Novel surface layer protein genes in *Bacillus sphaericus* associated with unusual insertion elements

Katrin Pollmann, Johannes Raff, Michaela Schnorpfeil, Galina Radeva and Sonja Selenska-Pobell

Institute of Radiochemistry, Forschungszentrum Rossendorf, D-01314 Dresden, Germany

The surface layer (S-layer) protein genes of the uranium mining waste pile isolate *Bacillus sphaericus* JG-A12 and of its relative *B. sphaericus* NCTC 9602 were analysed. The almost identical N-termini of the two S-layer proteins possess a unique structure, comprising three N-terminal S-layer homologous (SLH) domains. The central parts of the proteins share a high homology and are related to the S-layer proteins of *B. sphaericus* CCM 2177 and P-1. In contrast, the C-terminal parts of the S-layer proteins of JG-A12 and NCTC 9602 differ significantly between each other. Surprisingly, the C-terminal part of the S-layer protein of JG-A12 shares a high identity with that of the S-layer protein of *B. sphaericus* CCM 2177. In both JG-A12 and NCTC 9602 the chromosomal S-layer protein genes are followed by a newly identified putative insertion element comprising three ORFs, which encode a putative transposase, a putative integrase/recombinase and a putative protein containing a DNA binding helix–turn–helix motif, and the S-layer-protein-like gene copies *sllA* (9602) or *sllB* (JG-A12). Interestingly, both *B. sphaericus* strains studied were found to contain an additional, plasmid-located and silent S-layer protein gene with the same sequence as *sllA* and *sllB*. The primary structures of the corresponding putative proteins are almost identical in both strains. The N-terminal and central parts of these S-layer proteins share a high identity with those of the chromosomally encoded functional S-layer proteins. Their C-terminal parts, however, differ significantly. These results strongly suggest that the S-layer protein genes have evolved via horizontal transfer of genetic information followed by DNA rearrangements mediated by mobile elements.

INTRODUCTION

Regularly structured paracrystalline surface layers (S-layers) are one of the most common surface structures found in bacteria and archaea. Most of them are composed of protein or glycoprotein monomers with the ability to self-assemble in two-dimensional arrays (Sára & Sleytr, 2000; Sidhu & Olsen, 1997). In most cases, they are the major protein species, constituting up to 15% of the total protein produced by the cells (Kuen et al., 1994). As the outermost cell structure the S-layers serve as an interface between the bacterial cell and the environment. Several functions have been ascribed to S-layers: a molecular sieve (Sára & Sleytr, 1987), virulence factors in several pathogenic bacteria (Ishiguro et al., 1981; Lewis et al., 1987; Mesnage et al., 2001; Mignot et al., 2001), an attachment structure for high-molecular-mass extracellular proteins such as amylases (Matuschek et al., 1994; Egelseer et al., 1995, 1996), an adhesin with affinity for human epithelial cells and fibronectin (Hynönen et al., 2002), and sorption of toxic heavy metal ions (Raff, 2002; Merroun et al., 2005).

It is essential for micro-organisms to be able to respond rapidly to changes in the environment. One way in which they can do this is to alter their surface properties by varying protein expression through programmed DNA rearrangements (Borst & Greaves, 1987). Such DNA rearrangements, which can be mediated by a variety of mechanisms, affect the phenotypic characteristics of surface structures, e.g. for
evasion of host defences, antigenic variation or fine tuning of surface structures (Dybvig, 1993). Recombinational events have also been described for the regulation of the expression of S-layer protein genes from various bacterial strains. It is assumed that bacteria carrying an S-layer use S-layer variation to adapt to different stress factors (Jakav-Viljanen et al., 2002; Kuen et al., 1997; Mignot et al., 2001, 2002). Chromosomal rearrangements include the inversion of promoters located between two oppositely orientated S-layer gene cassettes (Dworkin & Blaser, 1996), recombination of S-layer gene segments (Blaser et al., 1994), and inversion of chromosomal segments containing two variants of S-layer-encoding genes orientated in opposite directions, resulting in the placement of the formerly silent gene behind the promoter (Boot et al., 1996a, b; Boot & Pouwels, 1996). In the case of Geobacillus stearothermophilus PV72, S-layer protein variation is mediated by the replacement of the chromosomally encoded active S-layer protein gene sbsA with the plasmid-encoded formerly silent S-layer protein gene sbsB, resulting in the activation of sbsB (Scholz et al., 2001).

The uranium mining waste pile isolate B. sphaericus JG-A12 is enveloped by an S-layer with a square symmetry, which is composed of identical protein monomers (Raff, 2002). The cells of B. sphaericus JG-A12, like those of the closely related B. sphaericus NCTC 9602, are able to bind selectively and reversibly high amounts of metals such as uranium, lead, copper, aluminium, gallium and cadmium from drain waters of uranium wastes (Selsenska-Pobell et al., 1999). Further analyses showed that the purified and recrystallized S-layers of both strains bind high amounts of uranium in a strain-specific way (Raff, 2002; Raff & Selenska-Pobell, 2004). This indicates that the S-layers play a key role in interactions of the bacteria with metals (Raff, 2002). Interestingly, the S-layers of the B. sphaericus strains JG-A12 and NCTC 9602 both have square symmetry with almost identical lattice constants, and their protein monomers have the same size and almost identical N-terminal SLH domains (Raff, 2002). However, they differ significantly in their pH and proteolytic stability and in their kinetics of uranium binding (Raff, 2002; Raff et al., 2004). Recently we demonstrated that the S-layer of strain JG-A12 possesses a significantly higher binding capacity to uranium in solution: in contrast to strain NCTC 9602, the S-layer of strain JG-A12 is able to bind almost all uranium dissolved in water in environmentally relevant concentrations of 5 mg l$^{-1}$ and below (Raff & Selenska-Pobell, 2004).

