JOURNAL OF BACTERIOLOGY, Nov. 1989, p. 6307–6315
0021-9193/89/116307-09$02.00/0

Vol. 171, No. 11

S-Layer Protein Gene of *Acetogenium kivui*: Cloning and Expression in *Escherichia coli* and Determination of the Nucleotide Sequence

JÜRGEN PETERS,* MARIETTA PETERS, FRIEDRICH LOTTSPEICH, AND WOLFGANG BAUMEISTER

Max-Planck-Institut für Biochemie, 8033 Martinsried, Federal Republic of Germany

Received 25 January 1989/Accepted 23 August 1989

***Acetogenium kivui* is an anaerobically growing thermophilic bacterium with a gram-positive type of cell wall structure. The outer surface is covered with a hexagonally packed surface (S) layer. The gene coding for the S-layer polypeptide was cloned in *Escherichia coli* on two overlapping fragments by using the plasmid pUC18 as the vector. It was expressed under control of a cloned *Acetogenium* promoter or the *lacZ* gene. We determined the complete sequence of the structural gene. The mature polypeptide comprises 736 amino acids and is preceded by a typical procaryotic signal sequence of 26 amino acids. It is weakly acidic, weakly hydrophilic, and contains a relatively high proportion of hydroxyamino acids, including two clusters of serine and threonine residues. An N-terminal region of about 200 residues is homologous to the N-terminal part of the middle wall protein, one of the two S-layer proteins of *Bacillus brevis*, and there is also an internal homology within the N-terminal region of the *A. kivui* polypeptide.**

Two-dimensional regular arrays of proteins on the cell surface, termed S layers, have now been realized to be ubiquitously present in the eubacterial and archaebacterial kingdoms (see reference 2 for a review). The fact that evolutionary pressure has obviously maintained these major cell constituents suggests that they play vital roles in the interaction of the cell with its environment. Recent three-dimensional reconstructions from electron micrographs have shed some light on both common and diverse structural features (see reference 2 for a review). Except for a few instances, however, their functions have remained rather enigmatic, and three-dimensional-structure determinations at or close to atomic resolution, as well as the knowledge of primary structures, will be required to understand their structure, function, assembly, and evolution.

The amino acid sequences of four different S-layer polypeptides from three phylogenetically very distant organisms (20, 31, 38, 39) and one partial sequence (13) have recently been reported, but more primary structures are clearly needed to address open questions.

Acetogenium kivui is a thermophilic, anaerobically growing bacterium which has a gram-positive cell wall structure (21, 22) and which is covered with a hexagonal S layer (21). The determination of the structure of the S layer in projection has been reported (33) and the three-dimensional structure has been determined at a resolution of 1.8 nm (unpublished results; a computer-generated model of the *A. kivui* S layer is shown in reference 2). The particularly clear domain structure indicated by the three-dimensional reconstruction should be well suited for both controlled fragmentation with proteases as well as immunoelectron microscopical studies.

We cloned the structural gene for the S-layer polypeptide in *Escherichia coli* by using the plasmid pUC18 (46) as the vector and determined the nucleotide sequence. We also report on sequence similarities among S-layer proteins and discuss some evolutionary implications.

* Corresponding author.

MATERIALS AND METHODS

Bacterial strains. *A. kivui* was obtained from the German Collection of Microorganisms (DSM/2030) and cultivated as previously described (33).

Isolation of *A. kivui* S layer. The S layer was isolated and purified as previously described (33) and was washed with 0.2 M KCl to remove bound lysozyme.

Isolation of DNA. Chromosomal DNA was prepared as previously described (27). Plasmid DNA was prepared as described by Birnboim (5).

Immunological procedures. Immunization of rabbits with purified S layer and labeling of purified immunoglobulin G (IgG) with fluorescein isothiocyanate were performed as previously described (30).

Colony screening and Western blot (immunoblot) analysis. These procedures were also conducted as described in reference 30 and references therein. For sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis, the Midget Electrophoresis System (LKB Instruments, Inc., Rockville, Md.) was used. As an alternative detection method for Western blots, the streptavidin system (Amersham Corp., Braunschweig, Federal Republic of Germany) was used in the following manner. After reaction of the Western blots with anti-*A. kivui* S-layer IgG and washing, biotinylated anti-rabbit IgG was added by following the protocol given in the Amersham manual. The following procedure is a slight modification of an unpublished protocol developed by Georg Wiegand (Martinsried, Federal Republic of Germany). Streptavidin (0.5 mg) and biotinylated alkaline phosphatase (1.0 mg) were each dissolved in 1 ml of 50 mM phosphate buffer, pH 7.0, and stored at −20°C in small aliquots. A suitable amount of these solutions was mixed at a ratio of 1.1:1 by volume. The preformed complex was added to blots equilibrated with 20 mM Tris hydrochloride (pH 7.5)–0.155 M NaCl–5 g of bovine serum albumin per liter (TBS) at a dilution of 1:200. Total liquid volumes of 15 to 20 ml were used for 7- by 8-cm blots. After incubation with constant, gentle agitation for 30 min, the blots were washed four times with TBS containing 0.05% Triton X-100 and rinsed once with 100 mM Tris hydrochloride (pH 8.8)–5 mM $MgCl_2$. Staining was then performed in 20 ml of this buffer by

addition of 10 μl of Nitro Blue Tetrazolium (75 mg/ml in 70% dimethylformamide) and 60 μl of 5-bromo-4-chloro-3-indolyl phosphate (50 mg/ml in dimethylformamide) and constant agitation for 1 to 4 h until color development was essentially complete. All chemicals and proteins used in this procedure were purchased from Sigma Chemical Co., München, Federal Republic of Germany.

Southern genomic analysis and colony hybridization. Oligonucleotide probes for Southern blot analysis and dideoxy sequencing were synthesized by D. Oesterhelt by using the phosphoramidite method (28). The 0.65-kilobase (kb) *Sst*I-*Hin*dIII fragment (see Fig. 2) was isolated by preparative agarose gel electrophoresis and electroelution. End labeling of oligonucleotides was performed with [γ-^{32}P]ATP by using T4 polynucleotide kinase (27). The 0.65-kb fragment was labeled with [α-^{32}P]CTP by using the random-hexamer method (11). Southern genomic analysis was conducted as described previously (43), and colony screening was done as described by Grunstein and Hogness (17). Nitrocellulose filters (Schleicher & Schuell, Dassel, Federal Republic of Germany) were used throughout, since they gave superior results when compared with GeneScreen. Filters were exposed to Kodak X-Omat AR film, and Quanta III intensifying screens (E. I. duPont de Nemours & Co., Inc., Wilmington, Del.) were optionally used.

