JOURNAL OF BACTERIOLOGY, Mar. 1994, p. 1224–1233
0021-9193/94/$04.00+0

Vol. 176, No. 5

# Domain Structure of the *Acetogenium kivui* Surface Layer Revealed by Electron Crystallography and Sequence Analysis

ANDREI LUPAS, HARALD ENGELHARDT, JÜRGEN PETERS, UTE SANTARIUS, SUSANNE VOLKER, AND WOLFGANG BAUMEISTER*

*Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany*

Received 23 September 1993/Accepted 13 December 1993

**The three-dimensional structure of the *Acetogenium kivui* surface layer (S-layer) has been determined to a resolution of 1.7 nm by electron crystallographic techniques. Two independent reconstructions were made from layers negatively stained with uranyl acetate and Na-phosphotungstate. The S-layer has p6 symmetry with a center-to-center spacing of approximately 19 nm. Within the layer, six monomers combine to form a ring-shaped core surrounded by a fenestrated rim and six spokes that point towards the axis of threefold symmetry and provide lateral connectivity to other hexamers in the layer. The structure of the *A. kivui* S-layer protein is very similar to that of the *Bacillus brevis* middle wall protein, with which it shares an N-terminal domain of homology. This domain is found in several other extracellular proteins, including the S-layer proteins from *Bacillus sphaericus* and *Thermus thermophilus*, Ompα from *Thermotoga maritima*, an alkaline cellulase from *Bacillus* strain KSM-635, and xylanases from *Clostridium thermocellum* and *Thermoanaerobacter saccharolyticum*, and may serve to anchor these proteins to the peptidoglycan. To our knowledge, this is the first example of a domain conserved in several S-layer proteins.**

*Acetogenium kivui* (19) is a hydrogen-oxidizing, acetogenic bacterium (18) that is moderately thermophilic and grows optimally at 66°C. In spite of its gram-negative staining behavior, its cell wall has gram-positive characteristics. Like many other bacteria, gram positive as well as gram negative, it is covered by a regularly arrayed surface layer (S-layer). This layer has a hexagonal structure and consists of a single 80-kDa protein whose gene has been cloned and sequenced (21). The S-layer protein is modified at four tyrosine residues by long glycan chains that are composed of glucose, galactosamine, and an as-yet-unidentified sugar-related component (22). Since these tyrosines are all preceded by valine, Val-Tyr may represent a recognition sequence for glycosylation in bacteria. The 200 N-terminal residues of the S-layer protein are 31% identical to the 200 N-terminal residues of the *Bacillus brevis* 47 middle wall protein. Within these, residues 29 to 85 and 94 to 150 represent a degenerate internal repeat (21). Recently, Fujino et al. (13) have sequenced three open reading frames 3′ of the *cipA* gene of *Clostridium thermocellum*. Each of these contains a C-terminal domain that is significantly homologous to the internal repeat region of the *A. kivui* and *B. brevis* proteins. Sequence comparisons indicate that one of these open reading frames may code for a structural component of the cellulosome.

A preliminary two-dimensional (2-D) analysis by electron microscopy in conjunction with image processing has shown that the *A. kivui* layer is composed of hexameric units creating a "toothed-wheel" motif (24). In this communication, we present two three-dimensional (3-D) reconstructions of the *A. kivui* S-layer, obtained by means of electron crystallography, and attempt to correlate the domain structure seen in the reconstructions with results obtained by proteolysis and sequence analysis. In addition, we show that the repeated sequence at the N terminus of the S-layer protein (named SLH for S-layer homology) is conserved in several other proteins and discuss possible implications for its function.

* Corresponding author. Mailing address: Molekulare Strukturbiologie, Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany. Phone: (089) 8578-2652. Fax: (089) 8578-2641.

## MATERIALS AND METHODS

**Bacterial strain and growth conditions.** *A. kivui* was obtained from the German collection of Microorganisms (DSM 2030), Braunschweig, Germany. Cells were grown anaerobically in the medium described by Leigh et al. (18), buffered with 50 mM phosphate (pH 6.5) and supplemented with yeast extract (2.0 g per liter), tryptone (2.0 g per liter), and glucose (5.0 g per liter). The growth temperature was between 60 and 64°C.

**S-layer preparation.** Cells were harvested in the logarithmic growth phase by centrifugation at 4,500 × *g* and washed once in distilled water. The peptidoglycan was digested by adding 10 to 20 mg of lysozyme to 100-ml aliquots of cell suspension and incubating the mixture for 6 to 8 h at room temperature. The suspension was then centrifuged at 39,200 × *g*, and the upper, white part of the pellet was resuspended in Tris-HCl buffer (25 mM, pH 7.5) containing 5 mM $MgCl_2$ while the lower, darker part of the pellet was discarded. This procedure was repeated four to five times. The isolated S-layer sheets were stored in the same buffer at 4°C.

**Specimen preparation and electron microscopy.** Isolated S-layer sheets were adsorbed to glow-discharged carbon-coated copper grids, rinsed with droplets of distilled water, and negatively stained with either 2% uranyl acetate (UA) (pH 4.0) or 1.5% Na-phosphotungstate (PTA) (pH 6.8). Electron microscopy was performed with a Philips EM 420 electron microscope equipped with an unlimited-tilt specimen holder (4), at an operating voltage of 80 kV and a magnification of ×36,000. Continuous-tilt series spanning the nominal range from 0 to 78° with increments in proportion of cos (Ψ) were recorded (29).

**Image processing.** Micrographs were examined in an optical diffractometer to select the single-layered areas with the best crystalline order and to monitor astigmatism and defocus. Images were subjected to densitometry with an Eikonix model

1412 camera system (Eikonix Corporation) as arrays of 1,024 × 1,024 pixels and with a pixel size of 15 μm, corresponding to 0.417 nm at the specimen level. Image processing was done with the SEMPER system (30). Projections were subjected to correlation averaging (27), and the 3-D reconstruction was accomplished by the hybrid Fourier space/real-space approach (29), starting with unit cells extracted from correlation averages of the tilt series. Three-dimensional models were generated by appropriate thresholding by means of "surface shading" of solid representations (26).

**Two-dimensional polyacrylamide gel electrophoresis (PAGE).** S-layer partially digested with proteinase K (22) was separated on Servalyt-precotes (Serva, Heidelberg, Federal Republic of Germany), pH range 3 to 6, in the first dimension and on 13% polyacrylamide gels in the second dimension.

**FTIR spectroscopy.** Spectroscopic analysis of the intact S-layer protein was performed by means of attenuated-total-reflection Fourier transform infrared (FTIR) spectroscopy. The spectra were recorded with a Nicolet 740 FTIR spectrometer. About 50 μg of the protein, dissolved in deionized water, was dried on a germanium crystal as previously described in detail (11). A total of 256 scans were accumulated at a resolution of better than 2 $cm^{-1}$. The amide I and II bands between 1,700 and 1,480 $cm^{-1}$ were subjected to band shape analysis and quantitative secondary-structure analysis as described by Kleffel et al. (17).

**Sequence analysis.** Secondary-structure predictions by the methods of Garnier et al. (14) and Chou and Fasman (5) were made by using the GCG package (6). Predictions by the method of Rost and Sander (25) were obtained by electronic mail (PredictProtein@EMBL-Heidelberg.de). For the sequences in Fig. 7, predictions were gathered for all the sequences and then each position of the alignment was assigned the most frequently predicted structure.