In order to elucidate the underlying causes of these differences, we studied the primary structures of the two S-layer proteins. In the present work the chromosomally encoded S-layer protein genes of B. sphaericus JG-A12 and B. sphaericus NCTC 9602 and their upstream and downstream regions were sequenced. In addition to the functional S-layer genes, a second silent, plasmid-located S-layer-protein-like gene was found and characterized in both the strains. The chromosomal S-layer protein genes are followed by a putative mobile gene element carrying a truncated S-layer-protein-like gene with the same sequence as the plasmid-encoded silent gene. To our knowledge this is the first report of an insertion element encoding an S-layer-protein-like gene as well as a transposase.

**METHODS**

**Bacterial strains and culture conditions.** B. sphaericus JG-A12 was recovered from a soil sample collected from the uranium mining waste pile Haberland near the town of Johannegeorgenstadt, Germany (Selenska-Pobell et al., 1999). B. sphaericus NCTC 9602 was purchased from the National Collection of Type Cultures, London, UK. Both B. sphaericus strains were routinely grown in nutrient broth (NB) medium consisting of 5 g peptone l$^{-1}$ and 3 g meat extract l$^{-1}$. For isolation of RNA from bacterial cells treated with Cd$^{2+}$, the strains were grown in defined medium (Chen et al., 1973). In the early exponential growth phase a solution of CdCl$_2$ was added to give a final concentration of 50 $\mu$M Cd$^{2+}$. The cells were harvested after 3 h incubation at 30 °C.

**DNA manipulations.** For PCR amplification genomic DNA of B. sphaericus NCTC 9602 and B. sphaericus JG-A12 was isolated and purified using the Nucleospin kit (Machery-Nagel) according to the manufacturer’s instructions. For the construction of vectorette libraries and for Southern hybridizations, genomic DNA was prepared as described by Ausubel et al. (1993). Digestion of DNA with restriction endonucleases (Promega, Invitrogen) and separation of DNA fragments by agarose gel electrophoresis were performed by published methods (Sambrook et al., 1989).

**PCR, oligonucleotides, vectorette libraries and sequence analysis.** PCR amplifications were carried out using a T3 Thermocycler (Biometra). Oligonucleotide synthesis was done by MWG-Biotech AG. The sequences of all oligonucleotides used are listed in Supplementary Table S1 with the online version of this paper. All amplification reactions were performed in a volume of 20 $\mu$l using Taq DNA polymerase (Promega), and the conditions were optimized for each primer pair. The Universal Vectorette System UVS-1 (Sigma-Aldrich) was chosen to perform asymmetric PCR after host-independent cloning. For the construction of different vectorette libraries, 1 $\mu$g genomic DNA from each of the strains B. sphaericus NCTC 9602 and JG-A12 was digested with the restriction endonucleases BglIII, ClaI, EcoRI, EcoRV, HaeIII, HindIII, SstI or TsuI. The corresponding libraries were constructed according to the manufacturer’s instructions. PCR amplifications of the ligated fragments were carried out in a 20 $\mu$l reaction volume containing 1 $\mu$l of the vectorette library as template, 1 nmol each of dATP, dCTP, dGTP and dTTP (Invitrogen), 2.5 mM MgCl$_2$, 1.6 $\mu$g BSA, 20 pmol Vectorette Primer, 20 pmol sequence-specific primer, and 1 unit Taq polymerase in 1 x reaction buffer.

For cloning of PCR products the TOPO-TA cloning kit (Invitrogen) was used according to the manufacturer’s instructions. Plasmid isolation was done using Wizard Plus SV Miniprep (Promega). The plasmids were used as template for PCR amplification of S-layer gene fragments using insert-specific primers. The PCR products obtained were purified using the Quickstep 2 PCR Purification Kit (Edge Bio Systems) and sequenced from each strand using a Perkin-Elmer Applied Biosystems 377 instrument.

**Isolation of RNA.** Total RNA was isolated from exponentially growing B. sphaericus NCTC 9602 and JG-A12 harvested at an OD$_{600}$ of 0.6 (NB medium) and from cells grown in defined...
medium treated with Cd(NO₃)₂ using the RNase Mini Kit (Qiagen) according to the manufacturer’s instructions. The remaining DNA was digested on-column by using the RNase free DNase Set (Qiagen). RNA isolation and blotting was repeated three times.

**Northern and Southern hybridization.** After digestion of total DNA of *B. sphaericus* JG-A12 and NCTC 9602, DNA fragments were separated on a 1% agarose gel (Seakem LE). Southern blots were performed on positively charged nylon membranes (Roche) using a TurboBlotter (Schleicher & Schuell), following the manufacturer’s instructions. Northern blotting of total RNA was performed according to published methods (Sambrook *et al*., 1989), by applying 2 μg total RNA of Cd(NO₃)₂-treated cells, and 5 μg total RNA of NB-grown cells of JG-A12 or 9602, respectively. For hybridization of the Southern and Northern blots, DNA probes specific for S-layer protein genes were labelled with DIG-dUTP by incorporation during PCR using the PCR DIG Probe Synthesis Kit (Roche).

The filters were hybridized with a 922 bp PCR-generated probe obtained with primer pair Mobi1_F/Mobi1_R (5’-ATCGCCCATTTCGTTGCTTGGGA-3’/5’-ATGTTACGTCTACCGAAAATACTCGA-3’), comprising a part of the putative transposase gene of JG-A12 and 9602, with a PCR-generated probe obtained with primer pair V83/R71 (5’-ATGCCTTTATCTCCAATAATATC-3’/5’-AAATTGTAGTACTGGGTACCTGCGCAGA-3’), comprising the putative recombination genes of 9602 and JG-A12, with an 857 bp PCR-generated probe obtained by using primer pair V8/R30 (5’-AAGAAATCGGAGGGATG-3’/5’-TTTTRTAAGTACCATTTGAATTTI-3’) comprising the central region of the S-layer protein gene of JG-A12, or with PCR-generated probes obtained with primer pair V37/R64 (5’-GTTGCTGGTTACCTGTAATAG-3’), comprising the 3’-region of the *slf* genes. Hybridization temperature was 38 °C. Northern and Southern blots were developed using the DIG Nucleic Acid Detection Kit (Roche). The size of the DNA fragments was estimated using the DIG-dUTP-labelled DNA Molecular Weight Marker II (Roche). The size of the S-layer-specific mRNA was estimated using the DIG-dUTP labelled RNA Molecular Weight Marker II (Roche).

