JOURNAL OF BACTERIOLOGY, Nov. 1992, p. 7419–7427
0021-9193/92/227419-09$02.00/0

Vol. 174, No. 22

S-Layer Protein Gene of *Lactobacillus brevis*: Cloning by Polymerase Chain Reaction and Determination of the Nucleotide Sequence

GABRIELE VIDGRÉN,[1]† ILKKA PALVA,[1]‡ RAIMO PAKKANEN,[2] KARI LOUNATMAA,[3,4] AND AIRI PALVA[1]‡*

Institute of Biotechnology[1] and Department of Electron Microscopy,[3] University of Helsinki, and Academy of Finland,[4] Helsinki, and Valio Bioproducts, Ltd., Turku Technology Centre, Biocity, Turku,[2] Finland

Received 27 May 1992/Accepted 4 September 1992

The surface (S)-layer protein of *Lactobacillus brevis* was isolated, purified, and characterized. The S-layer protein is the major protein of the cell, with an apparent molecular mass of 46 kDa in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Immunogold electron microscopy with polyclonal antiserum against the isolated 46-kDa protein was used to confirm the surface location of this protein. N-terminal amino acid sequences of the intact 46-kDa protein and its tryptic peptides were determined. The gene of the S-layer protein was amplified from the genome of *L. brevis* by polymerase chain reaction with oligonucleotides, synthesized according to the N-terminal amino acid sequences, as primers. The polymerase chain reaction fragments containing the entire S-layer gene and its regulatory regions were sequenced. Nucleic acid sequence analysis revealed one open reading frame with a capacity to encode a protein of 48,159 Da. From the regulatory region of the gene, two subsequent promoters and a ribosome binding site, showing typical features of prokaryotic consensus sequences, were found. The coding region contained a characteristic gram-positive-type signal peptide of 30 amino acids. Removal of the signal peptide results in a polypeptide of 435 amino acids, which is in excellent agreement with the size of the S-layer protein determined by SDS-PAGE. The size and the 5′ end analyses of the S-layer transcripts confirmed the monocistronic nature of the S-layer operon and the functionality of the two promoters found.

Regular crystalline arrays of proteinaceous subunits forming surface layers (S layers) have been commonly found in bacteria, and their morphological and chemical characteristics and assembly have been extensively studied (3, 31). In most cases, the S layer consists of a single protein or glycoprotein subunit ranging from 40 to 200 kDa in size, and the subunits are arranged on the cell surface with repeating hexagonal, tetragonal, or linear symmetry (31). In some microbes, such as *Bacillus brevis*, more than one S-layer protein has been found (1, 39). Although the functional significance of S-layer proteins is not yet fully elucidated, the S layers are assumed to have an important role in nature, because a substantial part of the synthetic capacity of the cell is used for their production (33). The following functions have been shown or presumed for S layers: (i) protective barrier against environmental hazards, (ii) control of the transfer of nutrients and metabolites, (iii) promoter for cell adhesion and surface recognition, and (iv) maintenance of cell shape and envelope rigidity (32). The knowledge of the primary structure of S-layer proteins is still limited, and, thus far, genes of S-layer proteins from only a few species, including *B. brevis* strains (6, 35), *Halobacterium halobium* (14), *Deinococcus radiodurans* (23), *Caulobacter crescentus* (7), *Aeromonas salmonicida* (4), and *Acetogenium kivui* (24), have been isolated and analyzed. The degrees of amino acid sequence homology of S layers among these species vary from total unrelatedness to partial similarity (6, 24, 25). The S-layer proteins commonly contain a large proportion of acidic, hydrophobic, and hydroxyl-containing amino acids and are low in sulfur-containing amino acids (24, 32).

In addition to being important structural units in prokaryotes, S layers also possess substantial potential for biotechnological applications. The requirement of large numbers of S-layer subunits per cell implies highly efficient synthesis and transport. In fact, with the promoters and signal sequence of the *B. brevis* surface protein, efficient syntheses of several heterologous proteins have been demonstrated (40, 41).

In *Lactobacillus brevis* (ATCC 8287), the S layer is composed of tetragonally arranged subunits, which are composed of a protein with a molecular weight of about 51,000 (16). The subunits dissociate from the cell wall with guanidine hydrochloride, and they can be reassembled into a native-like array (17).

In the present study, we have purified the *L. brevis* S-layer protein, amplified its gene by polymerase chain reaction (PCR), and sequenced the structural and regulatory regions of the S-layer gene. Our special emphasis was to analyze the expression and secretion signals of this protein in order to study its synthesis and regulation.

