View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Regensburg Publication Server

JOURNAL OF BACTERIOLOGY, Dec. 1990, p. 7111–7118
0021-9193/90/127111-08$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 172, No. 12

Primary Structure and Glycosylation of the S-Layer Protein of *Haloferax volcanii*

MANFRED SUMPER,* ELKE BERG, REINER MENGELE, AND ISOLDE STROBEL

Lehrstuhl Biochemie I, Universität Regensburg, 8400 Regensburg, Federal Republic of Germany

Received 8 June 1990/Accepted 14 September 1990

The outer surface of the archaebacterium *Haloferax volcanii* (formerly named *Halobacterium volcanii*) is covered with a hexagonally packed surface (S) layer. The gene coding for the S-layer protein was cloned and sequenced. The mature polypeptide is composed of 794 amino acids and is preceded by a typical signal sequence of 34 amino acid residues. A highly hydrophobic stretch of 20 amino acids at the C-terminal end probably serves as a transmembrane domain. Clusters of threonine residues are located adjacent to this membrane anchor. The S-layer protein is a glycoprotein containing both N- and O-glycosidic bonds. Glucosyl-(1→2)-galactose disaccharides are linked to threonine residues. The primary structure and the glycosylation pattern of the S-layer glycoproteins from *Haloferax volcanii* and from *Halobacterium halobium* were compared and found to exhibit distinct differences, despite the fact that three-dimensional reconstructions from electron micrographs revealed no structural differences at least to the 2.5-nm level attained so far (M. Kessel, I. Wildhaber, S. Cohe, and W. Baumeister, EMBO J. 7:1549–1554, 1988).

Two-dimensional regular arrays of proteins on the cell surface (S layers) have now been recognized to be ubiquitously present in eubacteria as well as in archaebacteria (see reference 21 for a review). The primary structures of five different S-layer polypeptides from four phylogenetically very different organisms have recently been reported (6, 17, 18, 23, 24). The S-layer protein from *Halobacterium halobium* was the first procaryotic glycoprotein to be discovered (12). Its primary structure was derived from the cloned gene (6), and the chemical structures of the saccharide chains attached to this glycoprotein have been determined (see references 7 and 22 for a review). This detailed structural knowledge makes the halobacterial S layer an attractive object for the elucidation of its three-dimensional structure by electron microscopy and the techniques of image processing. However, electron microscopy of extreme halophiles has been hampered by the high salt concentrations required to maintain the integrity of the S layer. Fortunately, this is not true for the S layer of a moderate halophile from the Dead Sea, *Halobacterium volcanii* (renamed *Haloferax volcanii*). Recently, the corresponding three-dimensional structure at a resolution of 2 nm has been derived. These data were combined and integrated with the chemical data known from the S-layer glycoprotein of *Halobacterium halobium* into a model of a halobacterial cell envelope (4). To eliminate the unsatisfactory need for integrating experimental data from different species, we decided to start a detailed chemical characterization of the S-layer protein from *H. volcanii*. In this paper, we report the primary structure as derived from the cloned gene and describe the saccharides O-glycosidically linked to this glycoprotein.

MATERIALS AND METHODS

Bacterial strains. *H. volcanii* was obtained from M. Mevarech (Israel) and cultivated as described previously (13).

Isolation of the S-layer protein. An 800-ml portion of a stationary culture of *H. volcanii* cells was centrifuged at 7,000 × *g* for 30 min. The cell pellets were carefully resuspended in 200 ml of a modified medium (as growth medium; however, without yeast extract and tryptone and adjusted to 150 mM Mg^{2+}). After addition of an equimolar amount of 0.5 M EDTA (pH 6.8; 60 ml), the suspension was shaken for 30 min at 37°C, and the resulting spheroplasts were removed by three consecutive centrifugations: 15 min at 3,000 × *g*, 5 min at 7,000 × *g*, and 10 min at 13,000 × *g*. The supernatant containing the S-layer protein as the main component was concentrated by ultrafiltration, dialyzed, and applied to a 6% sodium dodecyl sulfate (SDS)-polyacrylamide gel. The glycoprotein was visualized by incubating the gel in 200 mM KCl, eluted from the homogenized gel slice with water, dialyzed, and precipitated with 10 volumes of cold ethanol. Deglycosylation with anhydrous hydrogen fluoride was performed as described previously (14).

Proteolytic digestion and separation of peptides. A 500-μg amount of S-layer protein was dissolved in 1.25 ml of buffer (0.1 M *N*-ethyl morpholino acetate [pH 6.8], 10 mM $CaCl_2$) and digested with 10 μg of trypsin at 37°C. After 4 h, an additional 10 μg of trypsin was added and incubation was continued for 10 h. After lyophilization, peptides were dissolved in 6 M guanidinium hydrochloride and separated by high-performance liquid chromatography on a LiChrospher 100 CH-18/2 (10 μm) reverse-phase column. Peptides were eluted by a 120-min linear gradient from 5 to 45% acetonitrile in 0.1% trifluoroacetic acid. Peptides were sequenced with an automated gas phase sequencer (Applied Biosystems, Inc., Foster City, Calif.) as described by Lottspeich (10).

Isolation of DNA. Chromosomal DNA and plasmid DNA were prepared as described previously (6).