**RT-PCR.** For reverse transcription of the S-layer mRNA followed by PCR the GeneAmp Reverse Transcriptase RNA PCR Kit (Applied Biosystems) was used. Reverse transcription and PCR were performed according to the manufacturer’s instructions. For amplification of the upstream regions of the genes, primer pair UR2F/UR2R (5’-CGCTGTTACCTGTAATAG-3’/5’-TAACGTATTGGAAMGTTAAC-3’) was used. For amplification of the 3’-regions of *slf* and *sll*, primer pairs R/V8/R71/29 (5’-ATCATCAGCATCTGATGGCGACGCACTGGTTACCTGTAATAG-3’), comprising the 3’-region of the *slf* genes. Hybridization temperature was 38 °C. Northern and Southern blots were developed using the DIG Nucleic Acid Detection Kit (Roche). The size of the DNA fragments was estimated using the DIG-dUTP-labelled DNA Molecular Weight Marker II (Roche). The size of the S-layer-specific mRNA was estimated using the DIG-dUTP labelled RNA Molecular Weight Marker II (Roche).

**Western blotting.** S-layer proteins of *B. sphaericus* NCTC 9602 and JG-A12 were purified as described previously (Raff, 2002). For digestion with the endoproteinase Glu-C (Sigma-Aldrich), 200 μg of purified S-layer proteins was resuspended in 32 μL of a 50 mM KH₂PO₄/Na₂HPO₄ buffer (pH 7.8). After addition of 2 μg of the endoproteinase Glu-C, the solution was incubated for 24 h at 37 °C. The resulting fragments were separated on a 12.5% SDS-PAGE (Laemmli, 1970) using a Mini-PROTEAN II (Bio-Rad) and transferred onto a PVDF membrane using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad), using 12 mM Tris base/8 mM CAPS/15% (v/v) methanol as cathode buffer and 12 mM Tris base/8 mM CAPS/1.14% (w/v) thioglycolate as anode buffer. The gel was blotted for 80 min at a current of 1.5 mA cm⁻². Visible protein bands were cut after staining the blotting membrane with Coomassie Brilliant Blue and destaining. The N-terminal amino acid sequences of the protein fragments were determined using an ABI 494A Precise HT sequencer (Applied Biosystems).

**Plasmid screening.** Aliquots of 500 μL *B. sphaericus* overnight cultures grown in NB were inoculated in 10 mL fresh NB and shaken at 30 °C for about 1 h to reach an OD₆₀₀ of 0.3. One millilitre of the fresh culture was pelleted by centrifugation at 6000 g. The cells were washed with TE buffer (10 mM Tris base, 1 mM EDTA, pH 8-0) and resuspended in 20 μL 25% sucrose in TE. Afterwards 20 μL of a lysis solution containing lysosome (4 mg mL⁻¹) and RNase A (2 mg mL⁻¹) was carefully added. Lysis of the cells was achieved directly in the wells of a low–percentage agarose gel and visualization of the plasmids was performed as described previously (Selenkaj-Trajkowa *et al*., 1990).

**Bioinformatic analysis of S-layer genes and proteins.** Sequence assembling was performed using AutoAssembler 2.0 (Applied Biosystems). For ORF predictions and restriction analyses BioEdit was used. Isoelectric point and molecular mass predictions were performed with the ExPASy Molecular Biology Server (http://www.expasy.org). Multiple alignments of protein sequences were performed using CLUSTALW. Predictions of secondary structures were done using the mfold program (Zuker *et al*., 1999; http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi). Computer searches and sequence comparisons were done using BLAST (Altschul *et al*., 1990).

**RESULTS**

**Characterization of the S-layer genes of *B. sphaericus* NCTC 9602 and JG-A12**

The entire sequence of the S-layer protein gene of *B. sphaericus* NCTC 9602 (*slfA*) (accession no. AJ849547) indicated one ORF of 3684 bp encoding a protein of 1228 aa. The ORF starts with ATG and is preceded by a typical ribosome-binding site (GGAGG) with a distance of 3 nt between the middle A of this sequence and the start codon. Two putative promoter sequences were identified at 106 nt (TTGACA, −35) and 83 nt (TATACT, −10) upstream of the start codon, which are characteristic for strong prokaryotic promoters. An inverted repeat of 34 nt was found at positions 57–90 downstream from the stop codon TAA of the ORF. This repeat forms the characteristic stem–loop of a terminating transcription signal, followed by an AT-rich region (Fig. 1a).

The entire S-layer protein gene sequence of the uranium mining waste pile isolate *B. sphaericus* JG-A12 (*slfB*) (accession no. AJ849549) indicated an ORF of 3714 bp, encoding a protein of 1238 aa. Two putative promoter sequences were identified at 106 nt (TTGACA, −35) and 83 nt (TATACT, −10) upstream of the start codon, which are characteristic for strong prokaryotic promoters. An inverted repeat of 34 nt was found at positions 57–90 downstream from the stop codon TAA of the ORF. This repeat forms the characteristic stem–loop of a terminating transcription signal, followed by an AT-rich region (Fig. 1a).