Cloning procedures. For expression cloning, chromosomal DNA of *A. kivui* was partially digested with restriction endonuclease *Sau*3A, size fractionated to 3.5 to 12 kb on a 0.7% agarose gel, electroeluted by using a Biotrap (Schleicher & Schuell) and recovered by precipitation. After ligation into *Bam*HI-cut and dephosphorylated pUC18 (46), *E. coli* JM109 was transformed to ampicillin resistance as described by Hanahan (19). Positive clones were identified by colony screening with fluorescein isothiocyanate-labeled antibodies as described above.

Exonuclease deletion. To construct an ordered subset of deletion derivatives of pMP221, this plasmid was cut with restriction endonucleases *Sph*I and *Bam*HI, subjected to an exonuclease III-VII deletion procedure (46), and transformed by the method of Hanahan (19). Clones were preselected by agarose gel electrophoresis of colony lysates and mapped by restriction with *Sst*I. In-frame deletion derivatives giving blue colonies in the color-forming assay (41) were preferred, since the intact nature of their reading frames could be conveniently monitored.

Nucleotide sequence analysis. Sequence analysis was performed with the dideoxy-chain termination method of Sanger et al. (34) by using the Sequenase kit (U.S. Biochemical Corp., Cleveland, Ohio) with [α-^{35}S]dATP (Amersham) and following the protocol supplied with this kit. By using both dITP and dGTP, all of the base positions could be identified. For other details see reference 31.

Amino acid analysis and amino acid sequence determination. For amino acid analysis of peptides, 0.1 to 0.5 μg of peptide material was hydrolyzed by gas-phase hydrolysis (40) by using a trifluoroacetic acid–6 M hydrochloric acid mixture (1:10 [vol/vol]) at 140°C for 2 h. The hydrolysate was derivatized by *o*-phthaldialdehyde–mercaptopropionic acid and subjected to reversed-phase high-pressure liquid chromatography. Amino acid sequence analysis was performed with a gas-phase sequenator (type 470A; Applied Biosystems) (25), and phenylthiohydantoin derivatives were analyzed with a high-pressure liquid chromatography system that separates all components isocratically (24). For amino acid analysis of the S-layer protein, 20 μg of protein was hydrolyzed with 5.7 N HCl at 110°C for 14 h, and amino acids were detected with ninhydrin on an automated amino acid analyzer (model 6300; Beckman Instruments, Inc., Fullerton, Calif.).

Polypeptide cleavage and peptide separation. Cyanogen bromide cleavage was performed in 70% formic acid for 4 h at room temperature by using a reagent-to-protein ratio of 5:1 by weight. The reaction was stopped by evaporation.

For cleavage with staphylococcal Glu-C protease (obtained from Boehringer Mannheim, Federal Republic of Germany), 4 mg of S layer was digested with 800 μg of protease for 72 h at room temperature in 50 mM Tris hydrochloride (pH 7.5)–0.4% sodium dodecyl sulfate (SDS). The S layer was denatured at 100°C for 20 min prior to digestion. SDS was removed by extraction with 5 volumes of isopentanol.

The peptide mixtures were dissolved in 70% formic acid and separated on a sieving column (TSK 2000 SW or TSK 3000 SW; LKB Instruments, Inc., or Pharmacia, Inc., Piscataway, N.J.). Suitable fractions of Glu-C peptides were further purified by reversed-phase high-pressure liquid chromatography as described previously (31).

Deglycosylation. Purified S-layer protein (100 μg) was vacuum dried from an aqueous suspension in the presence of 500 μg of SDS and treated with trifluoromethanesulfonic acid as described in reference 9. The samples were neutralized by addition of 10 volumes of 1 M sodium carbonate buffer, pH 9.2, at 0°C. After precipitation with 4 volumes of ethanol at 0°C and centrifugation at 13,600 × *g* for 3 min, the salt precipitate was dissolved in water and the protein pellet was dissolved in electrophoresis sample buffer.

Computer analysis. The computer programs for sequence similarity studies used in this study were available from the Protein Identification Resource (16).

RESULTS

Protein chemical studies. Preparations of the S layer of *A. kivui* contained one single polypeptide, as judged by amino acid sequence analysis. However, on SDS-polyacrylamide gels, the protein migrated in several more or less diffuse bands, the fastest-moving one corresponding to an approximate M_r of 82,000 (Fig. 1a). Variation of the temperature of solubilization with SDS between 40 and 100°C, denaturation with SDS and 8 M urea, or pretreatment with 88% formic acid did not significantly affect this band pattern. However, by peptide mapping with different proteases, these bands were indistinguishable and the amino acid analyses of electroeluted bands revealed no significant compositional differences (data not shown). After treatment of the S layer with trifluoromethanesulfonic acid, which is an established method for the peripheral degradation of O-linked carbohydrates (see reference 18), only one band with an approximate M_r of 86,000 appeared on SDS-polyacrylamide gels (Fig. 1a). Preliminary compositional analyses have revealed significant amounts of glucose, *N*-acetylglucosamine, and *N*-acetylgalactosamine. These data indicate that the S-layer polypeptide of *A. kivui* is heterogeneously glycosylated. More detailed investigations are beyond the scope of this paper.

Cloning of the S-layer gene in *E. coli*. Chromosomal *Sau*3A fragments were cloned as described in Materials and Methods. *E. coli* JM109 (46) was used as the cloning host. About 10,000 recombinant clones were subjected to an immunological screening procedure, with antibody directed against the purified S layer (see Materials and Methods). Of 16 initially positive clones, only one retained the expression of polypeptides cross-reactive with antibodies against the S layer after