Generation of a cDNA probe by PCR. The sense and antisense primer (see Table 1) were synthesized by using the phosphoramidite method (11). The polymerase chain reaction (PCR) was performed in 100 μl of 50 mM Tris hydrochloride (pH 8.5)–50 mM NaCl–5 mM $MgCl_2$–2 mM dithiothreitol–200 μM concentration of each deoxynucleoside triphosphate. A 100-pmol portion of each primer, 1.5 μg of chromosomal DNA, and 2.5 U of *Taq* polymerase (Perkin-Elmer-Cetus) were added. After 30 cycles of amplification (1 min at 94°C, 1 min at 50°C, and 1 min at 72°C), a product of

* Corresponding author.

the expected size (102 bp) was detected on a 3% NuSieve agarose gel. This fragment was eluted from the gel and purified by phenol-chloroform extraction. To ensure blunt ends, the fragment was treated with the Klenow enzyme in the presence of all four deoxynucleoside triphosphates. After phosphorylation with ATP, using T4 polynucleotide kinase, the fragment was ligated into the *Sma*I site of vector pUC18. The recombinant plasmid was purified from transformants of *Escherichia coli* JM 109 and sequenced by the dideoxy-chain termination method of Sanger et al. (20).

Cloning procedures. Chromosomal DNA of *H. volcanii* was digested with restriction endonuclease *Bam*HI, size fractionated to 5 to 6 kb on a 1% agarose gel, electroeluted, and recovered by precipitation. After ligation into *Bam*HI-cut and dephosphorylated pUC18, *E. coli* JM 109 was transformed and positive clones were identified by colony hybridization with radioactive PCR-generated and cloned insert DNA.

Nucleotide sequence analysis. Sequence analysis was performed with the dideoxy-chain termination method of Sanger et al. (20) by using [α-^{35}S]dATP (Amersham).

β-Elimination. The purified S-layer protein was incubated in 0.1 M NaOH–1 M $NaBD_4$ for 14 h at 37°C. After neutralization with 5 M acetic acid, the sample was applied to a column (2 ml) of Dowex AG-50W-X8(H^+) (Bio-Rad Laboratories). The effluent was dried and evaporated three times with 1 ml of methanol to remove all borate. To remove all anionic substances (e.g., contaminating RNA), the material was dissolved in water and applied to a column (2 ml) of Dowex AG1 (acetate). The effluent was used for further analysis.

Carbohydrate analysis. Sugar composition was determined by gas chromatography (GC) with a model 3700 GC (Varian Instruments, Palo Alto, Calif.) equipped with a Durabond 1701 capillary column (30-m; J & W Scientific, Folsom, Calif.). Total carbohydrate analysis was performed by using pentafluoropropionic acid derivatives of the sugar methylglycosides as described in reference 8.

Permethylation. The deuterium-reduced saccharide obtained after β-elimination was permethylated by the procedure of Hakomori (2), as modified by Waeghe et al. (25). The permethylated saccharide was purified with Sep-pak C_{18} cartridges (Waters Inc.) by elution with 20% acetonitrile. After hydrolysis in 4 M trifluoroacetic acid for 2 h at 100°C, the methylated sugars were reduced with $NaBH_4$, peracetylated, and analyzed by GC-mass spectrometry (MS).

GC-MS. All GC-MS analyses were performed with a Hewlett-Packard model 5995 GC equipped with a Durabond 1701 fused silica capillary column (30-m; J & W Scientific). The carrier gas was helium at 1 ml/min. The temperature gradient for methylated alditol acetates was 2°C/min, starting at 140°C.

The assignment of methylated alditol acetates was accomplished by using reference mass spectra (3).

Amino acid derivatives were analyzed at an initial temperature of 60°C. After 5 min, the temperature was raised at 5°C/min to a final temperature of 200°C. The single ion scanning trace was adjusted to 155 atomic mass units.

Determination of threonines involved in O-glycosidic linkages. In the course of β-elimination with $NaBD_4$, the threonine residues linked to saccharides become converted to deuterated α-aminobutyric acid. Amino acids obtained by hydrolysis of the β-eliminated protein were converted to their corresponding methyl esters and treated with trifluoroacetic anhydride (19). These derivatives were analyzed by GC-MS. The deuterated α-aminobutyric acid derivative was identified by comparing the retention times with an authentic standard and by a shift of the main mass peak from 154 to 155 atomic mass units. Exact quantification was impossible, because the substance was not completely separated from the glycine derivative.

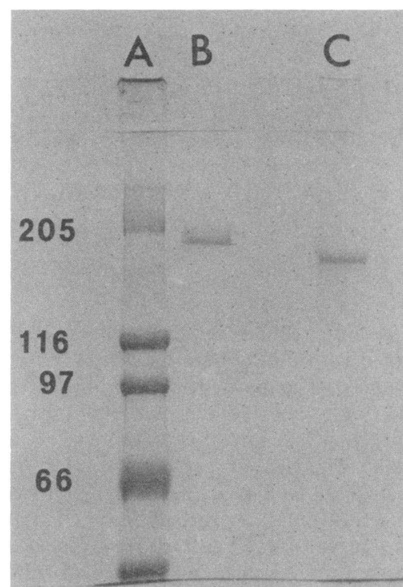

FIG. 1. SDS-polyacrylamide gel patterns of the *H. volcanii* S-layer glycoprotein. Lanes: A, molecular mass standards (in kilodaltons); B, purified S-layer protein; C, S-layer protein after deglycosylation by treatment with anhydrous hydrogen fluoride.

To get an approximate value for the extent of O-glycosylation of threonines, β-eliminated glycoprotein was hydrolyzed for 24 h at 100°C in 6 M HCl. The resulting amino acids were determined with an automatic amino acid analyzer (Biotronic) equipped with a postcolumn ninhydrin detection system.