The region upstream of the *slfB* gene is identical to that of the *slfA* gene of strain 9602, comprising the same ribosome-binding site and the same putative promoters (Fig. 1a). Similarly, the first 630 nt of the two ORFs are almost identical, whereas the central regions of the genes from position 630 to position 2550 share a decreasing identity
from 90% to 70%, and after position 2550 the genes differ significantly.

The sizes of the mRNAs transcribed from the two S-layer protein genes were determined by Northern blotting using an S-layer-protein-gene-specific DIG-labelled PCR-generated DNA probe (primer pair V8/R30: see Methods). Transcripts of about 3900 bp were detected in both strains, which is in accordance with the sizes of the two above-described ORFs. The expression of the slfA and slfB genes was confirmed by RT-PCR using primer pairs specific to the upstream regions and the 5'-terminal sequences and by N-terminal sequencing of five (strain JG-A12) and nine (strain 9602) protein fragments resulting from a digestion of purified S-layer proteins by endoproteinase Glu-C (data not shown).
The S-layer protein genes of both strains are preceded by an ORF of 714 bp that shows an identity of 42% to the *Bacillus cereus* ATCC 14579 *N*-acetylglucosaminidylphosphoheptanuronic acid-β-N-acetylmuramidase gene (accession no. NP_835080), which is involved in the biosynthesis of teichoic acid linkage units in bacterial cell walls, and also 42% identity to the teichoic acid synthesis protein A (accession no. AAS09508) of *Lactococcus lactis* NCC 533. Interestingly, similar to the upstream promoter-carrying regions of the S-layer protein genes, the nucleotide sequences of the putative teichoic acid synthesis protein genes of the strains JGA12 and 9602 share 100% identity. To our knowledge this is the first report of an S-layer protein gene located downstream of a gene encoding an enzyme which is involved in cell wall synthesis. Whether this localization is connected to the regulation of cell wall synthesis has to be elucidated in future.

The expressed S-layer proteins SlfA and SlfB have identical signal peptides of 31 aa that are responsible for the protein secretion to the cell surface. The predicted site for the cleavage of the signal peptides is located between the alanine residues 31 and 32. The theoretical molecular masses of the mature proteins are 126 624 Da for SlfA and 126 254 Da for SlfB, which are in accordance with the masses previously determined by SDS-gel electrophoresis (Raff, 2002). The calculated pI values of the mature proteins are 5.2 (SlfA) and 5.1 (SlfB). Analyses of the hydrophobicity profiles of the functional proteins (Kyte & Doolittle, 1982) show in the central regions of the proteins – between residues 320–696 (SlfA) and 444–889 (SlfB) – a remarkably low content of hydrophobic amino acids.

The primary structures of the two functional S-layer proteins studied demonstrate, similar to many other S-layers (Bahl et al., 1997; Engelhardt & Peters, 1998; Ilk et al., 2002; Kuen et al., 1997; Mesnage et al., 2001), a high content of acidic amino acids (aspartate and glutamate) and of serine and threonine residues. The low content of histidine, tryptophan and methionine, as well as the absence of cysteine, is characteristic not only of the studied S-layer proteins but also of all other described S-layer proteins of *Bacillaceae* (Sára & Sleytr, 2000; Sidhu & Olsen, 1997). The N-terminal regions of SlfA and SlfB were found to be identical, containing three acidic S-layer homologous (SLH) domains, starting at positions 3, 68 and 135 after the end of the signal peptide. These SLH domains are larger than those of the other known S-layer proteins of *B. sphaericus* and their primary structures are completely different. They are probably involved in anchoring the S-layer subunits to the cell wall components as demonstrated for the SLH domains of other S-layer proteins (Engelhardt & Peters, 1998; Ilk et al., 1999; Sára & Sleytr, 2000).

The two proteins share a very high, slowly decreasing identity from 96% to about 60% in their central region up to positions 873 of SlfA and 869 of SlfB. In contrast, the following C-terminal parts of the proteins show no similarity between each other (Fig. 2). The amount of glutamic acid and aspartic acid residues, carrying carboxylic groups, and of serine and threonine residues, carrying hydroxyl groups, is strikingly high in these regions. However, the number of serine and threonine residues is significantly higher in SlfB than in SlfA. Stretches of these amino acids are found throughout the central and C-terminal parts of both

![Fig. 2. Comparison of the primary structures of the B. sphaericus S-layer proteins studied to date. The percentage amino acid identity between the known S-layer proteins of *B. sphaericus* strains is shown. NTD, N-terminal domain; CD, central domain; CTD, C-terminal domain; Pos., position; trunc., truncated; IS, putative IS element. Symmetry data are according to Raff (2002).](http://mic.sgmjournals.org)
proteins. However, they are differently organized and distributed in the two proteins. A remarkable feature is the occurrence of such accumulations in the variable regions of the more conserved central parts: e.g. an insert with the sequence GAVTTTSYT in SfB, an insert with the sequence TTVDKD in SfA, and the sequence DTTETDKFT in SfB.

Localization and analysis of a second S-layer-protein-like gene copy on large plasmids

The abrupt change to low identity of the S-layer protein genes of strains 9602 and JG-A12 (Fig. 1a) may indicate a horizontal gene transfer between different organisms and/or DNA rearrangements. Bearing in mind that plasmids play an important role in horizontal gene transfer and that they are widely distributed within the genus Bacillus, we studied the plasmid contents of B. sphaericus NCTC 9602 and JG-A12 and checked for the presence of S-layer protein gene copies on them. After electrophoretic separation of cell lysates of strains 9602 and JG-A12 on low-percentage agarose gels, large plasmids were visualized in both strains. Southern hybridization using an S-layer-protein-gene-specific probe generated by primer pair V8/R30 (see Methods) showed that a second S-layer gene copy is located on large plasmids in both strains (data not shown).