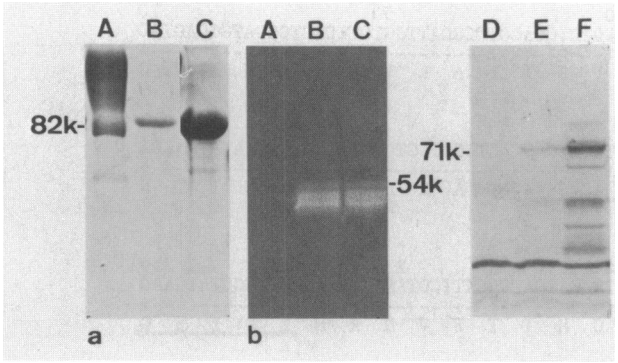


FIG. 1. SDS-polyacrylamide gel and Western blot patterns of the *A. kivui* S-layer polypeptide. (a) SDS-polyacrylamide gel electrophoresis of the purified S layer. The gel was stained with Coomassie blue. Numbers followed by k, apparent M_r (10^3). Lanes: A, S layer, 30 μg; B, deglycosylated S layer (see Materials and Methods), 3 μg; C, as in B, 30 μg. (b) Expression of the cloned *A. kivui* polypeptide in *E. coli*. Lanes A, B, and C were stained with anti-*A. Kivui* S-layer IgG and fluorescein isothiocyanate-labeled anti-rabbit IgG. Lanes D through F were stained by using a variation of the streptavidin system, using anti-*A. kivui* S-layer IgG as first antibody (see Materials and Methods). Total cell protein from 0.4 ml of culture (A_{550} = 1.0, 1-cm path length) was loaded on each lane. Lanes: A, *E. coli* JM109 (pUC18); B, MP111; C, MP111, grown with 1 mM IPTG; D, *E. coli* JM109(pUC18); E, MP211; F, MP211, grown with 1 mM IPTG. For the restriction maps of the corresponding plasmids, see Fig. 2.

recultivation. The plasmid, designated pMP100, contained a 6.2-kb insert. Western blot analysis with anti-*A. kivui* IgG (not shown) indicated that an antibody-reactive polypeptide with an M_r of at least 54,000 was encoded on this DNA fragment. We presumed that MP100 contained a truncated version of the S-layer protein gene. The size of the insert was reduced by partial digestion of pMP100 with *Eco*RI and religation. Subclones were screened by Western blot analysis. In the plasmid pMP111 (Fig. 2), two internal *Eco*RI fragments of pMP100 were deleted, but the Western blot polypeptide pattern (Fig. 1b) was unchanged. From the fact that expression was independent of isopropyl-β-D-thiogalactopyranoside (IPTG), we inferred that the cloned gene fragment was expressed from a cloned promoter and that the 3′ end of the gene rather than the 5′ end was missing. From restriction mapping, we further inferred that the gene was located at the end of the insert facing the *lacZ* gene. In order to identify a chromosomal fragment containing the entire gene, Southern genomic analysis was performed by using as probes the 0.65-kb *Sst*I-*Hin*dIII fragment from pMP111, which was believed to lie within the coding region of the gene (see Fig. 2), and two synthetic oligonucleotides derived from known amino acid sequences of the S-layer polypeptide (see the next paragraph). Since the major part of the gene was known to be contained in the 3.1-kb *Eco*RI-*Sau*3A fragment of pMP111 and since its orientation was also known, we concluded that a 5.7-kb *Eco*RI-*Kpn*I fragment hybridizing with all of these probes on Southern blots (data not shown) contained the entire gene.

Chromosomal DNA of *A. kivui* completely digested with these two restriction endonucleases was size fractionated by 4 to 7 kb by agarose gel electrophoresis and ligated into suitably cut pUC18. However, all attempts to clone the 5.7-kb fragment in *E. coli* were unsuccessful, as judged by a colony hybridization assay. In contrast to this, cloning of a 5.8-kb *Sst*I fragment was achieved by using a similar approach. By screening 500 clones, four clones that contained a 5.8-kb *Sst*I fragment in the same orientation as pMP200 (Fig. 2), as judged by restriction analysis, were identified. All of them hybridized with the 0.65-kb *Sst*I-*Hin*dIII fragment and a probe derived from a region of the protein near the N terminus (later shown to correspond to positions 298 to 314 of the nucleotide sequence; see Fig. 3). However, a probe derived from the N terminus itself (nucleotide positions 91 to 107) did not hybridize with any of these clones. These data