MATERIALS AND METHODS

Bacterial strains. *L. brevis*, designated herein GRL1, was obtained from the German Collection of Microorganisms (DSM/20556 = ATCC 8287) and cultivated in MRS broth (Oxoid) without shaking at 37°C.

Purification of the S-layer protein. GRL1 cells grown in MRS broth were collected by centrifugation at 10,000 × *g* for

* Corresponding author.
† Present address: National Public Health Institute, Helsinki, Finland.
‡ Present address: Agricultural Research Centre of Finland, Food Research Institute, 31600 Jokioinen, Finland.

5 min and washed once with 50 mM Tris-HCl, pH 7.5. The pellet, equivalent to 1 ml of culture, was dissolved directly in 200 μl of Laemmli sample buffer and applied to a preparative polyacrylamide gel (10%). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed by the method of Laemmli (13). After electrophoresis, the gel was treated with 1 M KCl to visualize protein bands. The band corresponding to the S-layer protein was excised and cut into pieces. The protein was eluted from the gel pieces in 1.5 ml of 6 M guanidine hydrochloride–0.5 M Tris-HCl–2 mM EDTA, pH 7.5, by incubating in an end-over mixer at room temperature for 10 h. The eluate was dialyzed against 10 mM Tris-HCl, pH 8.5, at +4°C for 10 h and freeze-dried.

To separate the S-layer protein for NH_2-terminal amino acid sequence analysis, SDS-PAGE with a 12% separation gel and a 5% running gel was used. From the gel the protein was transferred electrophoretically onto a polyvinylidene difluoride membrane (19).

Digestion and separation of peptides. The freeze-dried protein was suspended in 300 μl of distilled water, and 500 ng of lysylendopeptidase (WAKO, Dallas, Tex.) was added. The mixture was incubated at 35°C for 6 h, and the resulting peptides were separated by reversed-phase chromatography on a Vydack 218 TPB5 column (0.46 by 15 cm) connected to a Varian 5000 liquid chromatograph. The peptides were eluted with a linear gradient of acetonitrile (0 to 60% for 90 min) in 0.1% trifluoroacetic acid.

NH_2-terminal sequence analysis of the purified protein and peptides. For NH_2-terminal amino acid sequence analysis, the S-layer protein was degraded in a gas-pulsed-liquid sequencer (9) after transfer onto a polyvinylidene difluoride membrane. The peptides were sequenced after application on Polybrene (2 mg)-pretreated fiberglass filters.

Immunological procedures. S-layer protein was isolated from the intact GRL1 cells by preparative SDS-PAGE. The protein band was eluted from gel slices as described above. Ten micrograms of dialyzed protein was used for immunization of rabbits by standard procedures (National Public Health Institute, Helsinki, Finland). The antiserum was designated KH1225 and used in immunogold labelling electron microscopy.

Isolation and purification of DNA. Chromosomal DNA from GRL1 was isolated as follows. Cells were grown in 200 ml of MRS broth to mid-logarithmic growth phase (Klett [filter 66] = 80), collected by centrifugation at 8,000 × *g* for 5 min at room temperature, and suspended in 3 ml of 6 M guanidine hydrochloride solution. After incubation for 20 min at room temperature, the cells were collected by centrifugation, washed once with 0.15 M NaCl–0.1 M EDTA solution, and suspended in 7 ml of the same solution. Twenty milligrams of lysozyme (Sigma), 200 μl of mutanolysin (15,000 U/ml; Sigma), and 3 μl of 1 M $CaCl_2$ were added, and the cells were incubated for 2 h at 55°C. After the addition of 8.75 ml of SDS, followed by incubation for 10 min at 65°C, 2.2 ml of 5 M sodium perchlorate was added and the cell lysate was extracted with chloroform-isoamylalcohol (24:1). The water phase was removed after centrifugation and precipitated with ethanol. The chromosomal DNA was collected around a glass rod. DNA was dissolved in 10 mM Tris-HCl–150 mM NaCl, pH 7.5. DNA fragments generated by PCR were purified after synthesis by proteinase K treatment (5) and agarose gel electrophoresis.

Oligonucleotide synthesis. The oligonucleotides were synthesized with an Applied Biosystems model 381A DNA synthesizer. Deoxyinosine was utilized for primers synthesized according to the N-terminal amino acid sequences of the S-layer protein and its peptides.