Both S-layer protein gene copies were analysed by PCR amplification of the different vectorette libraries using S-layer-protein-gene-specific primers and by cloning and sequencing of PCR products. The entire sequences of the plasmid-located S-layer protein-like gene copy slla (accession no. AJ849548) of strain 9602 and of the gene copy sllb (accession no. AJ849550) of strain JG-A12 indicate ORFs of 3297 bp and 3303 bp, encoding a protein of 1099 and 1101 aa, respectively. Thus, the two putative S-layer proteins should be significantly smaller than the functional proteins (Fig. 1b). Remarkably, the sequences of the two sll genes are almost identical. Furthermore, the 5’-sequence regions of the ORF of the two sll genes are identical to those of the sllf genes, whereas their 3’-terminal regions differ significantly from those of the functional genes (Fig. 2). Moreover, the regions of the ribosome-binding sites of the sll copies were found to be identical to those of the chromosomally encoded S-layer genes. A direct repeat is found at position 46–97 downstream from the stop codon TAA, forming a deduced stem–loop of 51 nt followed by an AT-rich region. Thus, the predicted transcripts should have a size of about 3500 nt, which is significantly smaller than those of the functional sllf genes. However, neither Northern blot analysis nor RT-PCR could prove expression of the S-layer-protein-like sll genes under the conditions described in this work. The exclusive synthesis of the Slf proteins was further confirmed by N-terminal sequencing of fragments of isolated S-layer proteins after digestion with the endoprotease Glu-C.

The N-terminal parts of the predicted silent proteins SlfA and SlfB are similar to those of the functional proteins SlfA and SlfB, possessing almost identical signal peptides with the exception of one amino acid in SlfA. The theoretical molecular masses of the mature putative S-layer proteins are 113 313 Da with a pI of 5.3 for SlfA, and 111 990 Da with a pI of 5.3 for SlfB. Both proteins show all typical features of S-layer proteins. The central regions of the putative proteins SlfA and SlfB are very similar to those of SlfA and especially SlfB. Interesting are several conserved regions shared by the putative proteins and SlfB. The almost identical C-terminal regions of SlfA and SlfB differ significantly from those of SlfA and SlfB and are shorter. Similar to the functional proteins, accumulations of serine, threonine, aspartate, and glutamate are found throughout the central and C-terminal parts, but with a different distribution and a significantly lower amount compared to SlfA and SlfB. However, since no expression of these proteins has yet been detected, their probable symmetries or interactions with metals remain unknown.

Identification and organization of a novel putative insertion element

The nucleotide sequences of the regions following the S-layer protein genes slla and sllb were determined. Following the highly divergent 3’-terminal parts of slla and sllb, large fragments of 7157 bp (strain 9602, accession number AJ866974) and 7139 bp (strain JG-A12, accession number AJ866975) were sequenced, and they showed a surprisingly high identity to each other. Sequence analyses predicted the presence of an ORF designated ORFA, followed by two divergently transcribed ORFs, designated ORFB and ORFC, and a sequence identical to the plasmid-encoded S-layer protein genes slla and sllb of strains JG-A12 and 9602 (Fig. 3). Analysis of the ORFA gene sequence showed a high

**Fig. 3.** Comparison of the organization of the putative IS elements ISBsph1 and ISBsph2, found on the chromosome of B. sphaericus NCTC 9602 and JG-A12, respectively, with IS657 of B. halodurans. Inverted repeats are shown in capitals and direct repeats are underlined.
identity of 68% with the IS605/IS200-like transposase of *Clostridium acetobutylicum* ATCC 824. In the case of strain 9602, the 6560 bp fragment comprising ORFA, ORFB, ORFC and an intact copy of sflA is flanked by a direct repeat of 2 bp (TT) and an imperfect inverted repeat of 14 bp. In the case of strain JG-A12, the 6313 bp fragment containing ORFA, ORFB, ORFC and a truncated sflB copy is flanked by a direct repeat of 7 bp (TTGTTGTG) and an imperfect inverted repeat of 8 bp (Fig. 3). On the basis of these results we suggest the presence of a new insertion element in each inverted repeat of 8 bp (Fig. 3). On the basis of these results we suggest the presence of a new insertion element in each inverted repeat of 8 bp (Fig. 3).

### Characterization of ORFA

ORFA, consisting of 450 bp, starts at position 317 downstream from the stop codon TAA of the S-layer protein gene sflB of strain JG-A12. It encodes a putative protein of 150 aa with a theoretical molecular mass of 17,716 Da and a calculated pI of 9.59. BLAST analysis of the primary structure of the protein showed 72% identity to an IS605/IS200-like transposase of *Clostridium acetobutylicum* ATCC 824, 71% to a transposase of *Clostridium thermocellum* ATCC 27405, 70% to an IS605/IS200-like transposase of *Staphylococcus epidermidis* ATCC 12228, and 67% to the IS657 transposase of the strain *Bacillus halodurans* C-125.

An almost identical ORF was identified in *B. sphaericus* NCTC 9602 starting 615 nt downstream from the stop codon of the S-layer protein gene sflA. Sequence comparisons of the primary structure of the predicted protein encoded by the ORF showed 97% identity to the putative protein encoded by the ORFA of strain JG-A12. Hence, the S-layer protein genes of both analysed strains are followed by a putative bacterial transposase-encoding gene.

In order to determine the number of putative transposase gene copies distributed within the genomes of the two strains, chromosomal DNA was digested using the restriction endonucleases EcoRV, Clal and SpeI. After electrophoretic separation, the fragments were transferred to a nylon membrane and hybridized with a DNA probe specific to the putative transposase gene. Southern blots of all digests gave one signal in each lane (Fig. 4), indicating the presence of only one gene copy on the chromosome of each strain. However, whether transposase genes are also associated with the sfl genes found on the plasmids remained unclear.