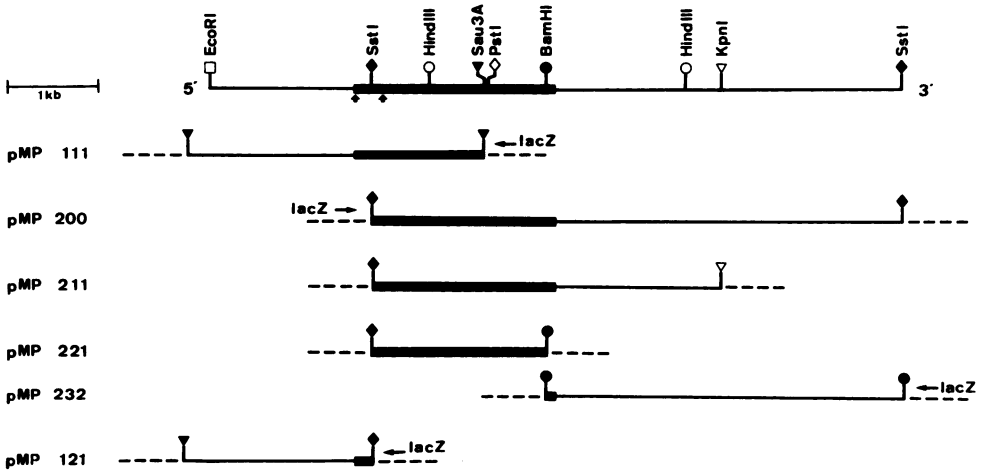
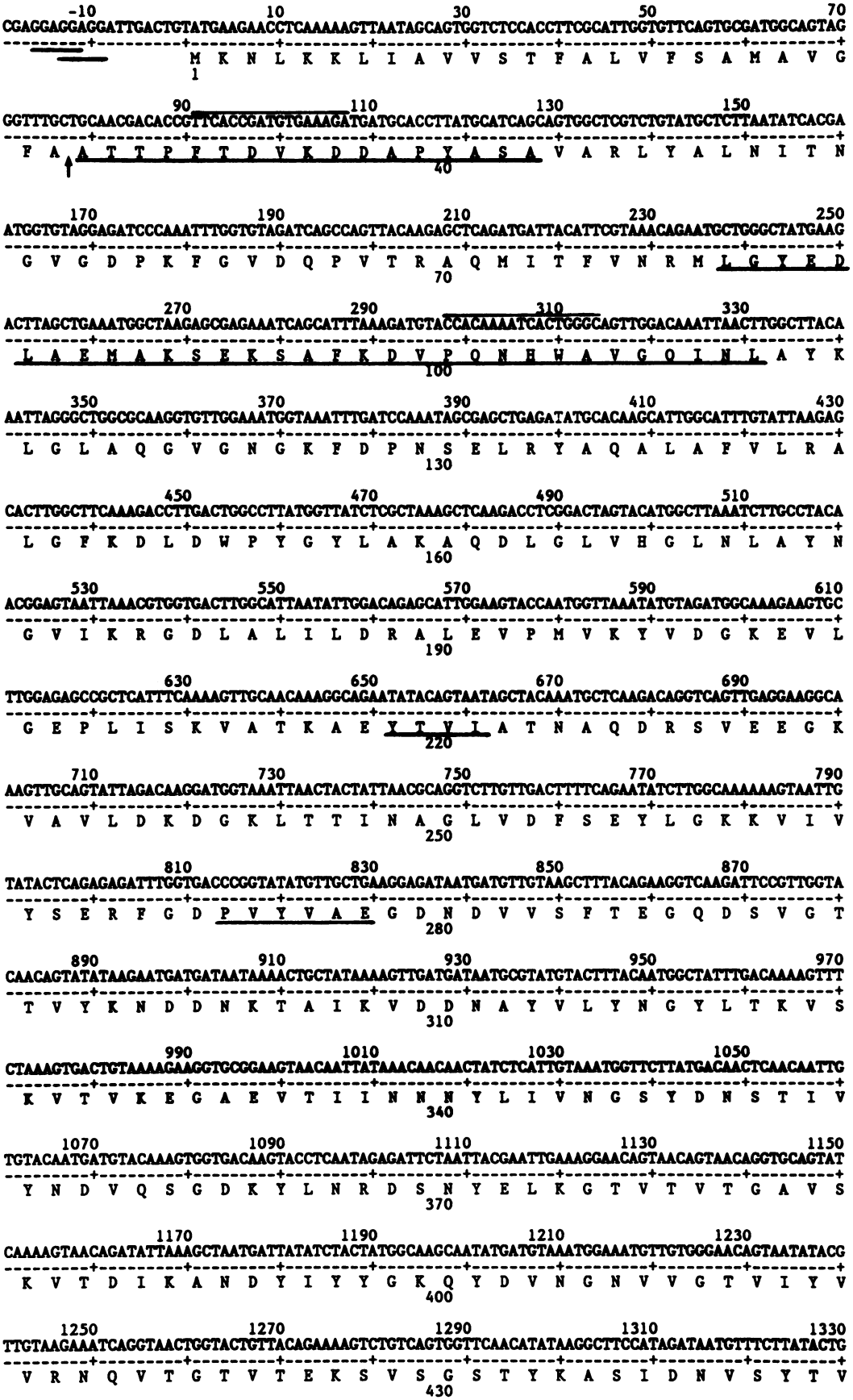


FIG. 2. Restriction maps of the *A. kivui* chromosomal DNA region containing the S-layer protein gene and of the plasmids used in this study. The uppermost drawing is a composite map obtained by restriction analysis of several overlapping subclone inserts; all restriction sites determined are shown. The positions at which oligonucleotide probes hybridized are marked with arrows. The direction of transcription of the *lacZ* gene is also indicated. In the plasmid maps below, only the terminal sites of the insert DNA are indicated. In each case, the insert is located in the multiple cloning site of pUC18 (represented by dashed lines). The *Sau*3A site of the 5′ end of pMP111 results from the deletion of two internal *Eco*RI fragments from the original *Sau*3A fragment in pMP100 (not shown). The solid bar represents the S-layer gene coding region.


-10 10 30 50 70
CGAGGAGGAGGATTGACTGTATGAAGAACCTCAAAAAGTTAATAGCAGTGGTCTCCACCTTCGCATTGGTGTTCAGTGCGATGGCAGTAG
M K N L K K L I A V V S T F A L V F S A M A V G
1
90 110 130 150
GGTTTGCTGCAACGACACCGTTCACCGATGTGAAAGATGATGCACCTTATGCATCAGCAGTGGCTCGTCTGTATGCTCTTAATATCACGA
F A A T T P F T D V K D D A P Y A S A V A R L Y A L N I T N
40
170 190 210 230 250
ATGGTGTAGGAGATCCCAAATTTGGTGTAGATCAGCCAGTTACAAGAGCTCAGATGATTACATTCGTAAACAGAATGCTGGGCTATGAAG
G V G D P K F G V D Q P V T R A Q M I T F V N R M L G Y E D
70
270 290 310 330
ACTTAGCTGAAATGGCTAAGAGCGAGAAATCAGCATTTAAAGATGTACCACAAAATCACTGGGCAGTTGGACAAATTAACTTGGCTTACA
L A E M A K S E K S A F K D V P Q N H W A V G Q I N L A Y K
100
350 370 390 410 430
AATTACGGCTGGCGCAAGGTGTTGGAAATGGTAAATTTGATCCAAATAGCGAGCTGAGATATGCACAAGCATTGGCATTTGTATTAAGAG
L G L A Q G V G N G K F D P N S E L R Y A Q A L A F V L R A
130
450 470 490 510
CACTTGGCTTCAAAGACCTTGACTGGCCTTATGGTTATCTCGCTAAAGCTCAAGACCTCGGACTAGTACATGGCTTAAATCTTGCCTACA
L G F K D L D W P Y G Y L A K A Q D L G L V H G L N L A Y N
160
530 550 570 590 610
ACGGAGTAATTAAACGTGGTGACTTGGCATTAATATTGGACAGAGCATTGGAAGTACCAATGGTTAAATATGTAGATGGCAAAGAAGTGC
G V I K R G D L A L I L D R A L E V P M V K Y V D G K E V L
190
630 650 670 690
TTGGAGAGCCGCTCATTTCAAAAGTTGCAACAAAGGCAGAATATACAGTAATAGCTACAAATGCTCAAGACAGGTCAGTTGAGGAAGGCA
G E P L I S K V A T K A E Y T V I A T N A Q D R S V E E G K
220
710 730 750 770 790
AAGTTGCAGTATTAGACAAGGATGGTAAATTAACTACTATTAACGCAGGTCTTGTTGACTTTTCAGAATATCTTGGCAAAAAAGTAATTG
V A V L D K D G K L T T I N A G L V D F S E Y L G K K V I V
250
810 830 850 870
TATACTCAGAGAGATTTGGTGACCCGGTATATGTTGCTGAAGGAGATAATGATGTTGTAAGCTTTACAGAAGGTCAAGATTCCGTTGGTA
Y S E R F G D P V Y V A E G D N D V V S F T E G Q D S V G T
280
890 910 930 950 970
CAACAGTATATAAGAATGATGATAATAAAACTGCTATAAAAGTTGATGATAATGCGTATGTACTTTACAATGGCTATTTGACAAAAGTTT
T V Y K N D D N K T A I K V D D N A Y V L Y N G Y L T K V S
310
990 1010 1030 1050
CTAAAGTGACTGTAAAAGAAGGTGCGGAAGTAACAATTATAAACAACAACTATCTCATTGTAAATGGTTCTTATGACAACTCAACAATTG
K V T V K E G A E V T I I N N N Y L I V N G S Y D N S T I V
340
1070 1090 1110 1130 1150
TGTACAATGATGTACAAAGTGGTGACAAGTACCTCAATAGAGATTCTAATTACGAATTGAAAGGAACAGTAACAGTAACAGGTGCAGTAT
Y N D V Q S G D K Y L N R D S N Y E L K G T V T V T G A V S
370
1170 1190 1210 1230
CAAAAGTAACAGATATTAAAGCTAATGATTATATCTACTATGGCAAGCAATATGATGTAAATGGAAATGTTGTGGGAACAGTAATATACG
K V T D I K A N D Y I Y Y G K Q Y D V N G N V V G T V I Y V
400
1250 1270 1290 1310 1330
TTGTAAGAAATCAGGTAACTGGTACTGTTACAGAAAAGTCTGTCAGTGGTTCAACATATAAGGCTTCCATAGATAATGTTTCTTATACTG
V R N Q V T G T V T E K S V S G S T Y K A S I D N V S Y T V
430

