by non-lethal procedures, as noted for Lactobacillus helveticus (Lortal et al., 1992), and Methanococcus voltae (Firtel et al., 1995). In contrast Corynebacterium glutamicum cells have patches of uncovered cell wall when grown on liquid medium to stationary phase (Chamet al., 1995). It is unclear whether synthesis and/or secretion of S-protein is limiting for this species at specific growth conditions, or whether the secreted S-protein just fails to constitute an S-layer and is released into the culture fluid.

Table 1. Properties of S-protein genes and deduced S-proteins.

<table>
<thead>
<tr>
<th>Species</th>
<th>Genbank Accession No.</th>
<th>No. of Amino Acid Residues encoded by ORF</th>
<th>N-terminal secretion signal</th>
<th>No. of transcription start points</th>
<th>Length (nt) of 5' untranslated leader</th>
<th>Silent genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetogenium kivui</td>
<td>M31069</td>
<td>762</td>
<td>26</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Aeromonas hydrophila</td>
<td>L37348</td>
<td>467</td>
<td>19</td>
<td>n.d.</td>
<td>n.d.</td>
<td>None</td>
</tr>
<tr>
<td>Aeromonas salmonicida</td>
<td>M64565</td>
<td>502</td>
<td>21</td>
<td>1</td>
<td>180</td>
<td>None</td>
</tr>
<tr>
<td>Bacillus anthracis</td>
<td>Z36946</td>
<td>814</td>
<td>29</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Bacillus brevis HPD31</td>
<td>D90050</td>
<td>1087</td>
<td>23 or 53</td>
<td>5</td>
<td>56 → 358</td>
<td>n.d.</td>
</tr>
<tr>
<td>Bacillus brevis 47 (MWP)</td>
<td>M19115</td>
<td>1053</td>
<td>23</td>
<td>5</td>
<td>115 → 358</td>
<td>None</td>
</tr>
<tr>
<td>Bacillus brevis 47 (OWNP)</td>
<td>M14238</td>
<td>1004</td>
<td>24</td>
<td>_b</td>
<td>_b</td>
<td>None</td>
</tr>
<tr>
<td>Bacillus sphaericus 2362</td>
<td>M23631</td>
<td>1176</td>
<td>30</td>
<td>n.d.</td>
<td>84</td>
<td>1c</td>
</tr>
<tr>
<td>Bacillus steatherophilus</td>
<td>X71092</td>
<td>1228</td>
<td>30</td>
<td>n.d.</td>
<td>n.d.</td>
<td>1c</td>
</tr>
<tr>
<td>Campylobacter fetus</td>
<td>L15800</td>
<td>933</td>
<td>_</td>
<td>1</td>
<td>114</td>
<td>6 to 8</td>
</tr>
<tr>
<td>Caulobacter crescentus</td>
<td>M84750</td>
<td>1026</td>
<td>_</td>
<td>1</td>
<td>61</td>
<td>None</td>
</tr>
<tr>
<td>Corynebacterium glutamicum</td>
<td>X69103</td>
<td>510</td>
<td>30</td>
<td>n.d.</td>
<td>n.d.</td>
<td>None</td>
</tr>
<tr>
<td>Deinococcus radodurans</td>
<td>M17895</td>
<td>1036</td>
<td>31 → 59</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Halobacterium halobium</td>
<td>J02767</td>
<td>852</td>
<td>34</td>
<td>1</td>
<td>111</td>
<td>n.d.</td>
</tr>
<tr>
<td>Haloferax volcanii</td>
<td>M62816</td>
<td>828</td>
<td>34</td>
<td>n.d.</td>
<td>n.d.</td>
<td>None</td>
</tr>
<tr>
<td>Lactobacillus acidophilus</td>
<td>X71412</td>
<td>444</td>
<td>31</td>
<td>1</td>
<td>191</td>
<td>1</td>
</tr>
<tr>
<td>Lactobacillus brevis</td>
<td>Z14250</td>
<td>465</td>
<td>30</td>
<td>2</td>
<td>130 or 43</td>
<td>None</td>
</tr>
<tr>
<td>Methanococcus voltae</td>
<td>M59200</td>
<td>565</td>
<td>12</td>
<td>3</td>
<td>104 → 245</td>
<td>n.d.</td>
</tr>
<tr>
<td>Methanosarcina mazeli</td>
<td>X77929</td>
<td>625</td>
<td>31</td>
<td>n.d.</td>
<td>n.d.</td>
<td>None</td>
</tr>
<tr>
<td>Methanothermus fervidus</td>
<td>X56297</td>
<td>593</td>
<td>22</td>
<td>1</td>
<td>33</td>
<td>None</td>
</tr>
<tr>
<td>Methanothermus sociabilis</td>
<td>X58296</td>
<td>593</td>
<td>22</td>
<td>1</td>
<td>33</td>
<td>None</td>
</tr>
<tr>
<td>Rickettsia prowazekii</td>
<td>M37647</td>
<td>1612</td>
<td>32</td>
<td>n.d.</td>
<td>n.d.</td>
<td>None</td>
</tr>
<tr>
<td>Rickettsia rickettsii</td>
<td>X16353</td>
<td>1300</td>
<td>34</td>
<td>1</td>
<td>130</td>
<td>None</td>
</tr>
<tr>
<td>Thermus thermophilus</td>
<td>X57333</td>
<td>917</td>
<td>27</td>
<td>1</td>
<td>127</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