### Characterization of ORFB

Following the putative transposase gene of strain JG-A12, an ORF of 525 bp was identified at position 217 downstream from the stop codon TAA of ORFA, encoding a putative protein of 175 aa with a theoretical molecular mass of 19,902 Da and a calculated pI of 9.66. Interestingly, the predicted gene is oriented in the opposite direction to sflB and ORFA (Fig. 3). BLAST analysis of the protein sequence showed an identity of 32% to a putative integrase of *Vibrio vulnificus* YJ016 and of 31% to the integrase/recombinase

---

**Fig. 4.** Southern hybridization. Genomic DNA of *Bacillus sphaericus* JG-A12 (a) and NCTC 9602 (b) was digested using different restriction endonucleases. After electrophoretic separation and blotting, the fragments were hybridized using a DIG-labelled PCR probe specific to the putative recombinase or the putative transposase of JG-A12 generated by the primer pairs V83/R71 and Mobil_F/Mobil_R, respectively (see Methods for primer sequences). Weak signals are indicated by arrows.

---

XerD of *Vibrio parahaemolyticus* RIMD 2210633. NCBI conserved domain search (Marchler-Bauer et al., 2003) showed high similarity to the domain INT-REC-C of the tyrosine recombinase/integrase family. Similar results were found for *B. sphaericus* NCTC 9602. A predicted ORF of 525 bp in the opposite orientation (Fig. 3) showed an identity of 97% to the putative integrase/recombinase protein encoded by the ORFB of *B. sphaericus* JG-A12.

To analyse the distribution of the putative integrase genes within the genomes of strains 9602 and JG-A12, chromosomal DNA of the two strains was digested with the restriction endonucleases EcoRV, Clal and SpeI. Southern hybridization of all digests gave one signal in each lane with a size corresponding to the signals detected using the transposase-gene-specific probe (Fig. 4). However, in the case of the Clal digests, a second small fragment giving a weak signal was detected, indicating the presence of a second recombinase copy within the genomes of 9602 and JG-A12. It remains to be determined whether this putative gene copy
is located on the chromosomes or on the large plasmids harboured by the strains.

**Characterization of ORFC**

After the putative recombining of *B. sphaericus* JG-A12 an ORF of 849 bp was found at position 385 after the stop codon of ORFB, in the opposite orientation (Fig. 3), encoding a protein of 283 bp with a theoretical molecular mass of 33 419 Da and a calculated pI of 5-68. NCBI Conserved Domain Search of the entire protein sequence found a high identity to helix–turn–helix XRE family-like proteins (HTH_XRE), which are prokaryotic DNA-binding proteins belonging to the xenobiotic response element family of transcriptional regulators (Luscombe *et al.*, 2000; Wintjens & Rooman, 1996; Wood *et al.*, 1999). A similar organization was found for strain 9602: after the putative recombinase gene an ORF of 843 bp is found, encoding a protein of 281 aa with 90 % identity to that of the predicted HTH_XRE-like protein of JG-A12.

**Identification of an additional silent and truncated S-layer protein gene copy**

Surprisingly, following the ORF of the HTH-like protein in both strains a gene sequence was found in the opposite orientation with a high identity to the functional, chromosomally encoded S-layer protein genes *sllA* and *sllB*, and especially to the silent plasmid-located S-layer-protein-like genes *sllA* and *sllB* (Fig. 3). However, some sequence differences were found. The entire sequence of the inverted S-layer-protein-like gene of strain 9602 determined in this study indicates an ORF of 3297 bp, encoding a protein of 1099 aa with a theoretical molecular mass of 116 472 Da, showing 100 % identity to the plasmid-encoded *sllA*. The potential ribosome-binding site preceding the ORF at 13 nt upstream of the start codon is incomplete (TGAGG), suggesting that this gene is silent. Further, the upstream region shows no identity to that of the functional gene *sllA* and putative promotors are found at a far distance from the start codon at –452 (–35, TTGAAT) and –426 (–10, TATAGA).

The corresponding sequence of the inverted S-layer-protein-like gene of strain JG-A12 represents a truncated form of the plasmid-encoded *sllB* possessing a large deletion of 391 bp of the 5’-terminal region (Figs 3 and 5).

Interestingly, the upstream region of this copy shows a high identity to the upstream region of the chromosomally located *sllA* copy of strain 9602, although a sequence of 65 bp containing the modified ribosome-binding site is truncated (Fig. 5). Additionally, a large part of the upstream region identical to that in 9602 was found to be repeated (Fig. 5).

**Characterization of the region following the silent S-layer protein gene sequences**

Following the truncated *sllB* copy of JG-A12, the nucleotide sequence of a 1407 bp fragment was determined. BLAST analysis of the translated sequence showed a high identity of 50 % of the last 513 bp to the 5’-terminal region (starting at aa 896) of a putative extracellular nuclease of *Deinococcus radiodurans* R1 with a size of 1067 aa. These results indicate the presence of a truncated ORF, probably disrupted by the insertion of ISBspH2.

**Comparative analysis of *B. sphaericus* S-layer proteins**

S-layer proteins of several *B. sphaericus* strains have been described previously: *B. sphaericus* CCM 2177 (Ilk *et al.*, 2002), P-1 (Deblaere *et al.*, 1995) and WHO 2362 (Bowditch *et al.*, 1989; Deblaere *et al.*, 1995). These strains and those of the present study belong to different DNA homology groups (Miteva *et al.*, 1999). The mosquito-pathogenic strain WHO 2362 belongs to DNA homology group II and possesses, like other pathogenic strains of the genus *Bacillus* (Lewis *et al.*, 1987; Mesnage *et al.*, 2001; Sidhu & Olsen, 1997), an oblique structured S-layer (Raff, 2002; Sidhu & Olsen, 1997), which is believed to contribute to their pathogenity (Lewis *et al.*, 1987; Mesnage *et al.*, 2001). In contrast, the S-layers of the non-pathogenic strains *B. sphaericus* CCM 2177 and P-1, which belong to DNA homology group III, possess square symmetry (Ilk *et al.*, 2002; Lepault *et al.*, 1986; Ohnesorge *et al.*, 1992). The S-layers of *B. sphaericus* NCTC 9602 and JG-A12 have square symmetry as well (Raff, 2002), although these strains belong to a different DNA homology group (group I) (Selenska-Pobell *et al.*, 1999).