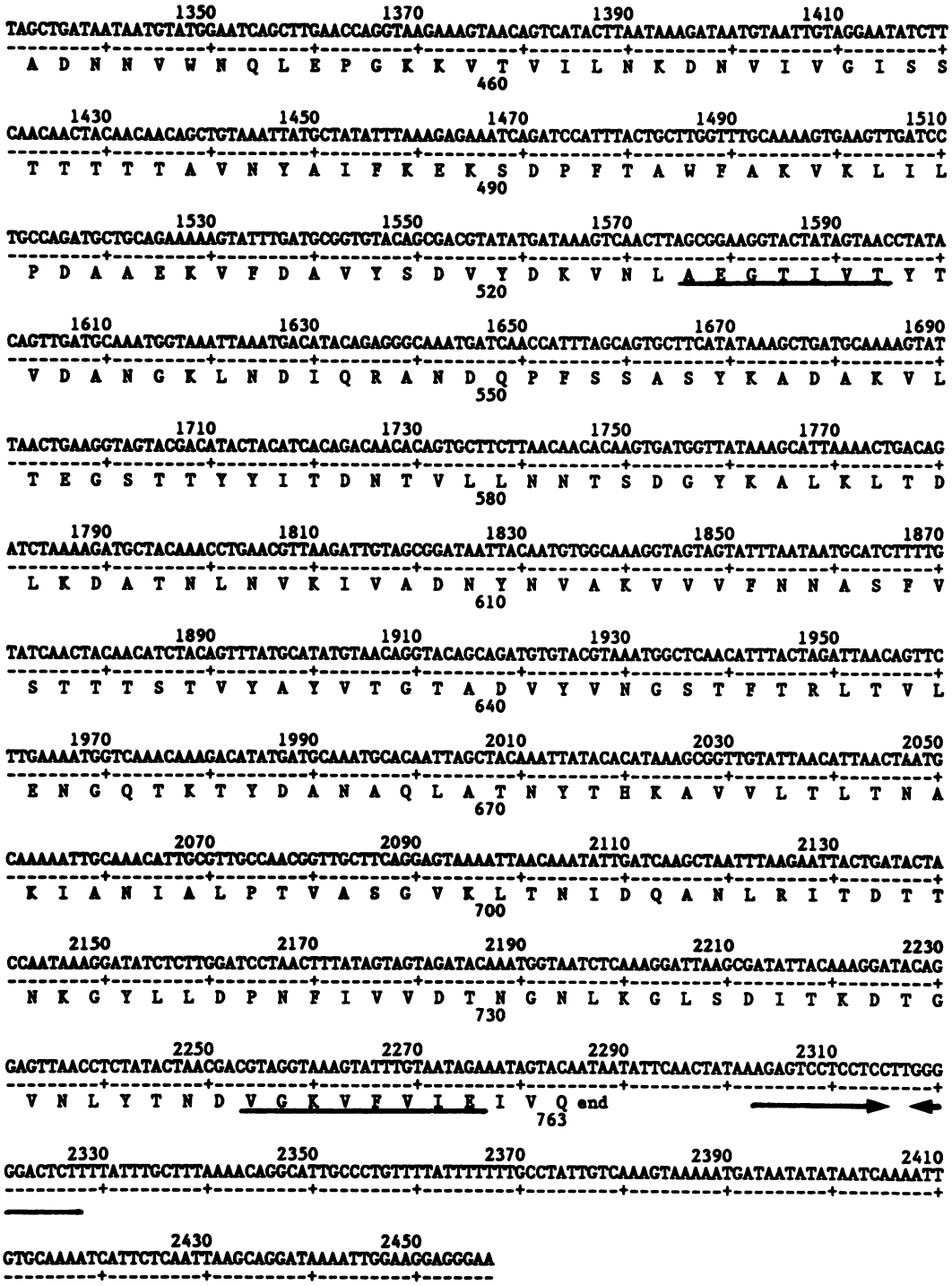


FIG. 3. Nucleotide sequence of the gene for the *A. kivui* S-layer polypeptide. A tandem of Shine-Dalgarno sequences is underlined. The signal peptidase cleavage site as derived from N-terminal amino acid sequence analysis of the mature S-layer polypeptide is marked with an arrow. The amino acid sequence coded in the single open reading frame is also displayed. Partial amino acid sequences determined by Edman degradation of the S-layer protein and of peptides obtained after cleavage with cyanogen bromide, staphylococcal protease, and pepsin are underlined. Partial nucleotide sequences corresponding to oligonucleotide probes are also marked with lines. The putative terminator palindrome (see text) is indicated with arrows.

enabled us to localize the position of the structural gene rather accurately.