a. Only the properties of expressed S-protein-encoding genes are shown in this table.
b. Polycistronic messenger.
c. At least one truncated, silent gene is present (Bowditch et al., 1989).
d. Based on protein analysis, it is postulated that additional gene(s) are present (Saira and Sleytr, 1994).

ORF, open reading frame; n.d., not determined.
S-protein represents approx. 10 to 15% of the total cellular protein of the bacterial cell. S-protein genes thus must be efficiently transcribed and/or yield stable mRNA. Multiple promoters have been found in front of the S-protein genes of several species, while S-protein genes of other species are preceded by a single promoter only (Table 1). Transcription directed by multiple promoters might contribute to a higher yield of mRNA compared to transcription directed by a single promoter. Multiple promoters may also provide an opportunity to regulate S-protein gene expression, if different promoters are active under different physiological conditions. One of the two promoters identified upstream of the S-protein gene of \( A. \) \( \text{salmonicida} \) is indeed under the control of a DNA-binding protein. This particular promoter is most active during the early stages of growth in a heterologous system (Noonan and Trust, 1995a). One of the five identified promoters in front of the S-protein gene in \( B. \) \( \text{brevis} \) is active only during the exponential phase of growth, while another promoter is active during all growth stages (Adachi et al., 1989).

The half-life of the S-protein mRNAs has only been determined for three species, \( C. \) \( \text{crescentus} \) (10 to 15 min; Fisher et al., 1988), \( A. \) \( \text{salmonicida} \) (22 min; Chu et al., 1993), and \( L. \) \( \text{acidophilus} \) (15 min; Boot et al., 1996a). Secondary-structure formation in the long, untranslated leader sequences, found for many S-protein mRNAs (Table 1), might contribute to such unusually long half-lives. Computer analysis predicts that the long, untranslated leader sequences of the S-protein mRNAs of \( A. \) \( \text{salmonicida} \) and \( L. \) \( \text{acidophilus} \) indeed can fold into stable secondary structures with energies of \(-218\) and \(-191\) kJ mol\(^{-1}\), respectively (Boot et al., 1996a; our unpublished results).

**Secretion of S-proteins**

Secretion of most extracellular proteins occurs via the general secretory pathway (GSP), involving a signal peptide/Sec-dependent transport route to cross the cytoplasmic membrane and, in Gram-negative bacteria, additional, substrate-specific branch pathways for transport across the outer membrane (Pugsley, 1993). Although most S-proteins possess an N-terminal signal peptide that is cleaved off to yield mature S-protein (Table 1), their secretion routes appear to have evolved independently. From the limited data on S-protein secretion, it appears that some bacteria use specific S-protein pathways.