RESULTS

Isolation of S-layer protein. Chelation of Mg^{2+}ions by adding EDTA to the growth medium converts cells of *H. volcanii* from their normal cupped, disk shape to a spherical form. This transition causes massive shedding of the S-layer protein, which can be recovered therefore from the modified growth medium in a highly enriched state. Purification to homogeneity was achieved by preparative SDS-polyacrylamide gel electrophoresis. After treatment of the S-layer protein with anhydrous hydrogen fluoride, selectively cleaving O-glycosidic bonds (14), a significant reduction of the apparent molecular mass was observed on SDS-polyacrylamide gel electrophoresis (Fig. 1). This strongly indicates the presence of covalently bound saccharides. A carbohydrate content of 12% (by weight) was determined by the phenol-sulfuric acid method (1).

Protein chemical studies. Automated Edman degradation of the purified protein resulted in the following N-terminal amino acid sequence: ERGNLDADSESF-KTIQ. The purified S-layer polypeptide was readily digested by trypsin, and the resulting peptide mixture was separated by reverse-phase C_{18} high-performance liquid chromatography. The materials of well-separated peaks were directly submitted to amino acid sequence analysis on an automated gas phase sequencer. All amino acid sequence data obtained are sum-

TABLE 1. Amino acid sequences of tryptic peptides from *H. volcanii* S-layer glycoprotein and selection of primers for generating a cDNA probe by PCR

Peptide	Amino acid sequence
1	TIQSGDR
2	TSGVSSSNSIR
3	SLTGTYR
4	VNADYNYESAEK
5	DLSDLQNDATTEQAK
6	TVEADDGDNTDTVNVEVVSER
7	NIGDTSEVGIA-SSAT-TSGSSTGPTVETADIAY
	primer primer
	Sense primer: ATT ATT GGN GAT AC C C
	Antisense primer: TA NGC GAT GTC NGC
8	YLDDSEVDLEVYD
9	LAESXTSIDSIYPDAAEAAGINPVATGE

marized in Table 1. Peptide 7 turned out to be an unusally large fragment from which 34 amino acid positions could be determined. Therefore, this sequence information was used to produce a highly specific DNA probe by a recently developed variant of the PCR technique (9).

Generation of a 102-bp DNA probe by PCR. The amino acid sequence of peptide 7 was used to synthesize a sense primer to amino acid sequence 1 to 5. An antisense primer was synthesized to the amino acid sequence 30 to 34 of the same peptide (Table 1). These primers were used to amplify the corresponding segment of the S-layer protein gene, using genomic DNA from *H. volcanii* as a template. Thirty cycles of amplification produced a DNA fragment of about 100 bp in length that was cloned by blunt-end ligation into vector pUC18. Sequencing of this insert revealed an open reading frame (ORF) matching the amino acid sequence of peptide 7. Therefore, this insert was used as a highly specific probe for the gene of the S-layer protein.

Cloning the S-layer gene. Genomic DNA of *H. volcanii* was digested with *Bam*HI and analyzed by a Southern blot experiment, using the PCR-generated fragment as the radioactive probe. A single hybridization signal was obtained at 5.5 kb. Therefore, the 5.5-kb fraction from a *Bam*HI digest of genomic DNA was ligated directly into vector pUC18. The resulting transformants were screened with the same probe. More than 1% of the transformants contained the desired gene. The 5.5-kb insert was sequenced in both directions, using synthetic oligonucleotides as primers. The restriction map of the cloned DNA fragment is shown in Fig. 2. Figure 3 summarizes the results of the sequencing experiments. The 2,820-bp segment shown covers the region from the *Bam*HI site upstream from the ORF to the second *Sst*I site. The ORF starts at position 53 with an ATG codon and ends at position 2536 with a TAA codon.

FIG. 2. Restriction map of the *H. volcanii* chromosomal DNA region containing the S-layer protein gene. Heavy bar represents the S-layer gene coding region.

Protein structure. From the N-terminal sequence of the mature protein and from known features of presequences, we conclude that the S-layer protein is synthesized as a precursor with a 34-amino-acid extension. The amino acid sequence at the potential cleavage site is Ala-Ala-Ala, a motif frequently used by signal peptidases (16). The mature polypeptide contains 794 amino acids, with a calculated M_r of 81,732. This value is much lower than the apparent molecular mass of the deglycosylated polypeptide derived from SDS-polyacrylamide gel electrophoresis (170 kDa; Fig. 1). Exactly the same abnormal electrophoretic behavior was found for the deglycosylated S-layer polypeptide of *Halobacterium halobium* (6). Both of these polypeptides may have a reduced capacity for SDS binding due to their extreme excess of hydrophilic amino acid residues. As a consequence, the reduced electrophoretic mobility would cause an overestimate of the molecular mass.

Nine stretches of the predicted amino acid sequence were confirmed by the amino acid sequence data derived from purified peptides. This allows the firm conclusion that translation of the ORF indeed represents the primary structure of the S-layer protein from *H. volcanii*. As demonstrated by hydropathy analysis (5), the entire polypeptide chain of the mature protein shows a single highly hydrophobic stretch of 20 amino acids only four amino acid positions away from the C terminus (Fig. 4). Probably this hydrophobic peptide serves as a membrane anchor. All other regions of the mature polypeptide consist mainly of polar amino acids and contain an excess of acidic residues. Next to the putative membrane anchor, four repeats of the amino acid sequence motif (D or E)-(threonine or serine)$_5$ are a remarkable feature. Seven potential N-glycosylation sites (N-X-S or N-X-T) are distributed throughout the polypeptide chain. Three of these sites were part of peptide (Table 1) and N-terminal sequences. These Edman degradation data confirmed the chemical modification of the corresponding asparagine residues, indicating the existence of N-glycosidically linked saccharides.