Amplification of DNA by PCR. The S-layer DNA was amplified by PCR as follows. The first PCR mixtures consisted of 250 ng of *L. brevis* GRL1 chromosomal DNA; 100 pmol of primers with deoxyinosine at the third unknown position; 0.2 mM each dATP, dCTP, dGTP, and dTTP; 1.5 mM $MgCl_2$; and 5 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus) in 100 μl of reaction buffer recommended by the manufacturer. Each amplification cycle included three reaction steps, i.e., 0.5 min at 95°C for denaturation, 1 min at 46°C for annealing, and 1.5 min at 72°C for extension. All together, 30 cycles were carried out for each sample by using a Perkin-Elmer Cetus DNA thermal cycler. Other PCRs were performed with Vent DNA polymerase (BioLabs) according to the manufacturer's instructions, using 3.5 mM $MgSO_4$ and 1 U of Perfect Match Enhancer (Stratagene). The template DNA for the PCRs, used to amplify the upstream and downstream regions of the S-layer gene with primers specific for the pBR322 vector and the S-layer gene (PCR2 and PCR3; see Fig. 3A), were generated as follows. GRL1 chromosomal DNA was digested with *Eco*RI, and the 2.5-kb fraction, carrying the S-layer gene, was isolated from agarose gel, treated with kinase, and ligated with *Eco*RI-*Bam*HI-digested pBR322 at the molar ratio of 1:1 by standard methods (26). The amount of ligated template in the PCRs was 30 ng in 50 μl of reaction volume.

Sequencing of the DNA. PCR fragments were sequenced by the dideoxy-chain termination method (27) by using the Sequenase kit (U.S. Biochemical) with α-^{35}S-dATP (Amersham).

RNA isolation, Northern (RNA) blot, and primer extension. Total RNA was isolated from *L. brevis* GRL1 cells grown to logarithmic growth phase essentially as described for *Bacillus subtilis* (21) with mutanolysin (Sigma) and lysozyme (Sigma) at concentrations of 900 U/ml and 5 mg/ml, respectively. RNA gel electrophoresis and Northern blot were performed as described previously (21). Hybridization probes were labelled with [α-^{32}P]dCTP (>3,000 Ci/mmol; Amersham) by using the protocol of the random-prime DNA labelling kit (Boehringer Mannheim). The 5′ ends of the S-layer mRNAs were determined by primer extension as described previously (10).

Electron microscopy. GRL1 cells were collected with glass rods from the MRS plates to the fixatives described below.

For immunoelectron microscopy, the cells were fixed either with 2.5% glutaraldehyde (Electron Microscopy Sciences, Washington, Pa.) in 0.1 M phosphate buffer (pH 7.5) or with 3.5% paraformaldehyde and 0.05% glutaraldehyde dissolved in 0.1 M phosphate buffer. Fixations were carried out at room temperature for 1 h. The cells were washed three times in the same buffer. The cells were then dehydrated in a series of ethanol concentrations (50, 60, 70, and 90%) and treated with 90% ethanol and LR White (London Resin) at a dilution of 2:1 for 1 h and at a dilution of 1:1 for 1 h before the final embedding in LR White. Resin (20) was polymerized for 24 h at 60°C. Thin sections were mounted on nickel grids which were placed on droplets of the following solutions for the indicated times: blocking solution (Tween 20 plus 1% bovine serum albumin [TBST-B]) for 10 min, antibody solution (1:200) for 2 h, blocking solution for 10 min, and a solution of protein A-gold (10-nm diameter; Zymed) diluted 1:200 in TBST-B for 1.5 h. The sections were then washed twice in Tween 20 for 10 min. Grids were extensively washed with distilled water on a shaker. After drying, the sections were stained with uranyl acetate and lead citrate

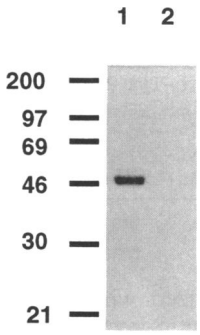

FIG. 1. SDS-PAGE analysis of *L. brevis* GRL1 proteins. Lanes: 1, intact cells; 2, culture medium. Numbers on the left are molecular size markers.

(Carlsberg's System; LKB). The preimmune serum of the same rabbit was used as a control.

Both of the micrographs were taken with a JEOL 1200EX transmission electron microscope at 60 kV.

Computer analysis. The DNA and the predicted protein sequences were analyzed with the PC/GENE set of programs (Genofit). With the FASTA program (22), the SWISSPROT, NBRF, and GenBank data bases were used for searching for homologous protein and nucleic acid sequences.

Nucleotide sequence accession number. The GenBank accession number of the 1,763-bp PCR fragment which contains the S-layer gene is Z14250.