S-proteins of \( A. \) \( \text{salmonicida} \) display significant sequence homology with the PulE family of proteins (Pugsley, 1993; Thomas and Trust, 1995). Protein ApsE, sharing sequence homology with the PulE family of proteins, which are a class of ATP-binding cassette (ABC) proteins believed to energize a terminal branch of the GSP, is necessary for S-protein transport in \( A. \) \( \text{salmonicida} \) (Noonan and Trust, 1995b). Insertional inactivation of \( \text{apsD} \) or \( \text{apsE} \) results in accumulation of S-protein in the periplasm, but does not affect secretion of other extracellular proteins.

The genes involved in synthesis and secretion of S-protein in \( A. \) \( \text{salmonicida} \) (\( \text{abcA} \), \( \text{apsE} \) and \( \text{vapA} \)) display a nucleotide composition, codon usage and AT drift that are strikingly different from that of other \( A. \) \( \text{salmonicida} \) genes (Noonan and Trust, 1995b; our unpublished observations). The nucleotide-composition features of these genes strongly resemble those of the AT-rich branch of Gram-positive bacteria, which might suggest that they were acquired from another bacterium. The fact that \( A. \) \( \text{salmonicida} \) has developed specific terminal branches of GSP for S-protein secretion may reflect the need of S-proteins to remain unfolded while in the periplasm, thereby preventing premature multimerization. This situation is reminiscent of the secretion of P-pilus fimbriae which are protected from multimerization by binding to chaperone proteins (Hultgren et al., 1991). The observation that ApsE shows significantly greater similarity to the \( P. \) \( \text{aeruginosa} \) fimbrial protein PilB, which is a cytoplasmic chaperone implicated in pilus assembly, than to the PulE family of proteins (34% versus 20–28% similarity), lends support to this notion. Because S-proteins, in general, lack cysteine residues, folding of S-proteins in the periplasm owing to S–S-bridge formation cannot occur. The absence of sufficient amounts of chaperone in the case of excessive production of S-proteins may cause premature, intracellular assembly, as was observed for the \( B. \) \( \text{stearothermophilus} \) S-protein in \( E. \) \( \text{coli} \) (Kuen et al., 1995).

The S-proteins of \( C. \) \( \text{crescentus} \) and \( C. \) \( \text{jejuni} \) are probably secreted by the Sec-independent type I secretion system (Wandersman, 1992), as these S-proteins are missing an N-terminal signal peptide. The assumed type I secretion of these S-proteins is supported by studies using fusions of part of the \( C. \) \( \text{crescentus} \) S-protein with a reporter protein. The C-terminal but not the N-terminal region of the S-protein was found to be sufficient to direct the secretion of hybrid proteins (Bingle and Smit, 1994). Secretion of the S-protein was also found to be dependent on a cluster of genes near the
S-protein gene, including an ABC transporter gene (J. Smit, personal communication).

Little is known about secretion of S-proteins of Gram-positive bacteria and Archaeabacteria. Based on theoretical considerations, it is thought that incorporation of S-protein into an existing S-layer occurs at places of crystal dislocation (Pumet al., 1991). Newly synthesized S-protein of B. stearothermophilus (Gruber and Sleytr, 1988) and Bacillus sphaericus (Howard et al., 1982) is incorporated at discrete bands along the lateral surface of the cell near the site where cell division takes place. Whether the S-protein is only secreted at the position where it is incorporated, or whether the S-protein is secreted throughout the whole cell and diffuses through the cell wall to a point where it can be incorporated, is unknown.