The alignment in Fig. 7 was generated by successive BLAST (1) searches of a nonredundant cumulative data base containing GenPept, Pir, and Swissprot. The searches were performed by electronic mail (blast@ncbi.nlm.nih.gov). The SLH sequence of *Thermotoga maritima* Ompα was used as a starting point because it occurs in single copy and is delimited clearly by the end of the signal sequence and the start of the coiled-coil rod domain (9). In each round, sequences whose similarity to a protein in the alignment was characterized by a smallest Poisson probability of less than $10^{-3}$ were added to the alignment and were used as starting points for the next round of searches. After a core alignment of 15 sequences had been reached and highly conserved positions had become apparent, sequences were inspected by eye for potential flanking partial repeats. The open reading frame protein from *Synechocystis* strain PCC6803 did not satisfy the cutoff Poisson probability by an order of magnitude because of two central insertions. Its sequence was added to the alignment after its membership in the family had been confirmed by profile searches. After the final alignment had been obtained, sequence profiles (15) were generated both for the complete alignment and for the 15 most similar sequences, and the Pir and Swissprot data bases were searched for further members of the family by using GCG programs. No other members were identified.

Searches for coiled-coil segments were performed by using the programs of Lupas et al. (20).

## RESULTS

**Structure of the S-layer.** The *A. kivui* S-layer is bound only weakly to the underlying peptidoglycan layer, and large sheets of it can be released by treating the bacteria with distilled water. Isolated S-layers have a tendency to self-associate, forming double layers with the two lattices usually in register. Such double layers are notoriously difficult to embed completely in negative stain. This may not cause problems for the 3-D reconstruction itself, provided that the two lattices are either perfectly in register (7) or well out of register (3, 23, 31), but it may cause serious problems in interpreting the reconstruction because of difficulties in defining the layer boundaries. Therefore we used only single-layered patches for 3-D reconstructions.

Since different negative stains can generate somewhat different representations of a protein structure (32), we used two negative stains in this study: the cationic UA (pH 4.0) and the anionic PTA (pH 6.8). In fact, there is a small but significant difference between the lattice constants of the *A. kivui* layer obtained with the two stains used. With UA we obtained a lattice constant of 19.6 ± 0.2 nm, and with PTA we obtained a lattice constant of 18.8 ± 0.1 nm. We therefore made 3-D reconstructions from both preparations.

The tilt series chosen for processing comprised 14 projections each. The actual tilt angles ranged from −0.3 to 78.3° (UA) and from 2.3 to 80.9° (PTA). No significant radiation damage was accumulated while recording the tilt series, as the power spectra of nominal 0° tilts at the beginning and end of the series showed no loss of information. Due to the high tilts available, resolution was nearly isotropic; it was marginally but consistently better in the case of UA-stained preparations (approximately 1.7 nm) than it was with PTA-stained preparations (approximately 2.0 nm), as estimated by the radial correlation function criterion (27). The agreement of the 3-D reconstructions with the projection data was examined by comparing averages of the untilted projection images with the corresponding projection through the 3-D reconstruction (12); there were no apparent differences.

Although the final 3-D reconstructions turned out to be quite similar, the diffractograms and the corresponding 2-D projection maps (Fig. 1) appear quite different. In the PTA-embedded preparations, the projection structure is dominated by the core and the connecting elements of the layer are rather poorly defined. In the projection maps from UA-stained preparations, the hexameric core is also the most prominent stain-excluding feature; however, here the network of structural elements providing lateral connectivity within the layer is clearly visible. In 2-D projections, the local variation in negative-stain depth (i.e., the variation in mass thickness) determines the contrast and hence the appearance of a given structure. In 3-D reconstructions, the depth of the stain is not as critical; it is sufficient that a given feature is outlined by a thin layer of stain, allowing a stain boundary to be traced. Also, differences in the adsorption to the underlying carbon film may have contributed to different stain distributions in the two S-layer preparations.

Sections through the reconstructed UA-stained layer that run parallel to the layer plane are shown in Fig. 2, proceeding from the inner to the outer surface. The reconstructed UA- and PTA-stained layers are compared in Fig. 3, as viewed from the inside (A and C) and from the outside (B and D). The sidedness is inferred from freeze-etching data (not shown). As with all eubacterial S-layers investigated so far, the core protrudes inwards; if one chooses to describe it as funnel shaped, it is invariably the narrower end of the funnel which faces the peptidoglycan (see also references 10 and 31). Apart from the small though noticeable difference in resolution, the basic structure is virtually the same in both reconstructions. By the classification scheme proposed by Saxton and Baumeister

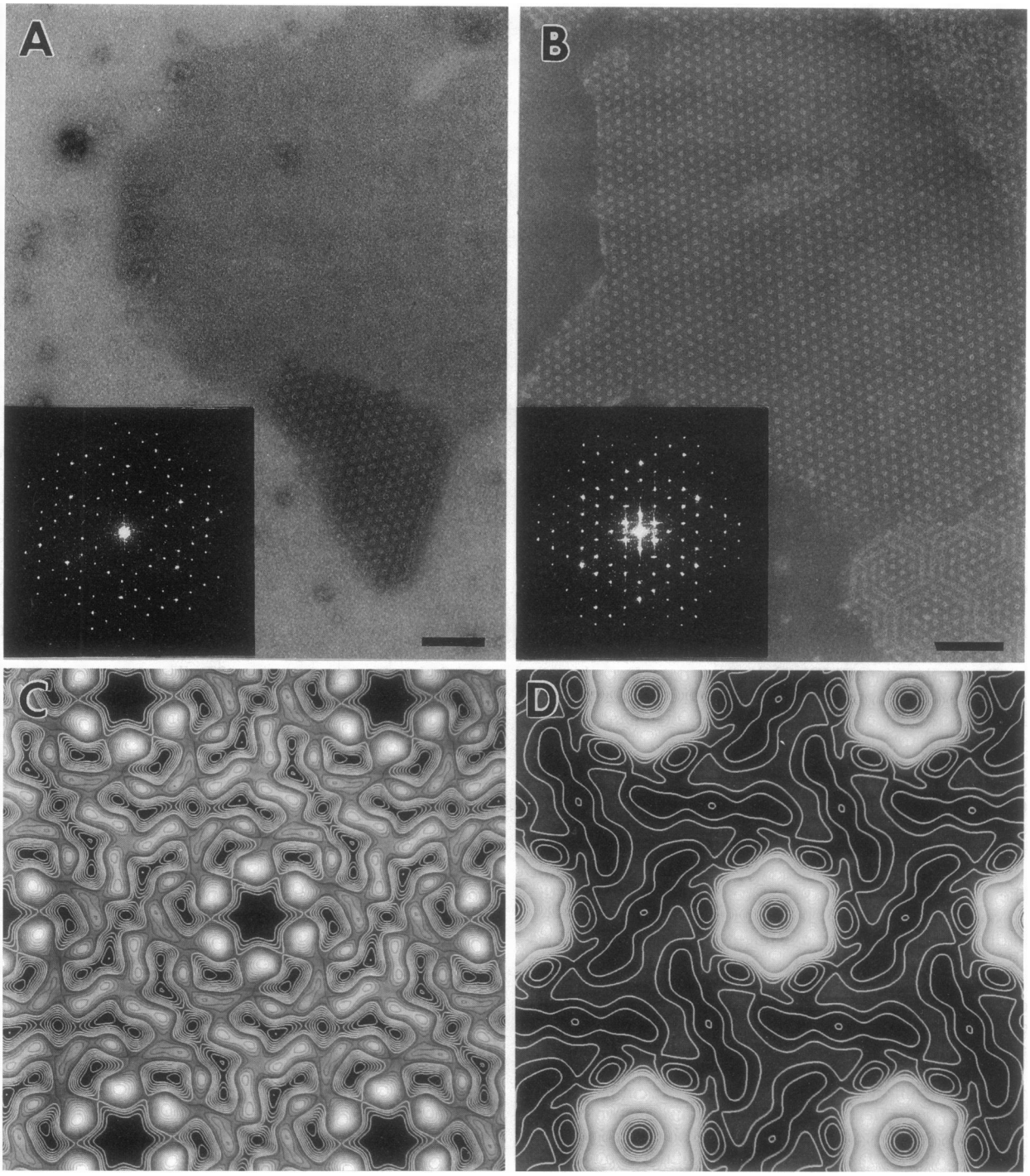


FIG. 1. Fragments of the *A. kivui* S-layer negatively stained with UA (A) and PTA (B). Panels C and D are 0° projections through the 3-D reconstructions and represent the average unit cell structure of the crystals depicted in panels A and B. The center-to-center spacings are 19.6 (A) and 18.8 (B) nm. The insets show the diffractograms (power spectra) of the S-layer patches used for the 3-D reconstructions. Bars, 100 nm.