RESULTS

Characterization of the S-layer protein. When intact cells of *L. brevis* GRL1 were boiled in Laemmli sample buffer and analyzed by SDS-PAGE, only one major band, with an apparent molecular mass of 46 kDa, was detected (Fig. 1). To confirm that this 46-kDa protein corresponded to the S-layer protein of *L. brevis*, the cells were treated with the antiserum (KH1255) raised against the isolated 46-kDa protein and analyzed by immunoelectron microscopy. The KH1255 antibody detected antigens which were mainly located in the outermost part of the cell wall. Only a few gold particles were found in the inner part of the cell wall or in the cytoplasm (Fig. 2A). In the control cells treated with preimmune serum (Fig. 2B), no gold particles could be seen either in the cell wall structures or in the cytoplasm. The postembedding immunoelectron microscopy clearly showed that the S-layer protein is heavily enriched in the outermost part of the cell wall of *L. brevis* GRL1 cells.

For sequence analysis of the 46-kDa S-layer protein, the GRL1 cells were boiled in Laemmli sample buffer. The solubilized S-layer protein was separated by SDS-PAGE, transferred to a polyvinylidene difluoride membrane, and degraded in a gas-pulsed-liquid sequencer. The N-terminal

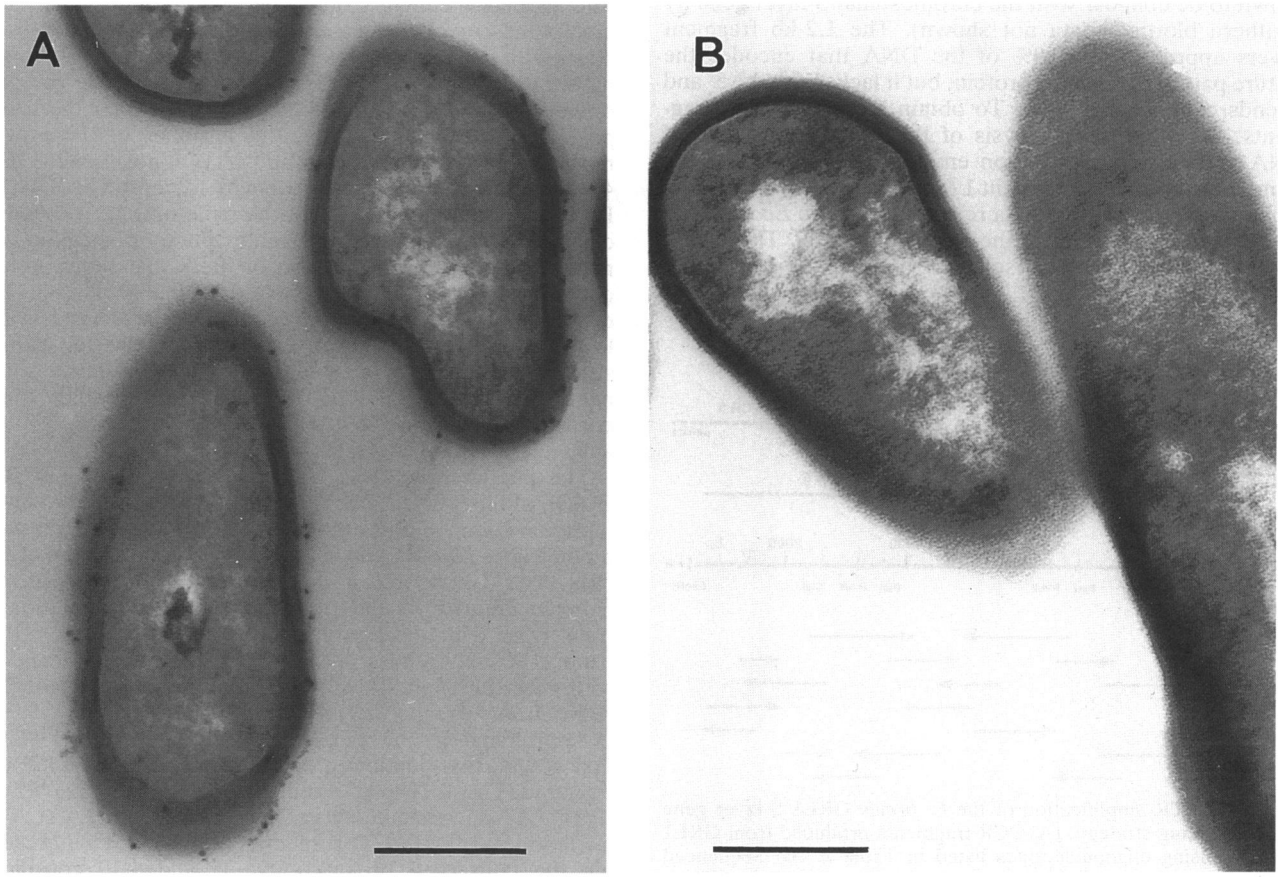

FIG. 2. Immunogold electron microscopy of *L. brevis* cells. (A) Postembedding labelling of cells with KH1255 antiserum raised against the 46-kDa protein and 10-nm colloidal gold particles. (B) Postembedding labelling of the same cells with preimmune serum and 10-nm colloidal gold particles. Bars = 0.2 μm.