O glycosylation of the S-layer protein. Degradation of the S-layer protein by alkaline β-elimination in the presence of $NaBD_4$ liberates a neutral saccharide fraction, containing galactose and glucose in a 1:1 stoichiometry. Deuterium was exclusively incorporated at the C-1 position of the galactose, indicating the presence of O-glycosidic linkages of the type Thr (or Ser)-Gal. The result of permethylation analysis of this β-eliminated reduced saccharide is documented in Fig. 5. The resulting sugar derivatives are 1,5-di-*O*-acetyl-2,3,4,6-tetra-*O*-methylglucitol and 2-*O*-acetyl-1,3,4,5,6-penta-*O*-methylgalactitol (deuterium labeled at C-1), proving the presence of the 1,2-linked disaccharide glucosylgalactose as a structural element of the S-layer protein.

Threonine residues appear to be involved in this O-glycosidic linkage as we determined the presence of 4 to 6 mol of α-aminobutyric acid per mol of β-eliminated glycoprotein.

As mentioned above, the predicted amino acid sequence contains a total of seven potential N glycosylation sites. A detailed structural analysis of these N-glycosidically linked saccharides is beyond the scope of this paper.

DISCUSSION

In this study we report the primary structure of the S-layer glycoprotein from *H. volcanii* and the structure of O-glycosidically linked disaccharides. Since a three-dimensional reconstruction from electron micrographs of negatively stained cell envelopes of *H. volcanii* turned out to be

```
       GGATCCGCGTTATAGCGAACTGCGCGGAAAATACTGGCGCTACAAGTACACTATGACA     58
                                                           M  T        2

  59   AAGCTCAAAGATCAAACGCGTGCGATACTTCTCGCCACGCTGATGGTAACTTCCGTCTTT   118
   3   K  L  K  D  Q  T  R  A  I  L  L  A  T  L  M  V  T  S  V  F      22

 119   GCGGGAGCCATCGCGTTCACGGGAAGTGCCGCCGCGGAGCGTGGAAACCTCGACGCTGAC   178
  23   A  G  A  I  A  F  T  G  S  A  A  A  E  R  G  N  L  D  A  D      42

 179   AGCGAATCGTTCAACAAGACAATCCAATCCGGCGACCGCGTCTTCCTCGGTGAGGAGATT   238
  43   S  E  S  F  N  K  T  I  Q  S  G  D  R  V  F  L  G  E  E  I      62

 239   AGCACCGACGCCGGCCTCGGTGCTTCGAATCCGCTTCTGACGGGTACCGCGGGCAACTCG   298
  63   S  T  D  A  G  L  G  A  S  N  P  L  L  T  G  T  A  G  N  S      82

 299   GAAGGTGTCTCTCTCGACCTTTCCTCGCCCATCCCGCAGACTACGGAGAACCAGCCGCTT   358
  83   E  G  V  S  L  D  L  S  S  P  I  P  Q  T  T  E  N  Q  P  L     102

 359   GGCACGTACGACGTTGACGGCTCTGGCTCCGCCACGACCCCCAACGTCACGCTGCTGGCG   418
 103   G  T  Y  D  V  D  G  S  G  S  A  T  T  P  N  V  T  L  L  A     122

 419   CCGCGCATCACTGACAGCGAAATCCTGACGTCCTCCGGTGGCGACGTCACCGGCTCCGCT   478
 123   P  R  I  T  D  S  E  I  L  T  S  S  G  G  D  V  T  G  S  A     142

 479   ATCAGCTCCAGCGACGCTGGTAACCTCTACGTCAACGCTGACTACAACTACGAGTCGGCC   538
 143   I  S  S  S  D  A  G  N  L  Y  V  N  A  D  Y  N  Y  E  S  A     162

 539   GAGAAGGTCGAAGTGACCGTCGAAGACCCGAGCGGAACGGACATCACCAACGAGGTCCTG   598
 163   E  K  V  E  V  T  V  E  D  P  S  G  T  D  I  T  N  E  V  L     182

 599   AGCGGCACGGACACCTTCGTTGACGACGGCTCGATTGGCAGCACCTCGTCGACCGGCGGT   658
 183   S  G  T  D  T  F  V  D  D  G  S  I  G  S  T  S  S  T  G  G     202

 659   GGCGTCGGCATCGACATGTCCGATCAGGACGCTGGCGAGTACACGATCATCCTCGAAGGC   718
 203   G  V  G  I  D  M  S  D  Q  D  A  G  E  Y  T  I  I  L  E  G     222

 719   GCTGAAGACCTCGACTTCGGTGACGCTACTGAAACGATGACCCTCACGATCTCGTCGCAG   778
 223   A  E  D  L  D  F  G  D  A  T  E  T  M  T  L  T  I  S  S  Q     242

 779   GACGAGATCGGCATCGAACTCGACAGCGAATCCGTGACGCAGGGTACTGACGTTCAGTAC   838
 243   D  E  I  G  I  E  L  D  S  E  S  V  T  Q  G  T  D  V  Q  Y     262

 839   ACCGTCACGAACGGCATCGACGGCAACGAGCACGTTGTCGCTATGGACCTCTCCGACCTC   898
 263   T  V  T  N  G  I  D  G  N  E  H  V  V  A  M  D  L  S  D  L     282

 899   CAGAACGACGCCACGACGGAACAGGCCAAGGAGGTCTTCCGTAACATCGGCGACACCTCG   958
 283   Q  N  D  A  T  T  E  Q  A  K  E  V  F  R  N  I  G  D  T  S     302

 959   GAAGTCGGCATCGCGAACAGCAGCGCTACGAACACGAGCGGCTCCTCCACGGGCCCGACC  1018
 303   E  V  G  I  A  N  S  S  A  T  N  T  S  G  S  S  T  G  P  T     322

1019   GTTGAGACCGCTGACATCGCGTACGCAGTCGTCGAAATCGACGGTGCGAGCGCCGTTGGC  1078
 323   V  E  T  A  D  I  A  Y  A  V  V  E  I  D  G  A  S  A  V  G     342

1079   GGGATTGAAACGCAGTACCTCGACGACAGCGAAGTTGACCTCGAAGTCTACGACGCTGGC  1138
 343   G  I  E  T  Q  Y  L  D  D  S  E  V  D  L  E  V  Y  D  A  G     362

1139   GTGAGCGCGACTGCAGCAGTCGGCCAGGACGCGACGAACGACATCACGCTCACCATCGAA  1198
 363   V  S  A  T  A  A  V  G  Q  D  A  T  N  D  I  T  L  T  I  E     382

1199   GAAGGCGGTACGACGCTTTCGAGCCCGACCGGTCAGTACGTCGTTGGCTCGGAAGTCGAC  1258
 383   E  G  G  T  T  L  S  S  P  T  G  Q  Y  V  V  G  S  E  V  D     402

1259   ATTAACGGGACGGCTACCAGCTCCGACAGCGTCGCAATCTACGTCCGCGACGATGGCGAC  1318
 403   I  N  G  T  A  T  S  S  D  S  V  A  I  Y  V  R  D  D  G  D     422

1319   TGGCAGCTCCTCGAAATCGGTGGCGACAACGAAATCAGTGTCGACTCCGACGACACGTTC  1378
 423   W  Q  L  L  E  I  G  G  D  N  E  I  S  V  D  S  D  D  T  F     442

1379   GAAGAGGAAGACATCGCGCTCTCGGGGCTCTCCGGCGACGGTAGCAGTATCCTGTCGCTG  1438
 443   E  E  E  D  I  A  L  S  G  L  S  G  D  G  S  S  I  L  S  L     462

1439   ACGGGTACCTACCGCATCGGTGTCATCGACGCGTCGGACGCTGACGTTGGCGGCGACGGT  1498
 463   T  G  T  Y  R  I  G  V  I  D  A  S  D  A  D  V  G  G  D  G     482
```