(28), the *A. kivui* layer would be designated an $M_6C_3$-type layer, indicating that it has a hexameric core centered on the sixfold axis, with connectivity provided by other structural elements joining near the threefold axis of symmetry.

In both reconstructions, monomer and domain boundaries can be traced without difficulty throughout most of the structure (Fig. 2 and 4). Such clarity is unusual at this level of resolution, delineation of monomer shape and domains generally being rather arbitrary at 2 nm. Each monomer contributes a globular domain of about 3.5 nm in diameter (the C-domain) to the core; this forms the innermost part of the S-layer and connects to the peptidoglycan. From each globular domain, a Y-shaped structure (Y-domain) with unequal arms radiates out and towards the surface. Together, six Y-domains

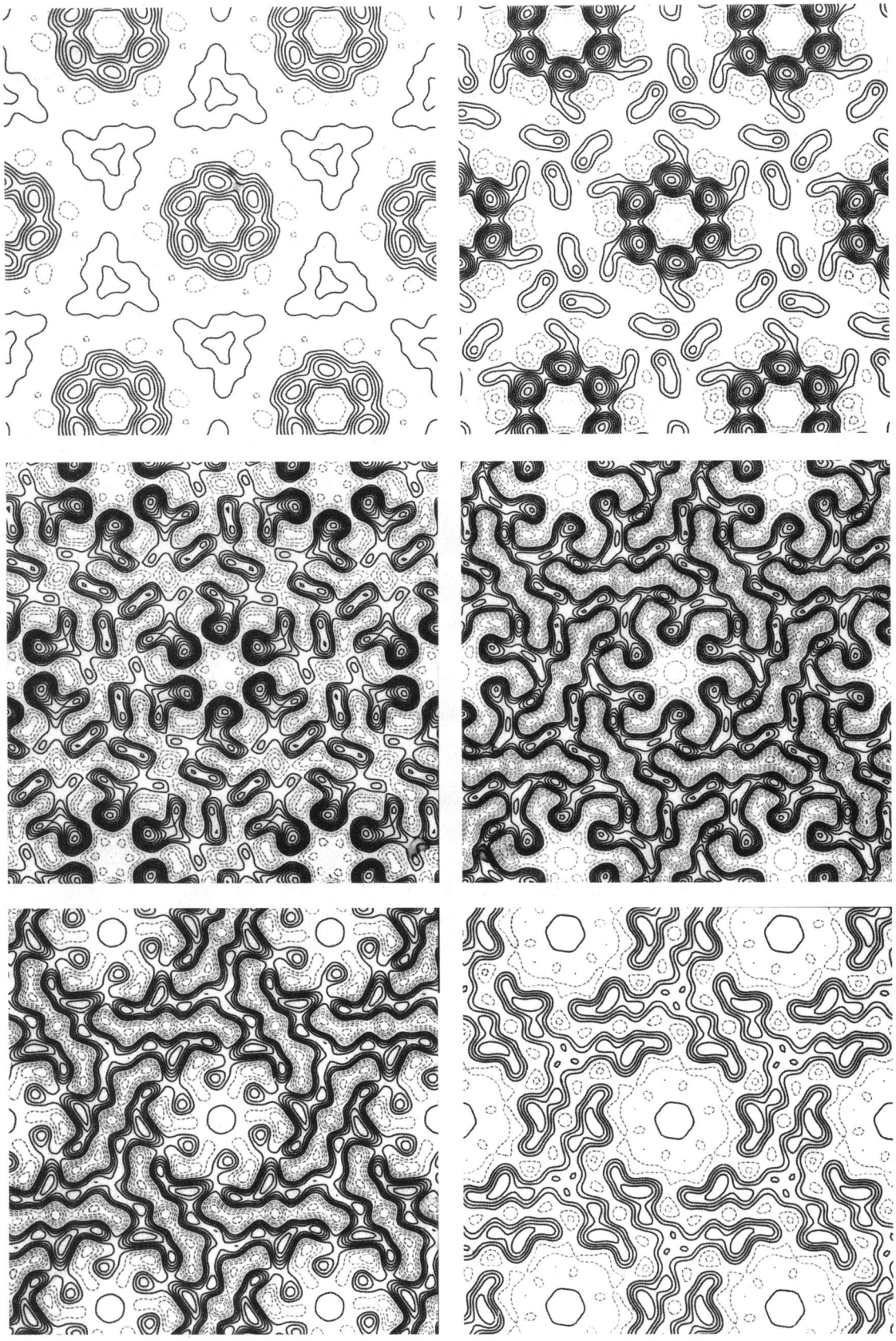

FIG. 2. Horizontal sections of the 3-D reconstruction from the UA-stained S-layer, viewed from the inner surface (top left) to the outer surface (bottom right). The sections have a spacing of 0.45 nm. The sampling of the sections is 0.15 nm in each direction.