TABLE 1. Peptide sequences of the S-layer protein[a]

Peptide no.	Sequence
1	Ala-Tyr-His-Tyr-Thr-Tyr-Thr-Tyr-Asn-Lys
2	Ser-Ala-Asp-Tyr-Phe-Arg-Ala-Tyr-Gly-Val-Lys
3	Tyr-Arg-Gly-Tyr-Val-Tyr-Gly-Gly-Lys
4	Thr-Thr-Asn-Arg-Gly-Ser-Val-Tyr-Tyr-Arg-Val-Val-Thr-Met
5	Ser-Tyr-Ala-Thr-Ala-Gly-Ala-Tyr-Ser-Thr-Leu-Lys

[a] For details, see Materials and Methods.

sequence of the intact S-layer protein was NH_2-Lys-Ser-Tyr-Ala-Thr-Ala-Gly-Ala-Tyr-Ser. For peptide analysis, the purified S-layer protein was subjected to lysylendopeptidase digestion and the resulting peptides were separated by reversed-phase chromatography. Five peptides were chosen for sequence analysis (Table 1). Peptide 5 was found to be the NH_2-terminal peptide of the S-layer protein without the NH_2-terminal lysine residue that was most probably removed in the lysylendopeptidase digestion.

Isolation of the S-layer gene. The S-layer gene was isolated by PCR since the gene detected from a *L. brevis* gene library in lambda gt10 turned out to be unstable in *Escherichia coli*. The first PCR fragments were generated by using oligonucleotide primers corresponding to the N-terminal amino acid sequence of the S-layer protein and its peptides (Table 1). These PCR oligonucleotide primers mostly contained deoxyinosine at the third position of each codon. The largest fragment generated was 1.2 kb (Fig. 3A, PCR1), and it was shown to be colinear with the chromosomal S-layer gene by Southern blotting (data not shown). The 1.2-kb fragment covers approximately 80% of the DNA that encodes the mature part of the S-layer protein, but it lacks both the 5′ and 3′ ends of the intact gene. To obtain these unknown fragments, Southern blot analysis of the GRL1 chromosomal DNA with various restriction enzymes was performed by using the 1.2-kb PCR fragment 1 (PCR1) as the hybridization probe (data not shown). As a result, one 2.5-kb *Eco*RI band was shown to cover all of the relevant regions. The 2.5-kb *Eco*RI fragment pool was isolated from a gel, ligated to the pBR322 vector linearized with *Eco*RI and *Bam*HI, and amplified by PCR. The S-layer gene and pBR322-specific oligonucleotides were used as primers to generate PCR2 and PCR3 (Fig. 3A) covering the missing upstream and downstream regions. Furthermore, additional PCRs with sequence-specific primers of the S-layer gene were performed directly from chromosomal GRL1 DNA to overlap the S-layer gene region and to confirm the sizes of the previous PCR fragments (Fig. 3A). The oligonucleotide sequences used for generation of the essential PCR fragments are listed in Table 2.

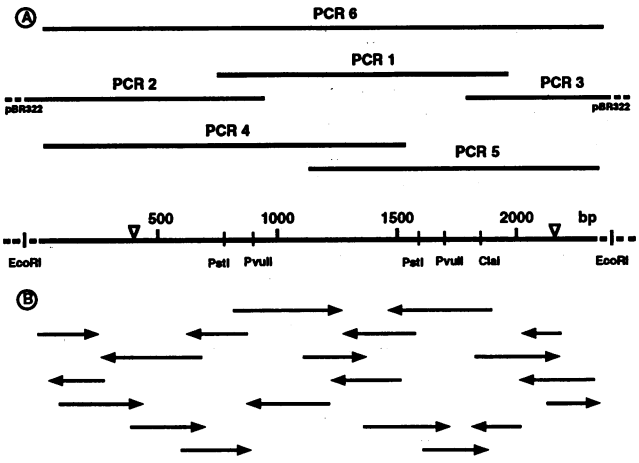

FIG. 3. PCR amplification of the *L. brevis* GRL1 S-layer gene and sequencing strategy. (A) PCR fragments produced from GRL1 DNA by using oligonucleotides listed in Table 2. (B) Sequenced fragments. Between panels A and B, the physical map of the S-layer region with relevant restriction enzyme sites is shown. The open arrowheads refer to the endpoints of the DNA sequence shown in Fig. 4.