Plasmids pMP211 and pMP221 were derived from pMP200 by deletion of the 1.9-kb *Kpn*I fragment or the 3.9-kb *Bam*HI fragment, respectively. The 3.9-kb *Bam*HI fragment was cloned into the *Bam*HI site of pUC18 to give pMP232 (see Fig. 2; the second *Kpn*I and *Bam*HI sites were located in the multiple cloning region of the vector.)

MP200, MP211, and MP221 were stable throughout more than 10 cycles of cultivation. As verified later by DNA

TABLE 1. Amino acid composition of the *A. kivui* S layer

Amino acid	Composition (mol%) from:	
	Amino acid analysis	DNA sequence
Cys	0	0
Asx	16.1	16.4
Thr	9.0	9.6
Ser	4.3	4.9
Glx	6.5	6.0
Pro	2.5	2.2
Gly	7.5	7.2
Ala	8.8	8.7
Val	12.1	12.5
Met	0.5	0.5
Ile	4.6	4.7
Leu	8.2	7.9
Tyr	6.1	6.1
Phe	2.9	2.7
His	0.5	0.5
Lys	7.9	7.8
Arg	1.9	1.9
Trp	ND[a]	0.5

[a] ND, Not determined. For the calculation, this value was set to 0.5.

sequencing, pMP221 contained an in-frame fusion of the S-layer gene with the *lacZ* gene. The plasmids pMP200 and pMP211 contained N-terminal *lacZ* fusions. Indeed, Fig. 1b demonstrates that the level of polypeptide produced was strongly enhanced by IPTG, although it was still below expectations based on *lacZ* promoter strength. MP221 gave blue colonies in the color-forming assay (41). This is plausible, since both the 3′ end and the 5′ end of the structural gene are missing in pMP221, but it should be noted that the size of the insert is as much as 1.95 kb. The gel pattern in Fig. 1b indicates that the S-layer fragment encoded in pMP211, which has a calculated M_r of 75,666, is subject to partial proteolytic digestion in *E. coli*. The approximate M_r of the major polypeptide observed is 71,000. The susceptibility of even the native protein to proteolytic degradation was also demonstrated in vitro. Treatment of the native S layer with trypsin resulted in the disintegration of the S layer, and the approximate M_r of the polypeptide was reduced to 72,000. When purified S layer was incubated with ultrasonicated *E. coli* JM109 cells, the 82,000-M_r band vanished and a 71,000-M_r band appeared as the major digestion product, as detected by Western blotting (data not shown).

Nucleotide sequence analysis. The bulk of the structural gene, which is present in pMP221, was sequenced from an ordered set of deletion derivatives generated with exonucleases III and VII (46). As mentioned above, this plasmid caused the formation of blue clones in the color-forming assay for *lacZ* fusion proteins, and this was also true for many of its deletion derivatives. The occasional formation of frameshift point mutations, resulting in white colonies, could thus be conveniently monitored. The sequence of the 5′ end was obtained from pMP121, and that of the 3′ end was obtained from pMP232. Sequences overlapping the *Sst*I or *Bam*HI site were obtained by using sequence-derived oligonucleotide primers with pMP100 or pMP211, respectively. The counterstrand was also sequenced by using synthetic oligonucleotides. The nucleotide sequence is shown in Fig. 3. A tandem of overlapping Shine-Dalgarno sequences (36) precedes a single open reading frame of 2,289 nucleotides. The other two forward frames contain 76 or 83 stop codons, respectively. From the N-terminal sequence of the mature protein and from known features of presequences (42, 44), we conclude that the S-layer polypeptide is synthesized as a precursor, with a 26-amino-acid extension. The mature polypeptide comprises 736 amino acids, with a calculated M_r of 80,046. The distance between the Shine-Dalgarno sequence and the putative transcriptional start site is within the optimal range (see references 35 and 37), and, therefore, it is unlikely that methionine 21 is used as an alternative start site.

The 3′ end is followed by a strong palindrome (p = 5.75, s = 15) which was identified by the TERMINATOR algorithm (6). This and the immediately ensuing very deoxyribosylthymine-rich stretch are typical of factor-independent transcription termination signals (6).

The codon usage of the structural gene is extremely biased towards A-T. The overall A-T content is 64%, whereas that of codon position 3 is 76%. This significantly nonrandom codon usage favored a statistical analysis of the coding probability (12) which strongly indicated that the region downstream of the 3′ end of the gene was noncoding.

The amino acid analysis of the nucleotide sequence-derived polypeptide is in excellent agreement with the composition of the purified S layer, as shown in Table 1.

```
  1' MKNLKKLIAVVSTFALVFSAMAVGFAA--TTPFTDVKDDAPYASAVARLYALNITNGVGD
       .::... : . ::.... ...::: ... :. : :: ...: :: ::....: :.
  1"     MKKVVNSVLASALALTVAPMAFAAEEAATTTAPKMDADMEKTVKRLEALGLVAGYGN

 59' PKFGVDQPVTRAQMITFVNRMLGYEDLAEMAKSEKSAFKDVPQNHWAVGQINLAYKLGLA
      ..:::...:::.. :.: :  : :. :..:. ... ..::  ..: .: .:.:   ...
 58" GEYGVDKTITRAEFATLVVRARGLEQGAKLAQFSNT-YTDVKSTDWFAGFVNVASGEEIV

119' QGVGNGKFDPNSELRYAQALAFVLRALGFK---DLDWPYGYLAKAQDLGLVHGLNLAYNG
     .:  . .:.:.... ::.:.....::::..   .  :: . ..:: .:....... . :.
117" KGFPDKSFKPQNQVTYAEAVTMIVRALGYEPSVKGVWPNSMISKASELNIARSITTPNNA

176' VIKRGDLALILDRALEVPMVKYVDG
     .. :::.  .::.:: : ... :.
177" AT-RGDIFKMLDNALRVDLMEQVEF
```

FIG. 4. Intersequence comparison of the *A. kivui* S-layer polypeptide and the MWP of *B. brevis* (38). The single primes denote positions in the *A. kivui* polypeptide and double primes refer to the MWP polypeptide. Identical matches are marked with double dots; conservative replacements (1) are marked with single dots. The displayed alignment was obtained with the FASTP program (23).

```
29                                                                  85
 TPFTDVKDDAPYASAVARLYALNITNGVGDPKFGVDQPVTRAQMITFVNRMLGYEDL
 ..: :: ..    . .  .: : :..:::. ::. .  .  :: ..:: : ::. ::
 SAFKDVPQNHWAVGQINLAYKLGLAQGVGNGKFDPNSELRYAQALAFVLRALGFKDL
94                                                                 150
```

FIG. 5. Intrasequence comparison of the N-terminal region of the *A. kivui* S-layer polypeptide. The region of strongest similarity as judged by the RELATE program (1) is displayed. Sequence positions are numbered as in Fig. 3.