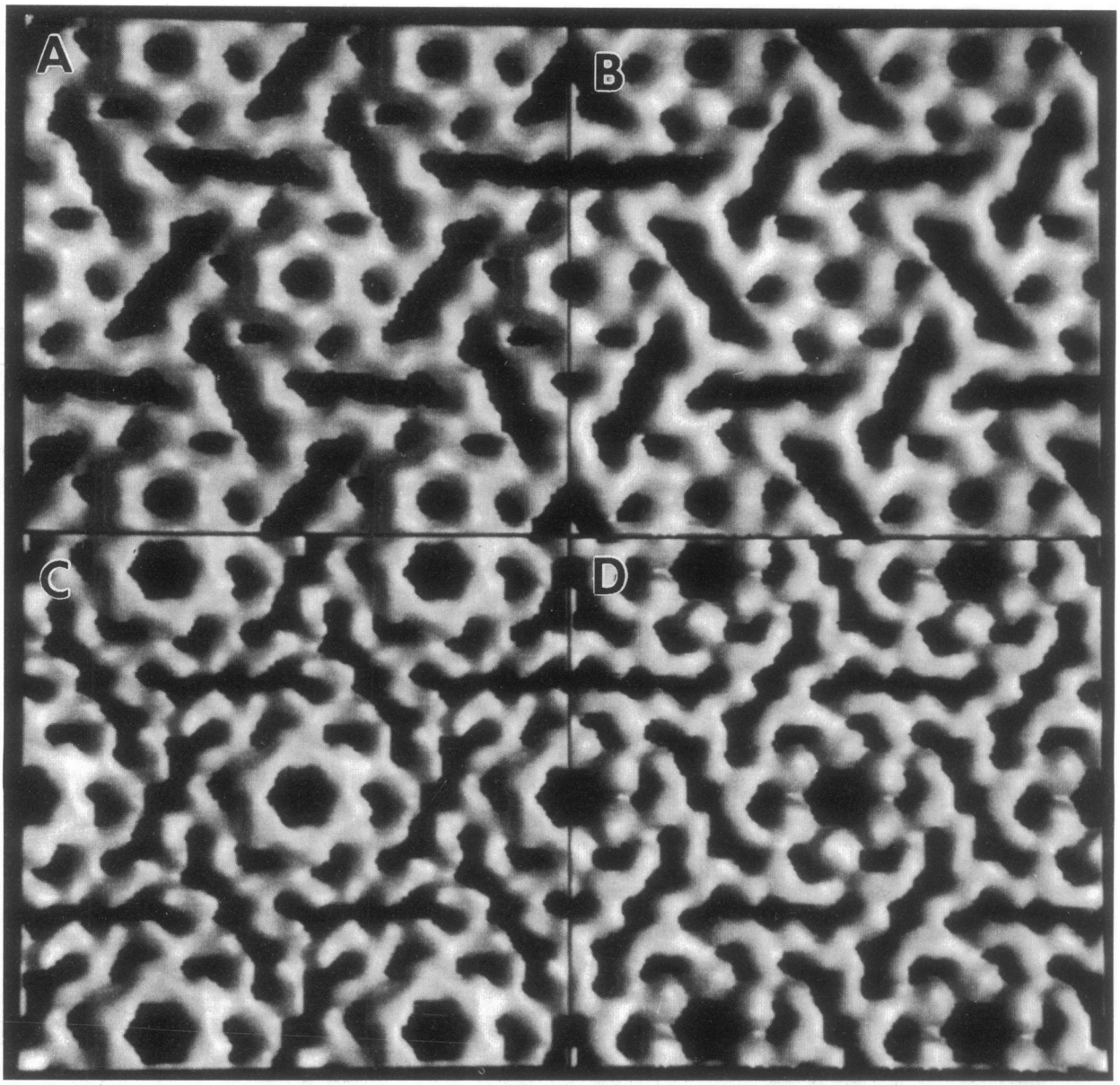


FIG. 3. Surface-shading representations of the 3-D reconstructions obtained for the PTA-stained S-layer (A and B) and the UA-stained S-layer (C and D). Panels A and C show the inner surface, and panels B and D show the outer surface.

form an oblique fenestrated rim of 12 nm in diameter and give the unit cell a funnel-shaped appearance. From the longer of the two arms of the Y-domain, a third, roughly L-shaped domain (L-domain) points towards the axis of threefold symmetry, where it connects to two L-domains from other unit cells. The L-domain is about 6 nm long and 3 nm in diameter.

The molecular volume of the monomer (as outlined in Fig. 4) is about 60 $nm^3$ for the UA preparation and about 76 $nm^3$ for the PTA preparation. These volumes are in fact only 62 and 78%, respectively, of the volume an 80-kDa protein can be expected to occupy, assuming a density of 1.37 $g/cm^3$. Volumes substantially lower than expected are commonly encountered in S-layer reconstructions; they are a consequence of setting restrictive thresholds for the sake of greater clarity in the view representations. Assuming that the thresholding affects all three domains equally, their fractions of the monomer volume can be calculated (Table 1). The domain boundaries shown in Fig. 4 were applied to all the horizontal sections of the 3-D reconstructions. Combining the molecular volumes of the well-preserved domains of the two reconstructions, we obtain molecular masses of 39 kDa for the core, 25 kDa for the Y-domain, and 16 kDa for the L-domain (Table 1).

**Secondary structure of the S-layer protein.** We determined the secondary-structure composition of the *A. kivui* S-layer protein by attenuated total reflection infrared spectroscopy (Fig. 5), using quantitative band shape analysis of the infrared spectra in the amide I and II regions between 1,700 and 1,480 $cm^{-1}$ (17). A comparison with the secondary-structure content predicted by the methods of Chou and Fasman (5), Garnier et al. (14), and Rost and Sander (25) is shown in Table 2. The value of about 40% β-sheet structure obtained is not unusual for S-layer proteins (2, 11). The α-helix content is lower and

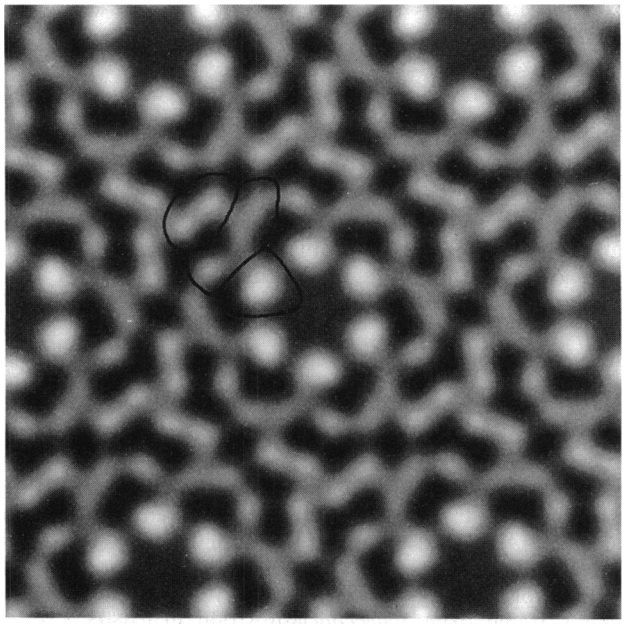

FIG. 4. Projection through the 3-D reconstruction of the UA-stained S-layer. The circled regions indicate the morphological domains. These masks were applied to horizontal sections for which thresholds had been set in order to obtain estimates of the volumes of the domains.

probably does not exceed 20 to 25% (Table 2). The band assigned to β-sheet absorptions is centered at 1,640.5 $cm^{-1}$; using the relationship between peak position and β-strand length given by Kleffel et al. (17), we calculate an average β-strand length of 5.8 amino acids. A similar value (6.7 ± 4.0) is obtained from the secondary-structure predictions. The amide I peak at 1,643 $cm^{-1}$, the basic absorption pattern, and the spectrum curve shape are very similar to the RNase A spectrum; this protein contains 27% α and 47% β structure, with an average β-strand length of 6.4 residues.

Although secondary-structure prediction for an individual sequence is too inaccurate to allow a detailed interpretation, three domains are identified that may correspond to structural domains of the protein: a mostly α-helical domain between residues 27 and 217, a domain composed mainly of β-strands between residues 218 and 480, and a domain of mixed α and β structure with a preponderence of β structure from residue 481 to the C terminus of the protein.

TABLE 1. Molecular masses of the morphological domains of the *A. kivui* S-layer protein derived from 3-D reconstructions

| Domain | Vol in 3-D model ($nm^3$) | | Relative vol (%) | Molecular mass estimate[a] (kDa) |
|---|---|---|---|---|
| | PTA stained | UA stained | | |
| L | (14)[b] | 16 | 20 | 16 |
| Y | 24 | 24 | 31 | 25 |
| C | 38 | (19)[b] | 49 | 39 |
| Total | 78 | | 100 | 80 |

[a] It is assumed that molecular mass is proportional to the volume occupied.
[b] Volume estimates in parentheses derive from incompletely stained domains and were not used to calculate the relative volume.