FIG. 3. Nucleotide sequence of the gene for the *H. volcanii* S-layer protein. The signal peptidase cleavage site as derived from N-terminal amino acid sequence analysis of the mature protein is marked with an arrow. The amino acid sequence coded in the single ORF is also displayed. Partial amino acid sequences determined by Edman degradation of the S-layer protein and of tryptic peptides are underlined. Boxed regions represent potential N-glycosylation sites. The putative membrane-binding domain is marked by a dotted line.

```
1499  TCCGTCGACGACTCGCTTACGACGTCCGAATTCACCAGCGGTGTGAGCAGCAGTAACTCG  1558
 483   S  V  D  D  S  L  T  T  S  E  F  T  S  G  V  S  S  S  N  S   502

1559  ATTCGTGTGACCGACCAGGCCCTTACGGGTCAGTTCACCACGATCAACGGTCAGGTCGCT  1618
 503   I  R  V  T  D  Q  A  L  T  G  Q  F  T  T  I  N  G  Q  V  A   522

1619  CCTGTCGAGACCGGCACGGTTGACATCAACGGTACCGCCTCCGGTGCCAACTCCGTCCTC  1678
 523   P  V  E  T  G  T  V  D  I [N  G  T] A  S  G  A  N  S  V  L   542

1679  GTCATCTTCGTCGACGAGCGCGGTAACGTCAACTACCAGGAAGTGAGCGTTGACAGCGAC  1738
 543   V  I  F  V  D  E  R  G  N  V  N  Y  Q  E  V  S  V  D  S  D   562

1739  GGTACCTACGACGAGGACGACATCACTGTCGGCCTCACGCAGGGCCGGGTCACCGCCCAC  1798
 563   G  T  Y  D  E  D  D  I  T  V  G  L  T  Q  G  R  V  T  A  H   582

1799  ATCCTCTCGGTCGGTCGCGACAGCGCTATCGGTGACGGCTCGCTGCCGTCCGGCCCGAGC  1858
 583   I  L  S  V  G  R  D  S  A  I  G  D  G  S  L  P  S  G  P  S   602

1859  AACGGCGCAACCCTCAACGACCTGACGGGCTACCTCGACACGCTCGACCAAAACAACAAC  1918
 603   N  G  A  T  L  N  D  L  T  G  Y  L  D  T  L  D  Q  N  N  N   622

1919  AACGGTGAGCAGATCAACGAGCTCATCGCCTCCGAAACGGTTGACGAGACGGCGAGCGAC  1978
 623   N  G  E  Q  I  N  E  L  I  A  S  E  T  V  D  E  T  A  S  D   642

1979  GACCTGATCGTCACCGAGACGTTCCGTCTGGCCGAGTCCTCGACGTCCATCGACTCGATC  2038
 643   D  L  I  V  T  E  T  F  R  L  A  E  S  S  T  S  I  D  S  I   662

2039  TACCCGGACGCCGCTGAAGCGGCTGGCATCAACCCGGTCGCAACCGGTGAAACGATGGTC  2098
 663   Y  P  D  A  A  E  A  A  G  I  N  P  V  A  T  G  E  T  M  V   682

2099  ATCGCTGGCTCGACGAACCTCAAGCCGGACGACAACACCATCAGCATCGAAGTGACGAAT  2158
 683   I  A  G  S  T  N  L  K  P  D  D  N  T  I  S  I  E  V  T  N   702

2159  GAAGACGGCACGTCCGTCGCACTCGAAGACACTGACGAGTGGAACAACGACGGCCAGTGG  2218
 703   E  D  G  T  S  V  A  L  E  D  T  D  E  W  N  N  D  G  Q  W   722

2219  ATGGTCGAAATTGACACCACTGACTTCGAGACGGGTACGTTCACCGTGGAAGCCGACGAC  2278
 723   M  V  E  I  D  T  T  D  F  E  T  G  T  F  T  V  E  A  D  D   742

2279  GGCGACAACACGGACACCGTGAACGTCGAAGTCGTCTCCGAGCGTGAGGACACCACGACG  2338
 743   G  D  N  T  D  T  V  N  V  E  V  V  S  E  R  E  D  T  T  T   762

2339  TCCTCCGACAACGCGACGGACACGACGACGACCACGGACGGACCCACGGAAACCACGACG  2398
 763   S  S  D [N  A  T] D  T  T  T  T  T  D  G  P  T  E  T  T  T   782

2399  ACCGCAGAGCCGACTGAGACCACCGAAGAGCCGACTGAGGAGACCACCACGTCCTCCAAC  2458
 783   T  A  E  P  T  E  T  T  E  E  P  T  E  E  T  T  T  S  S  N   802

2459  ACGCCCGGCTTCGGTATCGCAGTCGCTCTCGTCGCACTCGTCGGCGCGGCCCTTCTCGCA  2518
 803   T  P  G  F  G  I  A  V  A  L  V  A  L  V  G  A  A  L  L  A   822

2519  CTCCGCCGCGAGAACTAACGCCGCCACCACTCCTCGTCACTCCGACGAGGACGTGACAAC  2578
 823   L  R  R  E  N  *

2579  AGCAGACGAGTAATCGCCGGACTCGTTCCGGTCTTTTATTTTCTTTCCCTACTCCACCGA  2638

2639  GAGACATCTCCGAGAGACCGCCATGCGCTTCGCACGGAACTCAGCGACCGCTGGTGTGCA  2698

2699  GACTAGGGTCGTGCCGAGCGAGACGACTCGTTGCGGCCTTCCGGTGAAACGGGGGTGAAT  2758

2759  CCGGACGCCTGTGCAACCCGGTTCTGCGAAACGCGTCGTGCGACGGGAAGCGTCGAGAGC  2818

2819  TC
```