Nucleotide sequence of the S-layer gene. Sequencing of the S-layer gene region was performed from the PCR fragments shown in Fig. 3A by the strategy shown in Fig. 3B. Both strands of the PCR fragments were sequenced by Sanger's dideoxy method. Altogether, sequence information was gathered from a 2.3-kb DNA region of the *L. brevis* GRL1 genome, of which approximately 1.76 kb is shown in Fig. 4. The sequence contains only one open reading frame and covers the complete S-layer gene. The open reading frame starts with ATG at nucleotide position 247, which is followed by six other codons before the next ATG. Since the only obvious ribosome binding site locates between these two possible initiation codons, the N terminus of the protein most likely starts from the second ATG at position 268 (Fig. 4). Thus, the gene has coding capacity for a protein of 48,159 Da. The first 90 nucleotides of the structural gene encode a characteristic gram-positive signal peptide of 30 amino acid residues (36). At the 5′ region of the signal sequence are codons for two positively charged amino acids (K). The central part encodes hydrophobic amino acids and is followed by codons for a typical signal peptidase recognition site, alanine-serine-alanine. Removal of the putative signal peptide results in a polypeptide of 435 amino acids which corresponds well to the 46-kDa size of the S-layer protein analyzed by SDS-PAGE (Fig. 1).

The putative ribosome binding site (Fig. 4, GGAGGA [RBS]) of the S-layer gene is located 9 bases upstream of the initiation codon and resembles typical prokaryotic ribosome binding sites (29, 34). Since the sequence information of 16S rRNAs of *Lactobacillus* species is still very limited, the complementarity of the ribosome binding site of S-layer transcripts with the equivalent 3′ end of the 16S rRNA cannot be determined. The sequence, however, matches well with the 3′ end (3′-UCUUUCCUCCA-5′) of the 16S rRNA from *Lactococcus lactis* and *Bacillus subtilis* (15, 30).

Approximately 163 and 77 nucleotides upstream of the first codon, two putative promoter regions, P1 and P2, can be found (Fig. 4). The −35 and −10 regions of P1 and P2 are cTTGTAT–15 nucleotides–gcTATACT and tcTTAACA–15 nucleotides–gtTATACT, respectively, which closely resemble the conserved prokaryotic −35 and −10 consensus sequences, tcTTGACAt and tgnTATAAT, respectively (37). In addition to the canonical hexanucleotides and some conserved nucleotides adjacent to them, similarity with

TABLE 2. Oligonucleotides used for synthesis of PCR fragments

Fragment	Primer sequence[a]	Position[b] or characteristic
PCR1	5'-TA T/C GCIACIGCIGGIGCITA T/C T/A C/G IACI T/C TIAA-3'	364–395
	5'-TT A/G TT A/G TAIGT A/G TAIGT A/G TA A/G TG A/G TAIGC-3'	1541–1513
PCR2	5'-GTATCACGAGGCCCT-3'	pBR322 specific
	5'-TAGTCAGCTGACTTCTTTGAAGAAG-3'	536–512
PCR3	5'-ACAACTGACTTGACTGGTGA-3'	1398–1418
	5'-GTATCACGAGGCCCT-3'	pBR322 specific
PCR4	5'-CCTGAGTTATAGTGGCGTG-3'	−376–−358
	5'-TAGCATCTGCAGCATTAGGGTTAG-3'	1198–1175
PCR5	5'-AACGACTACTAAGGCTGATATGCC-3'	675–698
	5'-CGTGTTCTCCTCCAATGAAGC-3'	1972–1951
PCR6	5'-CCTGAGTTATAGTGGCGTG-3'	−376–−358
	5'-CGTGTTCTCCTCCAATGAAGC-3'	1972–1951

[a] The two primer sequences for each PCR fragment correspond to the coding and complementary strands. "I" refers to deoxyinosine, and T/C and A/G, etc., refer to the mixture of the two nucleotides.
[b] Position of the primer sequences at the S-layer region (Fig. 4).

promoter regions of different gram-positive bacteria (8) can be found in an A-rich region upstream of the −35 region of promoter P2 and in the presence of adenosine at positions −4 and −7 of P2 and −7 of P1.