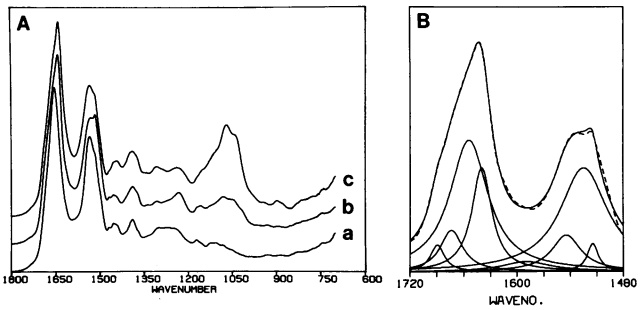


FIG. 5. (A) Infrared spectra of the *A. kivui* S-layer protein (b) together with those of the reference proteins bovine serum albumin (a) and bovine acid α1-glycoprotein (c). The absorption band centered at 1,050 $cm^{-1}$ derives from carbohydrates bound to the protein. (B) Amide I and II region of the *A. kivui* S-layer protein (dashed line). The band centered at 1,643 $cm^{-1}$ is assigned to β-strands, and the band at 1,655 $cm^{-1}$ is assigned to α-helices. These were used to estimate the relative secondary structure content.

**Glycosylated domains of the S-layer protein.** The infrared spectrum shows a significant absorption between 1,150 and 1,000 $cm^{-1}$ that originates from the carbohydrates bound to the S-layer protein (21, 22). By comparison with other glycoproteins, we estimate the carbohydrate content to be 8% of the protein weight in this particular preparation (Fig. 5), corresponding to 40 to 50 sugar molecules per protein monomer.

After digestion of the S-layer with proteinase K at 50°C, a high-molecular-weight fraction that contains glycans composed of glucose, galactosamine, and an unidentified sugar in equimolar ratio can be isolated. Further proteolysis yields peptides that are glycosylated at Tyr-297, Tyr-516, Tyr-520, and Tyr-632 (22). Analysis of the high-molecular-weight fraction by 2-D PAGE yields two diagonal patterns of spots (Fig. 6) whose pI value decreases from 4.5 to below 3 with increasing molecular weight. If the fraction is deglycosylated with trifluoromethanesulfonic acid prior to electrophoresis, only two bands of 33 and 15 kDa are observed. After electrophoresis and blotting, we determined the N-terminal amino acid sequences of the two polypeptides: the 33-kDa fragment starts at Lys-215, and the 15-kDa fragment starts at Ala-480. We conclude that the diagonal patterns seen in 2-D electrophoresis are formed by two protease-resistant domains of the S-layer protein that are modified by acidic repetitive unit glycans of heterogenous length. The 33-kDa fragment extends approximately to residue 475 and contains the glycosylation site at Tyr-297, while the 15-kDa fragment extends approximately to residue 634 and contains the three glycosylation sites at Tyr-516, Tyr-520, and Tyr-632. The 15-kDa fragment may have its C-terminal boundary determined by the glycosylation site at

TABLE 2. Secondary-structure composition of the *A. kivui* S-layer protein

| Structural element | Content (%) predicted by[a]: | | | Content determined by FTIR spectroscopy (SD) (%) |
|---|---|---|---|---|
| | RS | CF | GOR | |
| α-Helix | 26 | 29 | 23 | 20 (±10) |
| β-Sheet | 35 | 39 | 42 | 45 (±5) |
| Other | 39 | 32 | 35 | 35[b] |

[a] RS, Rost and Sander (25); CF, Chou and Fasman (5); GOR, Garnier et al. (14).
[b] This value is an estimate based on the values for the α-helix and β-sheet; it was not determined experimentally.

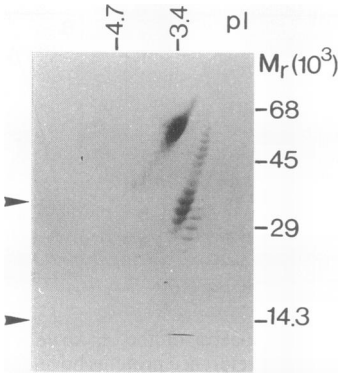


FIG. 6. Two-dimensional electrophoretic separation of a proteinase K-resistant fraction of the S-layer. The fraction was separated by isoelectric focussing in the first dimension (horizontal) and sodium dodecyl sulfate-PAGE in the second dimension (vertical). The pI values indicated are those obtained with Serva test mixture 9. The arrows mark the positions of the two bands that are obtained if the fraction is deglycosylated prior to electrophoresis.

Tyr-632, since the glycan may have prevented proteinase K from further degrading this domain. In a small fraction of the 15-kDa fragments, though, degradation may have proceeded beyond Tyr-632, since a minor diagonal pattern can be detected below the 15-kDa pattern in 2-D electrophoresis.

## DISCUSSION

**Structural similarity to other S-layer proteins.** The S-layer protein of *A. kivui* forms a lattice with p6 symmetry and a center-to-center distance of approximately 19 nm. The hexameric unit of the S-layer is a funnel-shaped structure with an inner ring-shaped core surrounded by an oblique fenestrated rim and six L-shaped spokes that provide lateral connectivity to other unit cells in the lattice. The putative monomer shows an unusually clear subdivision into domains.

The same structural organization is seen in the *B. brevis* 47 middle wall protein (31). This protein forms a hexagonal $M_6C_3$ lattice with a center-to-center distance of approximately 18.3 nm. As in the *A. kivui* S-layer protein, the core of the unit cell is ring shaped and contains a pore of about 2.5 nm in diameter. The monomers show a separation in C-, Y-, and L-domains, although the Y- and L-domains are considerably bulkier and the domain boundaries are less clear. Despite these minor differences, the overall similarity is striking and quite unexpected given that the two proteins do not have any significant sequence homology outside their N termini. They do, however, have the same domain structure in secondary-structure predictions: a mostly α-helical domain of 200 residues followed by a 300-residue domain composed mainly of β-strands and a C-terminal domain of mixed α and β structure that is 230 residues longer in the *B. brevis* protein than in the *A. kivui* protein and may account for the bulkier shape of the Y- and L-domains in this protein. It is noteworthy in this context that the similarity between the *B. brevis* 47 middle wall protein and its *B. brevis* HPD31 homolog follows the same pattern as the secondary-structure predictions. When calculating the average similarity of the two proteins with a scanning window of 100 residues, three domains with different levels of sequence identity are resolved fairly clearly: an N-terminal domain of about 180 residues (93% identity), a central domain of about 390 residues (71% identity), and a C-terminal domain of about 500 residues (46% identity).

Of the 10 S-layers from gram-positive species whose 3-D structure is known, one other, that of *Clostridium thermohydrosulfuricum*, belongs to the $M_6C_3$ class. This layer has a center-to-center distance of 16 nm and is formed by rod-shaped monomers that do not show an obvious separation into domains at 2.0-nm resolution (3).

**Sequence similarity to other proteins.** In the *A. kivui* S-layer protein, residues 30 to 94 are 32% identical to residues 95 to 151 and represent a sequence that was probably duplicated during the evolution of the protein. A third, partial copy of the sequence may exist at residues 152 to 168 (Fig. 7). This sequence is also found in several copies in the *B. brevis* 47 middle wall protein (21), in the *B. brevis* HPD31 hexagonal wall protein (8), and in three open reading frames 3′ of the *cipA* gene of *C. thermocellum* (13). We refer to it as the SLH domain.