possible (4), this halobacterial cell envelope now appears ideally suited for a three-dimensional analysis at high resolution. We have previously characterized in detail the S-layer glycoprotein from *Halobacterium halobium*. Common structural features of that glycoprotein and the S-layer polypeptide sequences known from eubacteria were recently analyzed (17). Comparison of the two halobacterial S-layer glycoproteins reveals common features but also substantial structural differences which may be summarized as follows.

(i) The schematic representation in Fig. 4. compares the S-layer proteins from *Halobacterium halobium* and *H. volcanii* with respect to the localization of potential glycosylation sites and putative membrane-binding domains. Both proteins exhibit at their C-terminal ends a stretch of 20 or 21

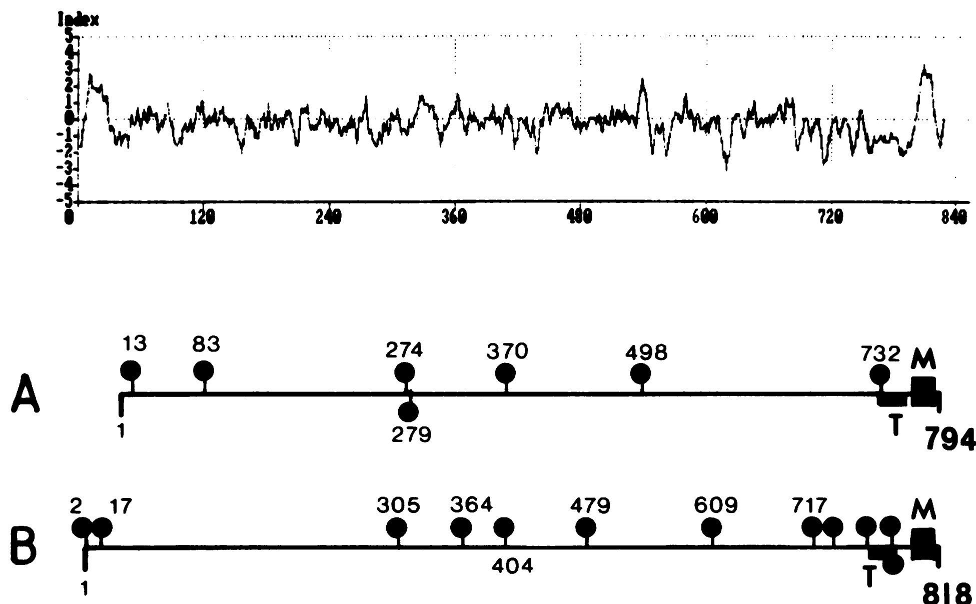

FIG. 4. Hydropathy analysis (5) of the *H. volcanii* S-layer protein (including signal peptide sequence) and schematic comparison of the mature S-layer proteins from *H. volcanii* (A) and *Halobacterium halobium* (B) with respect to location of glycosylation sites and membrane-binding domains. M, Membrane-binding domain; T, region of threonine clusters. Closed circles mark the positions of potential N-glycosylation sites.