Sequence comparisons of the different *B. sphaericus* S-layer proteins demonstrate the unique structure of the S-layer proteins of *B. sphaericus* NCTC 9602 and JG-A12. The primary structures of the SLH domains studied in this work differ significantly from those described for the S-layer

![Fig. 5. Organization of the chromosomally encoded copies of sllA and sllB. mRBS, modified ribosome-binding site; SP, signal peptide; NTD, N-terminal domain; CTD, C-terminal domain; DR, direct repeat.](image-url)
proteins of *B. sphaericus* CCM 2177, *B. sphaericus* P-1 and *B. sphaericus* WHO 2362 (Fig. 2), whose N-terminal SLH domains share a surprisingly high identity (Fig. 2). The region following the N-terminal domains of the S-layer proteins of *B. sphaericus* NCTC 9602 and JG-A12 shares a striking high (61%) and slowly decreasing identity with the S-layer proteins of strains CCM 2177 and P-1, whereas no significant homology was found to the S-layer protein of strain 2362. The high identity of the N-terminal and central parts of SlfA, SlfB, SllA and SllB changes abruptly in their C-terminal parts (Fig. 2). Whereas the C-terminal part of SlfA shows no similarity to any known S-layer protein, the corresponding region of SlfB shares a surprisingly high identity of 57% (Fig. 2) with that of the S-layer protein of *B. sphaericus* CCM 2177. This high identity is very interesting, since the C-terminal regions of the Bacillus S-layer proteins studied to date are highly variable. Further, sequence comparisons of S-layer protein genes of even closely related bacterial strains demonstrated a high diversity of these proteins (Hansmeier et al., 2005; Raff & Selenska-Pobell, 2004). The almost identical structure of the C-terminal regions of the putative S-layer proteins SllA and SllB is therefore surprising.

**DISCUSSION**

**Primary structure characteristics of the S-layer proteins of *B. sphaericus* NCTC 9602 and JG-A12**

The amino acid compositions of the two functional S-layer proteins studied demonstrate features typical of many other S-layers (Bahl et al., 1997; Engelhardt & Peters, 1998; Ilk et al., 2002; Kuen et al., 1997; Mesnage et al., 2001). A remarkable feature is the stretches of aspartic and glutamic acid residues as well as serine and threonine residues, which were found especially in the C-terminal parts of SlfA and SlfB as well as of the putative proteins SllA and SllB, with different amounts and distribution. In this context, the results of previous work in which the interactions of the S-layers of *B. sphaericus* JG-A12 and NCTC 9602 with uranium were studied (Merroun et al., 2005; Raff et al., 2004) are interesting. The analyses showed that the S-layer SlfB of strain JG-A12 is much more effective in uranium binding than the S-layer of strain NCTC 9602 (Raff, 2002; Raff & Selenska-Pobell, 2004). EXAFS (extended X-ray absorption fine structure) demonstrated that the uranium bound by the S-layers of both strains is coordinated mainly to phosphate and carboxyl groups of the proteins (Merroun et al., 2005; Raff et al., 2003, 2004). Analyses of post-translational modifications confirmed the phosphorylation of the proteins (Raff, 2002; Raff & Selenska-Pobell, 2004). In addition, ICP-MS (inductively coupled plasma mass spectrometry) and colorimetric methods showed that the S-layer protein SlfB of the uranium mining waste pile isolate JG-A12 possesses six times more phosphorus than the S-layer protein SlfA of strain NCTC 9602. Bearing in mind that threonine and serine are the most frequently phosphorylated residues in prokaryotic proteins (Cozzone, 1988; Saier et al., 1990; Thomas & Trust, 1995), the higher level of phosphorylation of SlfB in comparison to SlfA may be related to the differences in the content and distribution of threonine and serine residues in the central and C-terminal parts of the S-layer proteins, and to the accumulations of serine and threonine stretches found in these regions. Since phosphate groups possess extremely high metal complexing activity, the high ability of strain JG-A12 to bind uranium and other metals (Selenska-Pobell et al., 1999) may be explained by the high content of phosphate residues of SlfB, but also by the stretches of residues carrying carboxyl groups found in the protein.

**Identification of novel IS elements in *B. sphaericus***

As well as the plasmid-encoded silent S-layer gene copies slfA and slfB an additional truncated gene copy was found downstream of slfA and slfB, showing the same sequence as slfA and slfB (Fig. 6). Interestingly, the truncated S-layer protein gene copies were found to be associated with genes...
encoding proteins involved in genetic rearrangements, indicating the presence of two novel IS elements, designated ISBpsh1 and ISBpsh2 (Fig. 6).

Although BLAST analyses of the putative transposase indicated the affiliation of ISBpsh1 and ISBpsh2 to the IS605/IS200-family, the IS elements show some significant differences from members of this family. IS200 contains only a single ORF encoding a transposase and does not carry terminal inverted repeats (Beuzon & Casadesus, 1997; Lam & Roth, 1983a, b). In the case of IS605, a second transposase is associated with a second divergently transcribed ORF, which is related to a putative transposase (Kersulyte et al., 1998, 2000; Murai et al., 1995). The putative transposase found in the present study is closely related to an ORF found in IS657 of the alkaliphilic strain Bacillus halodurans C-125 (Takami et al., 2001). IS657 contains only one ORF and no inverted repeats were found (Fig. 3). Thus, the insertion elements found in the present work are quite unusual. Their size and organization are more indicative of a transposon rather than an insertion element, although BLAST analysis showed no similarities to a known transposon and no similar gene element was found in any bacterial genome sequenced so far. To our knowledge, this is the first description of a mobile element encoding an S-layer-protein-like gene.