We have made an exhaustive search of sequence data bases for further occurrences of the SLH domain and have identified seven additional proteins that contain it (Fig. 7). The SLH sequences are strongly divergent, with an average identity of only 27%. A single residue, Gly-28, is universally conserved. Three of the proteins, Ompα from *Thermotoga maritima*, the S-layer protein of *Thermus thermophilus*, and a 44-kDa open reading frame from *Synecchocystis* strain PCC6803, form a subgroup within the family: they contain a single SLH domain followed by a predicted coiled-coil domain, and their SLH domains are most closely related to each other. Since *T. maritima* is the second deepest known branch within the eubacteria, this subgroup of proteins may represent an ancestral form of SLH proteins. The other members of the family contain several SLH domains and no predicted coiled-coil segments. Their SLH sequences are generally more similar between proteins than within proteins, indicating that the event which led to multiple repeats must be fairly ancient.

Consensus secondary-structure predictions by the methods of Garnier et al. (14), Chou and Fasman (5), and Rost and Sander (25) (Fig. 7) agree with respect to the main structural elements of the SLH domain: two α-helices flanking a β-strand. This may indicate that several SLH sequences associate to form a native structure, since a single β-strand is structurally meaningless. Four insertion or deletion sites are found; all of them lie in predicted loop or turn regions. In the three open reading frame proteins from *C. thermocellum*, the repeats start with the predicted β-strand and have the first helix added at their C terminus. Since these repeats are likely to fold into the same structure as the repeats in the other proteins, this arrangement may simply reflect a change in connectivity.

SLH sequences appear to form an independently folding domain that can be appended to proteins in a modular way. They are found either at the very beginning or at the very end of proteins. The modular character is most clearly seen in the alkaline cellulase from *Bacillus* strain KSM-635, which is 71% identical to the alkaline cellulase from *Bacillus* strain 1139 but contains 180 additional residues at its N terminus that form the three SLH repeats (Fig. 8A). Another modular protein is the S-layer protein of *Bacillus sphaericus*, which is homologous to the *B. brevis* 47 middle wall SLH region at its N terminus and homologous to the outer wall protein over the rest of the sequence (Fig. 8B). This suggests that the S-layer in *B. sphaericus* represents a telescoped version of the S-layer of *B. brevis* 47.

Several considerations make it likely that SLH sequences serve as an anchor to the peptidoglycan. (i) All proteins that contain SLH domains also contain a signal sequence, and many are known to be attached to the cell wall. (ii) The most highly

```
            CF     HHH      HHHHHHHHHHHH  SSSSS                 HHHHHHHHHHHHHHHHHHHHHH
            GOR             HHHHHHHHHHHH  SSS                   HHHHHHHHHHHHHHHHHHHHH
            RS            HHHHHHHHHHHHH    SS                    HHHHHHHHHHHHH

                  1        10         20         30         40         50         60         70
Tm Omp         21  FFPDVPK-DHWAYEYVWKLWQR-GIFIGYPDGEFKGD-------RYITRYEAATAVSRLLDFIEQKMLAGAS
Tt SLP         24  QFSDVPA-GHWAKEAVEALAAK-GIILGFPDGTFRGN-------ENLTRYQAALLIYRLLQQIEEELKTQGT
Ak SLP         30  PFTDVKD-DAPYASAVARLYAL-NITNGVGDPKFGVD-------QPVTRAQMITFVNRMLGYEDLAEMAKSEKS
                   AFKDVPQ-NHWAVGQINLAYKL-GLAQGVGNGKFDPN-------SELRYAQALAFVLRALGFKDLD
                   ----------WPYGYLAKAQDL-GLVHG
Bb MWP         29  TTTAPKM-DADMEKTVKRLEAL-GLVAGYGNGEYGVD-------KTITRAEFATLVVRARGLEQGAKLAQFSN
                   TYTDVKS-TDWFAGFVNVASGE-EIVKGFPDKSFKPQ-------NQVTYAEAVTMIVRALGYEPSVKGV
                   ----------WPNSMISKASEL-NI
Bb HWP         57  TTTAPKM-DAAMEKTVKRLEAL-GLVAGYGNGDFGVD-------KTITRAEFATLIVRARGLEQGAKLAQFNT
                   TYTDVRS-TDWFAGFVNVASGE-EIVKGFPDKSFKPQ-------NQVTYAEAVTMIVRALGYEPSVRGV
                   ----------WPNSMISKGSEL-NI
Bs SLP         32  QLNDFNKISGYAKEAVQSLVDA-GVIQGDANGNFNPL-------KTISRAEAATIFTNALELEAEGDV
                   NFKDVKA-DAWYYDAIAATVEN-GIFEGVSATEFAPN-------KQLTRSEAAKILVDAFELEGEGDLS
                   EFADASTVKPWAKSYLEIAVAN-GVIKG
B Cel          40  PFSDVKK-TSWSFPYIKDLYEQ-EVITGTSATTFSPT-------DSVTRAQFTVMLTRGLGLEASSKDY
                   PFKDRK---NWAYKEIQAAYEA-GIVTGKTNGEFAPN-------ENITREQMAAMAVRAYEYLENELSLPEEQR
                   EYNDSSSISTFAQDAVQKAYVL-ELMEGNTDGYFQPK-------RNSTREQSAKVISTLL
Ct XynX       907  TFNDIKD--NWAKDVIEVLASR-HIVEGMTDTQYEPS-------KTVTRAEFTAMILKLLNIKEEAYNG
                   EFSDVKN-GDWYANAIEAAYKA-GIIEGDGKN-MRPN-------DSITREEMTSIAMRAYEMLTSYKEENIGAT
                   SFNDDKSISDWAKNVVANAAKL-GIINGEPSNVFAPK-------GIATRAEAAAIIYGLLEKSNNL*
Ts XynA      1055  TFDDIK--NSWAKDAIEVLASR-HIVEGMTDTQYEPN-------KTVTRAEFTAMILRLLNIKEEQYSG
                   EFSDVNS-GDWYANAIEAAYKA-GIIEGDGKN-ARPN-------DSITREEMTQ*
Ct Orf2       1453                        AYLRGYPDGSFRPE-------RNITRAEAAVIFAKLLGADESYGAQSAS
                   PYSDLAD-THWAAWAIKFATSQ-GLFKGYPDGTFKPD-------QNITRAEFATVVLHFLTKVKGQEIMSKLATIDISNP
                   KFDDCV--GHWAQEFIEKLTSL-GYISGYPDGTFKPQ-------NYIKRSESVALINRALERGPLNGAPK
                   LFPDVNE-SYWAFGDIMDGALD
Ct Orf3       482                        SYLTGYPDKMFRPE-------KSITRAEAAVIFAKLLGANENTKINYNV
                   SYTDVDS-SHWASWAIKFVSYK-KLFTGYPDGSFKPN-------QNITRAEFSTVVFKLLVSEKGLKEEKIEKS
                   KFGDTK--GHWAQQFIEQLSDL-GYINGYPDGTFKPN-------NNIKRSESVALINRAMGRGPLHGAPQ
                   VFEDVPQ-THWAFKDIAEGVLN
Ct Orf4       241                        PFLKGYPGGLFKPE-------NNITRAEAAVIFAKLLGADENSAGKNSSI
                   TFKDLKD-SHWAAWAIKYVTEQ-NLFGGYPDGTFMPD-------KSITRAEFATVTYKFLEKLGKIEQGTDVKT
                   QLKDIE--GHWAQKYIETLVAK-GYIKGYPDETFRPQ-------ASIKRAESVALINRSLERGPLNGAVL
                   EFTDVPV-NYWAYKDIAEGVIY
S Orf          64  ELRDVQP-SDWAFAALQSLVERYGCLVGYPDRTYRGDAEGTLRARPLSRYEFAAGLNACLNTIEQLL
```