By using PC/GENE programs, a putative transcription terminator sequence can be found 24 nucleotides downstream from the two stop codons of the S-layer gene (Fig. 4). The hairpin structure is AU rich and consists of a stem of 19 bp (11 UT pairs, 5 GC pairs, and 3 unpaired U's and G's) and a loop of 6 bp (ΔG, −24.9 kcal/mol [1 cal = 4.184 J]). The hairpin loop is followed by an AU-rich region (Fig. 4). Approximately 460 bp upstream of the structural gene, a typical rho-independent-type transcription terminator rich in GC pairs (ΔG, −15.8 kcal/mol) can also be found (sequence not shown), indicating that the S-layer gene is monocistronic.

Furthermore, in the coding region of the S-layer gene, 10 partly overlapping direct repeat sequences of 10 to 12 nucleotides can be found (Table 3). These DNA repeats correlate at the protein level to repeating stretches of two to four amino acids rich in threonine, alanine, glycine, and serine residues (Table 3).

The codon usage of the S-layer protein is clearly biased (Table 4) and resembles mostly that reported for highly expressed *B. subtilis* proteins (28). The only striking difference in the codon usage of the S-layer protein from that of *B. subtilis* proteins is the lysine codons. The ratios of AAA to AAG codons in the *L. brevis* S layer and *B. subtilis* proteins are 3:36 and 31:3, respectively.

The amino acid composition of the S-layer protein deduced from the DNA sequence is shown in Table 5. The typical features of the amino acid composition are the large number of hydrophobic amino acids, amino acids with hydroxyl groups, and the absence of cysteine. The amount of basic amino acids is larger than that of acidic amino acids, which is a notable exception to the general features of other characterized S-layer proteins (33).

Analysis of mRNA of the S-layer gene. The size of the mRNA transcribed from the S-layer gene was analyzed by Northern blotting with PCR1 (Fig. 3A) as the hybridization probe. The probe detected a 1.5-kb transcript (Fig. 5A). The size of the S-layer mRNA is in good agreement with the size of the S-layer gene, thus further confirming the DNA sequence analysis and the monocistronic nature of the S-layer operon. The primer extension mapping of the 5' ends of the S-layer mRNA revealed two 5' ends with a size difference of 86 bases (Fig. 5B). These start sites of the S-layer transcripts locate immediately downstream from the two putative promoter regions deduced from the DNA sequence (Fig. 4), thus confirming the existence and functionality of the found promoters. The primer extension was done with total RNA isolated from exponentially growing cells. The intensities of the RNA bands from the primer extension are similar, suggesting that both promoters are used with roughly equal efficiency at this growth phase.

DISCUSSION

In this work, we have analyzed the gene encoding the S-layer protein of *L. brevis*. The gene was amplified by PCR as separate, overlapping pieces and as a full-length fragment. Amplification was followed by DNA sequencing. For most PCRs, Vent polymerase was used to minimize the error frequency of the in vitro synthesis (12, 18). The DNA sequence of 2.3 kb revealed only one complete open reading frame, of which 1,395 bp encodes the structural part of the S-layer protein. Thus, the S-layer gene has the capacity to encode a protein of 48,159 Da. The NH_2-terminal sequence analysis of the intact protein indicates that the mature S-layer protein starts with the lysine residue at position 30 of the predicted amino acid sequence (Fig. 4). Removal of the signal peptide results in a polypeptide of 435 amino acids with a calculated molecular mass of 45 kDa. This is in excellent agreement with the apparent molecular mass of 46 kDa by SDS-PAGE analysis of the S-layer protein released from the cell wall of GRL1 cells. Furthermore, all five peptide sequences of the isolated S-layer protein (Table 1), which cover 12% of the whole amino acid sequence, can be found from the deduced sequence (Fig. 4). This further confirmed the correctness of the PCR fragments analyzed. Also, the sizes of S-layer mRNA and the transcription start sites (Fig. 5), as well as Southern blot analyses (data not shown), are in excellent agreement with the DNA data. We