Interaction between S-proteins and cell-wall components

S-proteins of several Gram-negative bacteria have been shown to interact with O-polysaccharide side-chains of the lipopolysaccharide (LPS) component of the outer membrane. A. hydrophila, Aeromonas sobria and C. crescentus strains carrying an S-layer contain homogenously-length O-polysaccharide side-chains, while strains which lack an S-layer are devoid of O-polysaccharide side-chains (Kokka et al., 1990; Walker et al., 1994). In line with this observation, it was found that a variant strain of A. salmonicida which had lost its smooth O-polysaccharide side-chains secretes the S-protein into the medium, while a variant strain with a partial loss of the O-polysaccharide side-chains possesses a disrupted S-layer (Griffiths and Lynch, 1990).

Two different types of C. fetus strains exist (A and B), which can be differentiated on the basis of differences in LPS composition. The S-proteins of these strains only show interaction with the homologous LPS type (Dworkin et al., 1995). S-proteins expressed in A and B strains differ in the N-terminal one-fifth of the amino acid sequences and are, except for a few (conservative) mutations, identical for the rest of the sequence. As S-proteins of A- and B-type strains show interaction only with the homologous LPS, it is clear that the N-terminal part of the S-protein is responsible for type-specific interaction. For A. hydrophila, it was found, however, that the C-terminal but not the N-terminal part of the S-protein is essential for interaction with the O-polysaccharide chains of the LPS (Thomas et al., 1992).

A cell-wall polymer of B. stearothermophilus consisting of glucose and N-acetylglucosamine is probably responsible for interaction with the S-protein (Sára et al., 1996). Approx. two-thirds of the S-protein of B. sphaericus P-1, counting from the C-terminus, could be deleted without detectable influence on the production, secretion and interaction of the truncated S-protein with the underlying cell wall (Deblaere, 1995). The N-terminal part of this S-protein is apparently necessary and sufficient for the protein to be retained on the cell wall.

SLH domains are not only found in S-proteins, but have also been found in some extracellular enzymes that are non-covalently linked to the bacterial cell wall, such as pullulanase and endoxylanase. The SLH domain was postulated to be the determinant for interaction of the S-protein and other extracellular proteins with the underlying peptidoglycan layer (Matuschek et al., 1994). Another explanation is that the SLH domain of such extracellular enzymes interacts with (SLH) domains of the S-protein (Egelseer et al., 1995).

Several bacterial species possess two S-layers with a different regular structure on top of each other. The two S-layers of B. brevis (the middle-wall protein (MWP) and outer-wall protein (OWP) S-layer) are formed by two different S-proteins (Tsuboi et al., 1988). The MWP S-layer interacts with the peptidoglycan layer and the OWP S-layer is attached onto the MWP S-layer. Expression of MWP in a heterologous host results in a protein capable of crystallization onto the isolated peptidoglycan layer of B. brevis, yielding an authentic crystalline structure. No such interaction was found when a truncated MWP with a deletion of 20% of the C-terminal end was used. Apparently, the C-terminal end of the MWP is important for interaction with receptors present in the peptidoglycan layer. The presence of two S-protein genes organized in an operon and giving rise to a double S-layer differs clearly from the multiple S-protein genes found in other bacterial species, where only one S-protein gene is expressed at a time (see below).

S-layer variation and silent S-protein genes

Some S-layers have been shown to be involved in the interaction of bacteria with receptors of the host mucosa. Furthermore, it has been shown that the S-protein, if present, is one of the most-dominant antigenic determinants of pathogenic bacteria. The combination of these properties makes S-layers a likely candidate for antigenic variation.

A family of S-proteins (100, 127 and 149 kDa) all having the same N-terminal amino acid sequence but yielding S-layers of different symmetry was reported for several C. fetus strains (Pei et al., 1988; Fujimoto et al., 1991). Cloning and sequencing of the S-protein genes of C. fetus 23D revealed that the different S-proteins originated from multiple S-protein genes. An active S-promoter was found upstream of the S-protein gene (sapA) for the wild-type strain (Dworkin and Blaser, 1996). Inversion of a 6.2 kb chromosomal segment containing the S-promoter