DISCUSSION

In this study we report the heterologous expression of the *A. kivui* S-layer polypeptide in *E. coli*. The gene and its product are not so toxic as to prevent stable expression of major gene fragments containing either the 5′ end or the 3′ end of the gene in *E. coli*. However, the failure to clone the entire gene for the *A. kivui* S layer directly from chromosomal DNA fragments leads one to suspect that the complete gene or its product may be toxic to the host cell. This view is also corroborated by the fact that N-terminal *lacZ* fusion proteins are produced in rather small amounts in the presence of IPTG.

With the advent of new nucleotide sequences, certain common structural features of S-layer polypeptides begin to emerge. All of the five S-layer polypeptides studied thus far, including that of *A. kivui*, contain an excess of acidic residues, although the *A. kivui* S-layer protein, like the hexagonally packed intermediate (HPI)-layer protein of *Deinococcus radiodurans* (31), is only weakly acidic (total charge at neutral pH = −13). As in the previously reported sequence of the HPI-layer polypeptide of *D. radiodurans* (31), we note clusters of serine-threonine residues (positions 473 to 479 and positions 626 to 631 in Fig. 3). A very threonine-rich sequence is also found in the S-layer polypeptide of *Halobacterium halobium* (20). Although these similarities do not represent statistically significant structural homologies as scored with the RELATE algorithm (1), these sequences are quite unique in procaryotic proteins. As in the HPI-layer polypeptide, the two serine-threonine clusters are flanked by moderately hydrophobic sequences. The structural significance of these serine-threonine-rich regions is not clear. In some eucaryotic proteins, such as the salivary glue protein of the fruit fly (14) or the contact site A protein of *Dyctostelium discoideum* (29), serine-threonine-rich sequences seem to be related to adhesion phenomena, and this might also apply to bacterial S layers. The distribution of positive and negative charges is not balanced throughout the polypeptide. In particular, the region from positions 190 to 380 contains a net negative charge of −15, whereas the region from positions 600 to 763 (C terminus) is positively charged (+3).

The prediction of secondary structure, as described by Chou and Fasman (8) and Garnier, Osguthorpe, and Robson (15), gave 33% β structure, 12% α helix, and 8% turns, or 41, 20, and 15%, respectively. The predicted values for β structure content are in good agreement with the value of approximately 35% determined by infrared spectroscopy (2). Spectroscopic measurements (infrared or circular dichroism) of the S layers or component proteins of *A. kivui* (2), *Sporosarcina ureae* (10), *Azotobacter vinelandii* (4), *Aeromonas salmonicida* (32), and the HPI layer of *D. radiodurans* (3) have yielded β structure contents around 30%. This might therefore be regarded as a structural characteristic of S layers.

Common architectural principles have also been identified on the tertiary-structure level by using three-dimensional electron microscopy. Thus, conspicuous structural similarities were found between the inner surfaces of the S-layers of *A. kivui*, *D. radiodurans*, and *Clostridium thermohydrosulfuricum* (2). It would be interesting to correlate such structural similarities, which seem to reflect common building principles, with primary structure similarities. However, by using the RELATE algorithm (1), we did not identify any statistically significant local-sequence similarity between the S-layer polypeptides of *D. radiodurans* and *A. kivui*. Recently three-dimensional structures of the *A. kivui* S layer (unpublished results; model shown in reference 2) and the *Bacillus brevis* middle wall protein (MWP) layer (Tsuboi, Engelhard, Santarius, Tsukagoshi, Udaka, and Baumeister, J. Ultrastruct. Mol. Struct. Res., in press) have become available. The reconstructions reveal a strikingly similar overall organization (both belong to the M_6 C_3 class [2]) and domain structure. An alignment of the two amino acid sequences with the FASTP program (23, 45) showed that the N-terminal regions are highly similar (see Fig. 4). The corresponding alignment score is 30.7 standard deviations (SD) above the random mean, as determined with the RDF program (23). This value may be taken to reflect a definite sequence homology (23). These are only four gaps of one to three amino acids, and the displacement of both sequences in the alignment is never higher than three residues. The rate of identity in a 201-amino-acid overlap is 31%, and the additional similarity, defined as the fraction of conservative replacements (1), is 43%. A dot matrix analysis (26) revealed an internal sequence homology within the N-terminal region of the *A. kivui* S-layer polypeptide. A statistical evaluation with the RELATE program yielded a SD value of 10.8 for the alignment shown in Fig. 5. This observed tandem arrangement of homologous sequences probably originated in an intragenic duplication. As might be expected, an internal sequence similarity was also observed in the MWP, but the level of significance was much lower (data not shown). The homology between the two polypeptides was found to be confined to the region shown in Fig. 4. However, by using the RELATE program, we found an additional region of much lower similarity between positions 527 through 640 in the *A. kivui* S-layer polypeptide and 646 through 759 in the MWP. Although the corresponding SD value is only 3.3, this result appears interesting in the light of the fact that similarities were also found between positions 758 through 797 in the MWP and positions 797 through 836 in the outer wall protein of *B. brevis* (39) (SD = 2.6.) and between positions 814 through 889 of the HPI-layer protein and positions 767 through 843 of the outer wall protein (SD = 5.1 [31]).

In conclusion, similarities of different qualities were detected between the C-terminal regions of all four eucaryotic S-layer polypeptides sequenced thus far. However, we do not yet understand the significance of these data. More amino acid sequences of S layers are clearly required to draw conclusions on phylogenetic relationships. In particular, the knowledge of S-layer protein sequences from organisms belonging to the same phylogenetic division as *A. kivui* and *B. brevis* should help us to discern whether the N-terminal sequence homology in Fig. 4 reflects an interspecies

gene transfer event or whether strong evolutionary pressure has conserved a region or domain in two related proteins. Furthermore, the identification of homologous amino acid sequences in the three-dimensional structures of S layers by topographical methods such as limited proteolysis or immunoelectron microscopy would bring us one step further in the understanding of S-layer architecture.

We are currently investigating the glycosylation of the S-layer protein of *A. kivui*. These studies are potentially of great interest, since there is no unequivocal proof so far of the existence of eubacterial glycoproteins.