amino acids exclusively composed of hydrophobic residues. Clusters of threonine residues adjacent to this putative membrane-binding domain are the most remarkable structural element in both proteins. Most if not all of these threonine residues are involved in O-glycosidic linkages to glucosylgalactose disaccharides in the *Halobacterium halobium* glycoprotein. Since the same type of covalently bound disaccharide is present in the *H. volcanii* protein, we assume by analogy that the clusters of threonine residues are also the sites of O glycosylation in *H. volcanii*. It was speculated (4)

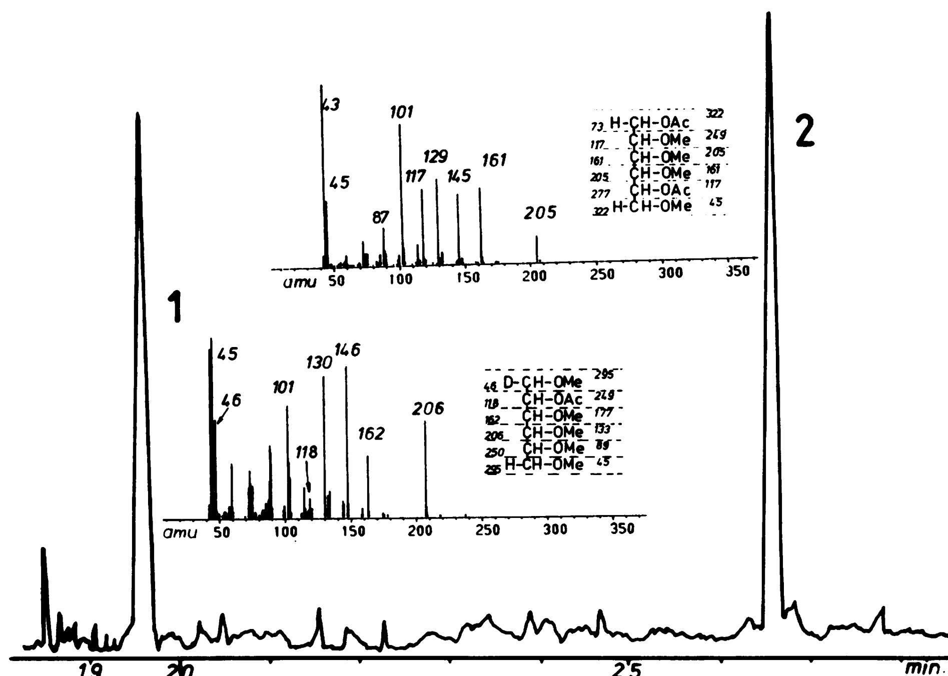

FIG. 5. GC-MS of products obtained by acid hydrolysis of reduced ($NaBD_4$) and permethylated saccharides β-eliminated from *H. volcanii* S-layer glycoprotein. Inserts show the mass spectra obtained from the material of the corresponding peaks. Peak 1, 2-*O*-Acetyl-1,3,4,5,6-penta-*O*-methylgalactitol (deuterium labeled at C-1); peak 2, 1,5-di-*O*-acetyl-2,3,4,6-tetra-*O*-methylglucitol. For details, see Materials and Methods.

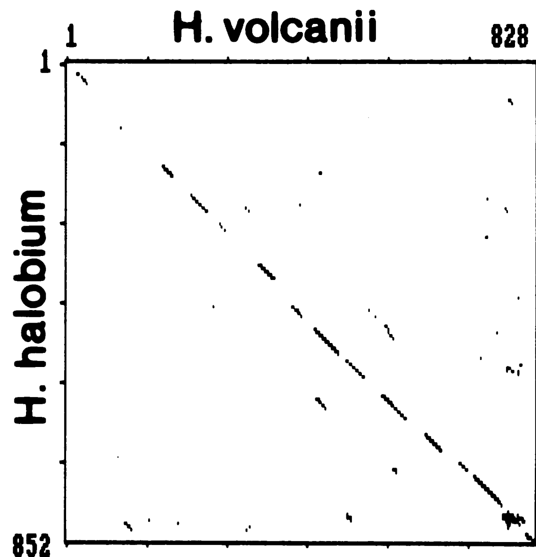

FIG. 6. Amino acid sequence homology between the *H. volcanii* and the *Halobacterium halobium* S-layer glycoproteins. The complete amino acid sequence (including the signal sequence) of the *H. volcanii* protein is represented on the horizontal axis; the corresponding sequence of the *Halobacterium halobium* protein is given on the vertical axis. The dot matrix plot was prepared with the PROSIS program (Pharmacia). Segments of 20 amino acid residues from the horizontal axis were compared with segments from the vertical axis, and a dot was placed whenever the number of matching amino acids was ≥10.

that this unusual structural element serves as a spacer between the membrane-binding domain and a more distant extracellular domain of the S-layer glycoprotein, thus creating an interspace which may be regarded as analogous to the periplasmic space of gram-negative eubacteria. In both proteins a number of potential N-glycosylation sites are scattered throughout the extracellularly located portions of the polypeptide chains, although fewer sites are recognized in the *H. volcanii* glycoprotein. Amino acid sequence data from isolated peptides indicate the existence of N-glycosidically linked saccharides at these sites in the *H. volcanii* glycoprotein. However, preliminary chemical data indicate major differences in the structure of N-glycosidically linked saccharides of the *H. volcanii* glycoprotein as compared with that of the *Halobacterium halobium* analog. We could demonstrate the presence of neither amino sugars nor covalently bound sulfate residues. This excludes the existence of the repeating unit pentasaccharide found at amino acid position 2 of the *Halobacterium halobium* glycoprotein as well as that of sulfated glucuronic acid-containing oligosaccharides typical for that glycoprotein (8, 15).