FIG. 7. Alignment of SLH sequences from *T. maritima* Ompα (Tm Omp) (GenBank entry name TMOMPAA), the *T. thermophilus* 100-kDa S-layer protein (Tt SLP) (GenBank entry name TTSLPP100), the *A. kivui* S-layer protein (Ak SLP) (Swissprot entry name SLAP_ACEKI), the *B. brevis* 47 middle wall protein (Bb MWP) (Swissprot entry name CWPM_BACBR), the *B. brevis* HPD31 hexagonal wall protein (Bb HWP) (Pir entry name A35129), the *B. sphaericus* 125-kDa S-layer protein (Bs Slp) (Pir entry name A33856), the *Bacillus* strain KSM-635 alkaline cellulase (B Cel) (Swissprot entry name GUN_BACS6), the *C. thermocellum* exoxylanase (Ct XynX) (GenBank entry name CLOCELXA), the *Thermoanaerobacter saccharolyticum* endoxylanase (Ts XynA) (GenBank entry name TEOENDXYLA), the *C. thermocellum* open reading frames 2, 3, and 4 (Ct Orf2, Ct Orf3, and Ct Orf4, respectively) (GenBank entry name CTANCA), and the *Synechocystis* strain PCC6803 hypothetical 44-kDa protein (S Orf) (Swissprot entry name YGLC_SYNY3). Numbering starts at the putative initiation codons and includes signal sequences. The C terminus of a protein is marked by an asterisk. Residues shown in boldface type are identical in more than half the sequences of the alignment. Secondary-structure elements (H [α-helix] and S [β-strand]) were predicted by the methods of Chou and Fasman (CF) (5), Garnier et al. (GOR) (14), and Rost and Sander (RS) (25).

conserved domain between the *B. brevis* 47 middle wall protein and its *B. brevis* HPD31 homolog is the SLH repeat (93% identity) (Fig. 8B), as would be expected of the anchoring domain. (iii) The S-layer protein of *B. sphaericus* 2362 is essentially an outer wall protein fused to the SLH domains of the middle wall protein. Since only the middle wall protein binds to the peptidoglycan, this function probably resides in the SLH domains. (iv) Proteolytic cleavage of an 18-kDa fragment from the S-layer protein of *B. sphaericus* 9602 yields monomers that assemble to native-like layers but fail to bind to the peptidoglycan (16). Thus, the anchoring domain in this protein is located at one of the termini of the protein, as the SLH sequences would be. For comparison, the expected mass of the SLH domains from *B. sphaericus* 2362 is 16 kDa. (v) Ompα forms a rod-shaped spacer between the inner and outer membranes of *T. maritima* cells and is inserted via a transmembrane sequence in the outer membrane (9). The connection to the inner membrane is abolished by lysozyme digestion, indicating that Ompα interacts with the peptidoglycan. This function is unlikely to be located in the coiled-coil segment and probably resides in the N-terminal SLH sequence. Other S-layers that do not possess SLH sequences, e.g., the outer wall protein of *B. brevis* or the S-layers of *Caulobacter crescentus* and *Aeromonas salmonicida*, also do not interact with the peptidoglycan but are in contact either with the outer membrane or, in the case of *B. brevis*, with the underlying middle wall protein. The only exception appears to be the S-layer from *T. thermophilus*, which does possess an SLH domain but is supposed to be in contact with the outer membrane.

**Domain structure of the *A. kivui* S-layer protein.** Probable domains of the *A. kivui* S-layer protein can be inferred from sequence homology, proteinase K digestion, and secondary-structure prediction (Fig. 9). A domain from residue 27 to approximately residue 200 (18 kDa) is defined by sequence homology to the outer wall protein of *B. brevis*. Proteolysis with proteinase K yields two further, protease-resistant domains, one from residue 215 to approximately residue 475 (33 kDa) and one from residue 480 to approximately residue 634 (15 kDa). The boundaries of these domains correspond closely to those of the α (residues 27 to 217), β (residues 218 to 480), and

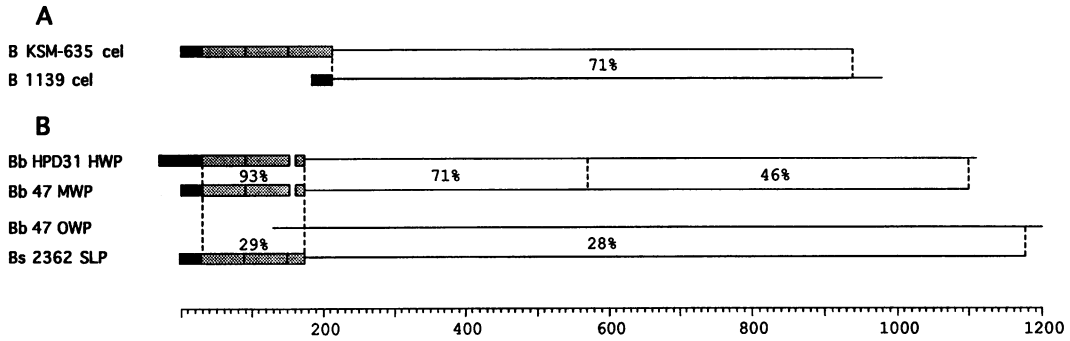


FIG. 8. Modular nature of SLH sequences. In both panels, signal sequences are shown in dark grey and SLH sequences are shown in light grey. (A) Alkaline cellulase from *Bacillus* strain KSM-635 (B KSM-635 cel) is 71% identical to alkaline cellulase from *Bacillus* strain 1139 (B 1139 cel) but contains three N-terminal SLH sequences that are not found in the other protein. (B) In *B. brevis* 47 (Bb 47), the S-layer is formed by a hexagonal middle wall protein (MWP) that is attached to the peptidoglycan and an outer wall protein (OWP) that connects to the middle wall protein. In *B. brevis* HPD31 (Bb HPD31), only a MWP homolog, the hexagonal wall protein (HWP), is found; this is 93% identical to the MWP in the SLH region, 71% identical in a domain predicted to consist largely of β-strands, and 46% identical over the rest of the sequence. In *B. sphaericus* 2362 (Bs 2362), the S-layer is formed by a protein that is 29% identical to the MWP in the SLH region and 28% identical to the OWP over the rest of the sequence. The scale at the bottom represents residue numbers.

mixed α and β (residues 481 to 736) domains defined by secondary-structure prediction. Corresponding domains can be defined in the *B. brevis* 47 middle wall protein and *B. brevis* HPD31 hexagonal wall protein from sequence conservation and secondary-structure prediction.

Also, three domains can be defined from the 3-D reconstruction. If we assume that the N-terminal SLH sequences interact with the peptidoglycan and that the domains are sequential in the protein sequence, then the following correlation can be made: the N-terminal SLH region forms the bottom of the core, the protease-resistant 33-kDa domain forms the upper part of the core and the short arm of the Y-domain, the protease-resistant 15-kDa domain forms the long arm of the Y-domain, and the protease-sensitive C terminus forms the L-domain. This arrangement is also most compatible with the location of the glycosylation sites, which then lie in the long arm of the Y-domain and thus in the most exposed part of the layer. In an inverse arrangement, they would lie in the core.

This domain organization and, in particular, the potential interaction of the SLH sequence with the peptidoglycan will be the subject of future studies.