ACKNOWLEDGMENTS

We thank F. Karrenberg for collaboration in the early stages of this work. Furthermore, we are indebted to D. Oesterhelt for synthesizing oligonucleotides and to J. Plambeck for performing amino acid analyses. We also thank M. Dûchene and G. Wiegand for stimulating discussions and B. Phipps for critically reading the manuscript.

This work was supported by grant Ba618/6-2 from the Deutsche Forschungsgemeinschaft.

LITERATURE CITED

1. **Barker, W. C., and M. O. Dayhoff.** 1972. Detecting distant relationships: computer methods and results, p. 101–110. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, D.C.
2. **Baumeister, W., and H. Engelhardt.** 1987. Three-dimensional structures of bacterial S-layers, p. 109–154. *In* J. R. Harris and R. W. Horne (ed.), Electron microscopy of proteins, vol. 6. Membranous structures. Academic Press, Inc. (London), Ltd., London.
3. **Baumeister, W., F. Karrenberg, R. Rachel, A. Engel, B. Ten Heggeler, and W. O. Saxton.** 1982. The major cell envelope protein of *Micrococcus radiodurans* R1. Structural and chemical characterization. Eur. J. Biochem. **125:**535–544.
4. **Bingle, W. H., J. L. Doran, and W. J. Page.** 1986. Characterization of the surface layer protein from Azotobacter vinelandii. Can. J. Microbiol. **32:**112–120.
5. **Birnboim, H. C.** 1983. A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol. **100:**243–255.
6. **Brendel, V., and E. N. Tifonov.** 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.
7. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4:**165–170.
8. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. **47:**46–148.
9. **Edge, A. S. B., C. R. Faltynek, L. Hof, L. E. Reichert, Jr., and P. Weber.** 1981. Deglycosylation of proteins by trifluoromethanesulfonic acid. Anal. Biochem. **118:**131–137.
10. **Engelhardt, H., W. O. Saxton, and W. Baumeister.** 1986. Three-dimensional structure of the tetragonal surface of *Sporosarcina ureae*. J. Bacteriol. **168:**309–317.
11. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
12. **Fickett, J. W.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10:**5303–5318.
13. **Fisher, J. A., J. Smit, and N. Agabian.** 1988. Transcriptional analysis of the major surface array gene of *Caulobacter crescentus*. J. Bacteriol. **170:**4706–4713.
14. **Garfinkel, M. D., R. E. Pruitt, and E. M. Meyerowitz.** 1983. DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68 C Glue gene cluster. J. Mol. Biol. **168:**765–789.
15. **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.
16. **George, D. G., W. C. Barker, and L. T. Hunt.** 1986. The protein identification resource (PIR). Nucleic Acids Res. **14:**11–15.
17. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72:**3961–3968.
18. **Hakimuddin, T. S., and P. Bahl.** 1987. A chemical method for the deglycosylation of proteins. Arch. Biochem. Biophys. **259:**52–57.
19. **Hanahan, D.** 1983. Studies on transcription of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–580.
20. **Lechner, J., and M. Sumper.** 1987. The primary structure of a procaryotic glycoprotein. J. Biol. Chem. **262:**9724–9729.
21. **Leigh, J. A., F. Mayer, and R. S. Wolfe.** 1981. *Acetogenium kivui*, a new thermophilic hydrogen-oxidizing, acetogenic bacterium. Arch. Microbiol. **129:**275–280.
22. **Leigh, J. A., and R. S. Wolfe.** 1983. *Acetogenium kivui* gen. nov., a thermophilic acetogenic bacterium. Int. J. Syst. Bacteriol. **33:**886.
23. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.
24. **Lottspeich, F.** 1985. Microscale isocratic separation of phenylthiohydantoin amino acid derivatives. J. Chromatogr. **326:**321–327.
25. **Lottspeich, F., J. Kellermann, A. Henschen, G. Rauth, and Müller-Esterl.** 1984. Human low-molecular mass kininogen. Eur. J. Biochem. **142:**227–232.
26. **Maizel, J. V., Jr., and R. P. Lenk.** 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA **78:**7665–7669.
27. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 122. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Matteucci, M. D., and M. H. Caruthers.** 1981. Synthesis of deoxyoligonucleotides on a polymer support. J. Am. Chem. Soc. **103:**3185–3191.
29. **Noegel, A., G. Gerisch, J. Stadler, and M. Westphal.** 1986. Complete sequence and transcript regulation of a cell adhesion protein from aggregating *Dictyostelium* cells. EMBO J **5:**1473–1476.
30. **Peters, J., and W. Baumeister.** 1986. Molecular cloning, expression, and characterization of the gene for the surface (HPI)-layer protein of *Deinococcus radiodurans* in *Escherichia coli*. J. Bacteriol. **167:**1048–1054.
31. **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. **169:**5216–5223.
32. **Phipps, B. M., T. J. Trust, E. E. Ishiguro, and W. W. Kay.** 1983. Purification and characterization of the cell surface virulent A protein from *Aeromonas salmonicida*. Biochemistry **22:**2934–2939.
33. **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.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
35. **Scherer, G. F. E., M. D. Walkinshaw, S. Arnott, and D. J. Morré.** 1980. The ribosome binding sites recognized by *E. coli* ribosomes have regions with signal character in both the leader protein coding segments. Nucleic Acids Res. **8:**3895–3907.
36. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16 S ribosomal RNA: complementary to nonsense triplets and ribosomal binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
37. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10:**2971–2996.
38. **Tsuboi, A., R. Uchihi, T. Adachi, T. Sasaki, S. Hayakawa, H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1988. Characteriza-

tion 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.

39. **Tsuboi, A., R. Uchihi, 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 wall proteins from protein-producing *Bacillus brevis* 47: complete nucleotide sequence of the outer wall protein. J. Bacteriol. **168**:365–373.
40. **Tsugita, A., T. Uchida, H. W. Mewes, and T. Ataka.** 1987. A rapid vapor-phase acid (hydrochloric acid and trifluoroacetic acid) hydrolysis of peptide and protein. J. Biochem. **102**: 1593–1597.
41. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutangenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
42. **Von Heijne, G.** 1985. Signal sequences: the limits of variation. J. Mol. Biol. **184**:99–105.
43. **Wallace, R. B., M. J. Johnson, T. Hirose, T. Miyake, E. H. Kawashima, and K. Itakura.** 1981. The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit β-globin DNA. Nucleic Acids Res. **9**:879–894.
44. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5159.
45. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.
46. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.