© 1996 Blackwell Science Ltd, Molecular Microbiology, 21, 1117–1123
leads to the expression of either one of the two S-protein genes (sapA or sapA2) surrounding this invertable segment (Fig. 1). A region of sequence identity (650 nucleotides (nt)), which contains a Chi-like sequence and a palindrome sequence (15 nt), is found in the 5'-part of each S-protein gene (Tummuru and Blaser, 1993). The recombination sites of the inverted segment are found within this identity region. How the S-proteins which are not surrounding the inverted segment are activated is presently unknown. As found for other pathogenic bacterial species, the S-protein of C. fetus is a major immunogenic determinant. Consistent with this notion is the observation that the immune response of an infected host is the selecting determinant for in vivo S-layer variation of C. fetus.

S-layer variation is also expected to occur in Lactobacillus acidophilus ATCC 4356. This strain contains two S-protein genes: the actively transcribed slpA gene and the silent slpB gene. The N-terminal and middle parts of the deduced proteins show reduced similarity, while the C-terminal one-third parts are almost identical (Boot et al., 1995). The two S-protein genes are located 3.0 kb apart on the chromosome, in a reverse orientation relative to each other (Fig. 2). The slpA gene is interchanged with the slpB gene through inversion of a chromosomal segment (Boot et al., 1996b). Inversion is taking place between two regions of sequence identity located between the S-promoter and the sequence encoding the mature part of the S-proteins. A 26 bp sequence having a high degree of similarity with the consensus recognition site of the Din family of invertases is present in the middle of this region, suggesting that inversion is catalysed by a member of this family. Lactobacilli are inhabitants of the gastrointestinal tracts and are claimed to be beneficial for the host. Thus, it seems unlikely that S-layer variation of non-pathogenic lactobacilli has the same function as S-layer variation for pathogenic organisms such as C. fetus, namely to circumvent an immune response of the infected host.

B. sphaericus 2362 contains a truncated S-protein gene located 11 kb downstream from the gene encoding the 125 kDa S-protein. This truncated gene, which lacks a translation start, can encode a polypeptide of 80 kDa (Bowditch et al., 1989). The N-terminal part of the hypothetical 80 kDa protein is almost identical to the middle part of the 125-kDa S-protein, while the C-terminal part of the hypothetical 80 kDa protein has a reduced similarity with the C-terminal part of the 125-kDa S-protein. Variation of the molecular mass of S-proteins of a B. sphaericus strain has also been noted (Hastie and Brinton, 1979; Howard and Tipper, 1973), but a detailed molecular study has not yet been performed to determine whether this variation is the result of a chromosomal recombination event between the expressed and (truncated) silent gene(s).

Variation of expressed S-proteins is also known for strains of B. stearothermophilus (Sára and Sleytr, 1994). Wild-type S-protein of B. stearothermophilus is replaced by a variant S-protein when an increased oxygen pressure is used during controlled growth in a fermenter. The biochemical properties of the wild-type and variant S-proteins are almost the same, but the structure of the S-layers formed is different. Based on proteolytic
degradation of wild-type and variant S-proteins, it is clear that the primary sequence of the S-proteins is different. It was postulated that the change of S-protein is accompanied by a chromosomal rearrangement, suggesting that chromosomal recombination is involved.

Conclusion

Despite their general occurrence, no general function is known for the bacterial S-layer. The efficient production of S-proteins, required for the formation of an intact S-layer on growing bacteria, requires a high energy input, indicating the importance of the S-layer during evolution. Recently, silent S-protein genes have been found in both Gram-positive and Gram-negative species, offering these bacteria the opportunity to vary their S-layer. S-layers have been shown to be important for efficient infection of several bacterial species. On the other hand (non-)specific blockage of epithelial receptors by S-layer-possessing, probiotic bacteria might help to prevent bacterial infections. More knowledge about S-proteins, S-layers or S-layer variation may possibly assist in defining strategies to prevent certain bacterial infections.

References


© 1996 Blackwell Science Ltd, Molecular Microbiology, 21, 1117–1123