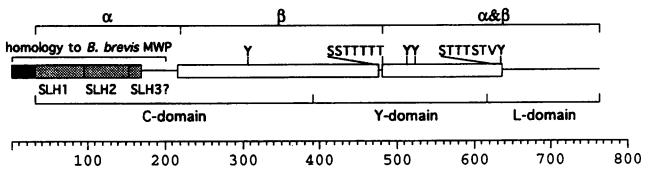


FIG. 9. Domain structure of the *A. kivui* S-layer protein. The signal sequence is shown in dark grey, the SLH sequences are shown in light grey, and the two protease-resistant fragments are unshaded. The four glycosylated tyrosines are labeled, as are two Ser-Thr clusters located at or immediately adjacent to the C-terminal ends of the protease-resistant fragments. The first cluster is significant at the 0.1% level, and the second is significant at the 1% level. The only protein of known structure that contains a Ser-Thr cluster at this level of significance is glucoamylase from *Aspergillus awamori*, also a glycoprotein, where the cluster (358 T-T-SSSSST-SS 370) is located mainly in a surface-exposed loop. Such a location is also likely for the Ser-Thr clusters in the *A. kivui* S-layer protein, given their protease accessibility and the proximity of the second cluster to a glycosylation site. α, α domain; β, β domain; α&β, mixed α and β domain. The scale at the bottom represents residue numbers.

## ACKNOWLEDGMENT

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (Ba 618/6-3).

## REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
2. **Baumeister, W., and H. Engelhardt.** 1987. Three-dimensional structure of bacterial surface layers, p. 109–154. *In* J. R. Harris and R. W. Horne (ed.), Electron microscopy of proteins, vol. 6. Academic Press, London.
3. **Cejka, Z., R. Hegerl, and W. Baumeister.** 1986. Three-dimensional structure of the surface layer protein of *Clostridium thermohydrosulfuricum*. J. Ultrastruct. Mol. Struct. Res. **96**:1–11.
4. **Chalcroft, J. P., and C. L. Davey.** 1984. A simply constructed extreme-tilt holder for the Philips eucentric goniometer stage. J. Microsc. (Oxford) **134**:41–48.
5. **Chou, P. Y., and G. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47**:251–276.
6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
7. **Dooley, J. S. G., H. Engelhardt, W. Baumeister, W. W. Kay, and T. J. Trust.** 1989. Three-dimensional structure of an open form of the surface layer from the fish pathogen *Aeromonas salmonicida*. J. Bacteriol. **171**:190–197.
8. **Ebisu, S., A. Tsuboi, H. Takagi, Y. Naruse, H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1990. Conserved structures of cell wall protein genes among protein-producing *Bacillus brevis* strains. J. Bacteriol. **172**:1312–1320.
9. **Engel, A. M., Z. Cejka, A. Lupas, F. Lottspeich, and W. Baumeister.** 1992. Isolation and cloning of Ompα, a coiled-coil protein spanning the periplasmic space of the ancestral eubacterium *Thermotoga maritima*. EMBO J. **11**:4369–4378.
10. **Engelhardt, H., Z. Cejka, and W. Baumeister.** 1988. Three-dimensional structure of surface layers from various *Bacillus* and *Clostridium* species, p. 96–100. *In* U. B. Sleytr, P. Messner, D. Pum, and M. Sára (ed.), Crystalline bacterial cell surface layers. Springer-Verlag, Berlin.
11. **Engelhardt, H., S. Gerbl-Rieger, D. Krezmar, S. Schneider-Voss, A. Engel, and W. Baumeister.** 1990. Structural properties of the outer membrane and the regular surface protein of *Comamonas acidovorans*. J. Struct. Biol. **105**:92–102.

12. **Engelhardt, H., W. O. Saxton, and W. Baumeister.** 1986. Three-dimensional structure of the tetragonal surface layer of *Sporosarcina ureae*. J. Bacteriol. **168:**309–317.
13. **Fujino, T., P. Beguin, and J.-P. Aubert.** 1993. Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. J. Bacteriol. **175:**1891–1899.
14. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120:**97–120.
15. **Gribskov, M., R. Lüthy, and D. Eisenberg.** 1990. Profile analysis. Methods Enzymol. **183:**146–159.
16. **Hastie, A. T., and C. C. Brinton, Jr.** 1979. Specific interaction of the tetragonally arrayed protein layer of *Bacillus sphaericus* with its peptidoglycan sacculus. J. Bacteriol. **138:**1010–1021.
17. **Kleffel, B., R. M. Garavito, W. Baumeister, and J. P. Rosenbusch.** 1985. Secondary structure of a channel forming protein: porin from *E. coli* outer membranes. EMBO J. **4:**1589–1592.
18. **Leigh, J. A., F. Meyer, and R. S. Wolfe.** 1981. *Acetogenium kivui*, a new thermophilic hydrogen-oxidizing, acetogenic bacterium. Arch. Microbiol. **129:**275–280.
19. **Leigh, J. A., and R. S. Wolfe.** 1983. *Acetogenium kivui* gen. nov., sp. nov., a thermophilic acetogenic bacterium. Int. J. Syst. Bacteriol. **33:**886.
20. **Lupas, A., M. Van Dyke, and J. Stock.** 1991. Predicting coiled coils from protein sequences. Science **252:**1162–1164.
21. **Peters, J., M. Peters, F. Lottspeich, and W. Baumeister.** 1989. S-layer protein gene of *Acetogenium kivui*: cloning and expression in *Escherichia coli* and determination of the nucleotide sequence. J. Bacteriol. **171:**6307–6315.
22. **Peters, J., S. Rudolf, H. Oschkinat, R. Mengele, M. Sumper, J. Kellermann, F. Lottspeich, and W. Baumeister.** 1992. Evidence for tyrosine-linked glycosaminoglycan in a bacterial surface protein. Biol. Chem. Hoppe-Seyler **373:**171–176.
23. **Phipps, B. M., R. Huber, and W. Baumeister.** 1991. The cell envelope of the hyperthermophilic archaebacterium *Pyrobaculum organotrophum* consists of two regularly arrayed protein layers: three-dimensional structure of the outer layer. Mol. Microbiol. **5:**253–265.
24. **Rasch, M., W. O. Saxton, and W. Baumeister.** 1984. The regular surface layer of *Acetogenium kivui*: some structural, developmental and evolutionary aspects. FEMS Microbiol. Lett. **24:**285–290.
25. **Rost, B., and C. Sander.** 1993. Prediction of protein structure at better than 70% accuracy. J. Mol. Biol. **232:**584–599.
26. **Saxton, W. O.** 1985. Computer generation of shaded images of solids and surfaces. Ultramicroscopy **16:**387–394.
27. **Saxton, W. O., and W. Baumeister.** 1982. The correlation averaging of a regularly arrayed bacterial surface protein., J. Microsc. (Oxford) **127:**127–138.
28. **Saxton, W. O., and W. Baumeister.** 1986. Principles of organization in S-layers. J. Mol. Biol. **187:**251–253.
29. **Saxton, W. O., W. Baumeister, and M. Hahn.** 1984. Three-dimensional reconstruction of imperfect two-dimensional crystals. Ultramicroscopy **13:**57–70.
30. **Saxton, W. O., T. J. Pitt, and M. Horner.** 1979. Digital image processing: the SEMPER system. Ultramicroscopy **4:**343–354.
31. **Tsuboi, A., H. Engelhardt, U. Santarius, N. Tsukagoshi, S. Udaka, and W. Baumeister.** 1989. Three-dimensional structure of the surface protein layer (MW layer) of *Bacillus brevis* 47. J. Ultrastruct. Mol. Struct. Res. **102:**178–187.
32. **Woodcock, C. L., and W. Baumeister.** 1990. Different representations of a protein structure obtained with different negative stains. Eur. J. Cell Biol. **51:**45–52.