(ii) Figure 6 shows the result of a homology plot based on the Harr method of constructing dot matrix homology plots. The complete amino acid sequence of the glycoprotein from *H. volcanii*, including the N-terminal signal peptide, is represented on the horizontal axis and the corresponding *Halobacterium halobium* analog is shown on the vertical axis. Remarkably, stretches of nearly complete homology are interrupted by stretches of unrelated amino acid sequences. The degree of homology strikingly drops towards the N terminus, indicating the possibility of different architectures of the outermost portions of the two halobacterial S layers.

ACKNOWLEDGMENTS

We thank R. Deutzmann, University of Regensburg, for sequencing peptides.

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 43).

LITERATURE CITED

1. **Dubois, M., K. A. Gilles, J. K. Hamilton, and P. A. Rebers.** 1956. Colorimetric method for determination of sugar and related substances. Anal. Chem. **28:**350–356.
2. **Hakomori, S.** 1964. A rapid permethylation of glycolipid and polysaccharide catalyzed by methylsulfinylcarbanion in dimethyl sulfoxide. J. Biochem. **55:**205–209.
3. **Jannson, E., M. Kenne, M. Liedgren, B. Lindberg, and J. Loenngren.** 1976. A practical guide to the methylation analysis of carbohydrates. Chem. Commun. **8:**1–74.
4. **Kessel, M., I. Wildhaber, S. Cohen, and W. Baumeister.** 1988. Three-dimensional structure of the regular surface glycoprotein layer of *Halobacterium volcanii* from the Dead Sea. EMBO J. **7:**1549–1554.
5. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
6. **Lechner, J., and M. Sumper.** 1987. The primary structure of a procaryotic glycoprotein. J. Biol. Chem. **262:**9724–9729.
7. **Lechner, J., and F. Wieland.** 1989. Structure and biosynthesis of procaryotic glycoproteins. Annu. Rev. Biochem. **58:**173–194.
8. **Lechner, J., F. Wieland, and M. Sumper.** 1985. Biosynthesis of sulfated saccharides N-glycosidically linked to the protein via glucose. J. Biol. Chem. **260:**860–866.
9. **Lee, C. C., X. Wu, R. A. Gibbs, R. G. Cook, D. M. Muzny, and C. T. Caskey.** 1988. Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. Science **239:** 1288–1291.
10. **Lottspeich, F.** 1989. Microscale isocratic separation of phenylthiohydantoin amino acid derivatives. J. Chromatogr. **326:**321–327.
11. **Matteucci, M. D., and M. H. Caruthers.** 1981. Synthesis of deoxyoligonucleotides on a polymer support. J. Am. Chem. Soc. **103:**3185–3191.
12. **Mescher, M. F., and J. L. Strominger.** 1976. Purification and characterisation of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarium*. J. Biol. Chem. **251:** 2005–2014.
13. **Mevarech, M., and R. Werczberger.** 1985. Genetic transfer in *Halobacterium volcanii*. J. Bacteriol. **162:**461–462.
14. **Mort, A. J., and D. T. A. Lamport.** 1977. Anhydrous hydrogenfluoride deglycosylates glycoproteins. Anal. Biochem. **82:**289–309.
15. **Paul, G., F. Lottspeich, and F. Wieland.** 1986. Asparaginyl-N-acetylgalactosamine-linkage unit of halobacterial glycosaminoglycan. J. Biol. Chem. **261:**1020–1024.
16. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167:**391–409.
17. **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.
18. **Peters, J., M. Peters, F. Lottspeich, W. Schäfer, and W. Baumeister.** 1987. Nucleotide sequence analysis of the gene encoding the *Deinococcus radiodurans* surface protein, derived amino acid sequence, and complementary protein chemical studies. J. Bacteriol. **167:**5216–5223.
19. **Pisano, J. J.** 1972. Gas liquid chromatographie of amino acid derivatives. Methods Enzymol. **25:**27–44.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
21. **Sleytr, U. B., and P. Messner.** 1988. Crystalline surface layers on bacteria, p. 160–186. *In* U. B. Sleytr, P. Messner, D. Pum, and M. Sara (ed.), Crystalline bacterial cell surface layers. Springer-Verlag, Berlin.

22. **Sumper, M.** 1987. Halobacterial glycoprotein biosynthesis. Biochim. Biophys. Acta **906**:69–79.
23. **Tsuboi, A., R. Tabata, Y. Takahashi, H. Hashiba, S. Sasaki, H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1986. Characterization of the genes coding for two major cell proteins from protein-producing *Bacillus brevis* 47: complete nucleotide sequence of the outer wall protein. J. Bacteriol. **168**:365–373.
24. **Tsuboi, A., R. Uchihi, T. Adachi, T. Sasaki, S. Hayakawa, H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1988. Characterization of the genes for the hexagonally arranged surface layer proteins in protein-producing *Bacillus brevis* 47: complete nucleotide sequence of the middle wall protein gene. J. Bacteriol. **170**:935–945.
25. **Waeghe, T. J., A. G. Darwill, M. McNeil, and P. Albersheim.** 1983. Determination, by methylation analysis, of the glycosyl-linkage composition of microgram quantities of complex carbohydrates. Carbohydr. Res. **123**:281–304.