**S-layer protein genes of B. sphaericus JG-A12 and NCTC 9602 have probably evolved by horizontal gene transfer and rearrangements**

The sequence comparisons (Fig. 2) strongly suggest an intragenic ‘patchwork’ evolution of the functional S-layer genes in strains of *B. sphaericus*. It is obvious that the N-terminal parts of the S-layer proteins of *B. sphaericus* strains 9602 and JG-A12 have evolved from a common ancestor, which is different from that of the other *B. sphaericus* strains studied. Similarly, the high identity of the N-terminal domains of the S-layer proteins of *B. sphaericus* strains CCM 2177, WHO 2362 and P-1 (Fig. 2) indicates a common derivation of these domains, although the strains belong to different DNA groups and possess different symmetries. On the other hand, the sequence comparisons demonstrate that the central parts of the S-layer proteins SlfA and SlfB and those of strains CCM 2177 and P-1 share a common origin, whereas the S-layer protein of strain WHO 2362 shows a unique structure. Further, the results indicate that the C-terminal parts of the S-layer proteins of strains JG-A12 and CCM 2177 have evolved from a common ancestor, whereas the C-terminal parts of strains 9602, P-1, and WHO 2362 have evolved from three different, unrelated ancestors.

The presence of the IS elements, consisting of a truncated *sll* copy, a putative transposase and a putative recombinase (Fig. 6), supports the above suggestion. Whereas transposases are needed for efficient transposition of the insertion sequence, functions of recombinases include the site-specific insertion of the mobile element and, most interesting in the present study, DNA inversions controlling expression of surface proteins (Sadowski, 1986; Stark et al., 1992).

Therefore, we may speculate that the insertion elements found were involved in the evolution of the S-layer protein genes in both strains studied. Parts of the S-layer protein genes may have been exchanged between the different gene copies by integrase/recombinase-mediated chromosomal rearrangements.

Shuffling of functional protein domains as well as horizontal gene transfer are well known and play an important role in microbial evolution. These processes generate proteins of a great variety, with similar functions but different specificities. Prominent examples of protein families produced by domain shuffling and horizontal gene transfer are the insecticidal crystal proteins of *Bacillus thuringiensis* (Bravo, 1997). Horizontal transfer and rearrangements of segments of S-layer protein genes have also been reported. It is assumed that the bacteria use S-layer variation for adaptation to different stress factors (Kuen et al., 1997; Jakava-Viljanen et al., 2002) or for their virulence (Blaser et al., 1994; Dworkin & Blaser, 1996; Mesnage et al., 2001; Mignot et al., 2003). Mechanisms of S-layer protein variation include the inversion of promoters, recombination of S-layer gene segments (Dworkin & Blaser, 1996), and the replacement of an active S-layer protein gene by a formerly silent gene (Boot et al., 1996b). *G. stearothermophilus* PV72 expresses two different S-layer protein gene variants (sbsA and sbsB) under different growth conditions. Whereas sbsA is located on the bacterial chromosome, sbsB is carried on a large plasmid harboured by the strain. For the expression of sbsB, the gene is translocated from the plasmid to a specific chromosomal expression site, whereas the coding region of sbsA is irreversibly transferred to the plasmid (Scholz et al., 2001). Due to the lack of significant sequence identity of the sbsA and sbsB genes, a replacement of the two genes by homologous recombination was excluded (Scholz et al., 2001).

In contrast, the chromosomally located (slfA and slfB) and the plasmid-located (sllA and sIlB) genes of *B. sphaericus* NCTC 9602 and JG-A12 share large regions with an extremely high identity. Moreover, the two plasmid-located genes sllA and sIlB are almost identical. To our knowledge, NCTC 9602 and JG-A12 are the first strains of *B. sphaericus* found to contain large plasmids encoding silent S-layer protein genes. However, to date we have not been able to induce the expression of any of the sl genes described in this work. Rearrangements of S-layer protein genes mediated by insertion elements have not been reported so far. In the case of *Aeromonas salmonicida* (Gustafson et al., 1994), *G. stearothermophilus* PV72 (Scholz et al., 2000) and *G. stearothermophilus* ATCC 12980 (Engelbrecht et al., 2000), transposable elements have been described which affect the expression of S-layer proteins by insertion. However, these elements were much smaller than those described in the present study, comprising only a transposase and not being associated with an S-layer protein gene. For *B. sphaericus* NCTC 9602, the occurrence of a spontaneous variant in a frozen stock culture was reported in an earlier study, expressing a
15–20 kDa smaller variant of the S-layer protein (Hastie & Brinton, 1979a, b). Both variants, the wild-type and the smaller variant, as well as several truncated forms were not able to bind to the sacculi of the S-layer-possessing related strain *B. sphaericus* P-1 (Hastie & Brinton, 1979b). From our sequence analyses this finding can be explained by the different N-termini of the wild-type S-layer protein of NCTC 9602 and possibly also of the smaller variant in comparison to the S-layer protein of P-1 (Deblaere et al., 1995). It can be suggested that the spontaneous mutation was caused by the replacement of slfA with slfA possibly mediated by ISBspht1. However, a transposition or inversion of ISBspht1 or ISBspht2 has not yet been observed, and if and under what conditions the IS elements are mobilized remains to be determined.

**ACKNOWLEDGEMENTS**

This study was supported by grant DFG/SE 671/7-2 from DFG, Bonn, Germany, by grant 7531.50-03.0370-01/5 from SMWK, Dresden, Germany, and by EC grant GRDI-2001-40424. We thank R. Getzlaff, I. Plumeier (GBF Braunschweig) and K. Flemming for technical assistance.

**REFERENCES**