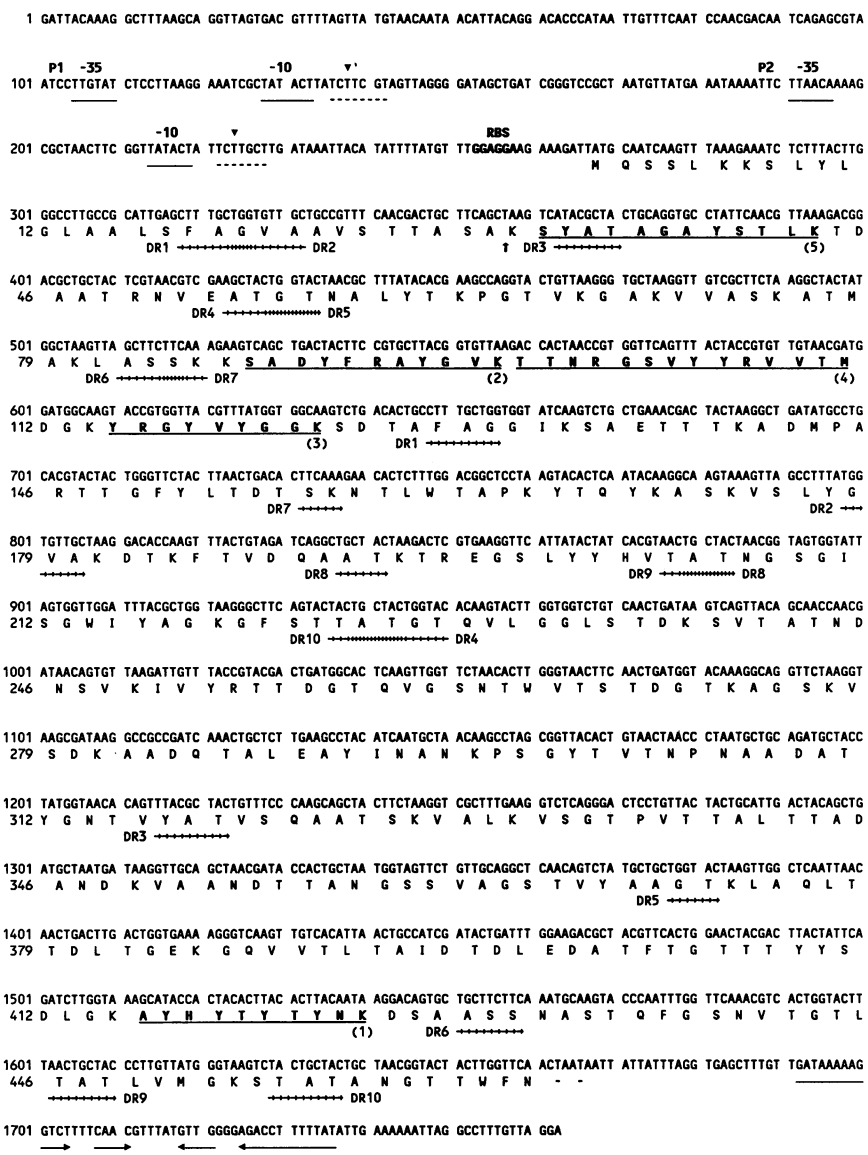

FIG. 4. Nucleotide sequence of the *L. brevis* S-layer gene. The predicted −10 and −35 regions of promoters P1 and P2 are underlined, and the 5′ ends of the transcripts found by primer extension (see Fig. 5) are marked with arrowheads. The cleavage site of the signal peptide is between amino acids 30 and 31 (↑). The N-terminal amino acid sequences of the intact S-layer protein and its tryptic peptides are underlined and numbered from (1) to (5) (Table 1). The deduced transcription terminator is shown with arrows. RBS refers to the predicted ribosome binding site. Direct repeats are marked with DR and a number and short continuous arrows.

thus conclude that the S-layer gene has been reliably synthesized by PCR.

In accordance with some other S-layer proteins (24), a special feature of the *L. brevis* S-layer protein is also a high content of amino acids with hydroxyl groups (threonine, tyrosine, and serine) (Table 5). This might indicate the presence of several inter- and/or intramolecular hydrogen bonds in the molecule. This is supported by the finding that S-layer protein could be released from the cells by guanidine hydrochloride. In some S-layer proteins, the hydroxyl-containing serine and threonine are found in clusters of five to six residues (24), whereas in the deduced sequence of the *L. brevis* S-layer protein, hydroxy amino acids are randomly distributed and only one cluster is present (TTTYYS at nucleotide position 1483). In our study, repeated stretches of two to four amino acids rich in threonine, alanine, glycine, serine, and tyrosine residues were found (Table 3). Some of them are partly overlapping, forming thus longer continuous stretches (Fig. 4). Many of the repeats are located partly or entirely at the predicted β-turn structures of the protein. In fact, 22 amino acids of 31 involved in the formation of the deduced β-turns are positioned at these repeats. The exact function of these repeats is unknown, but some of them might be involved in the formation of linkers between different domains of the polypeptide (2). Furthermore, like other S-layer proteins, the *L. brevis* S-layer protein contains no cysteine residues, which is also typical for other gram-positive exoproteins. In contrast to other S-layer proteins, the *L. brevis* S-layer protein contains an unexpectedly large amount of basic amino acid residues (lysine and arginine)

TABLE 3. Direct repeats in the *L. brevis* S-layer gene

DR[a]	Nucleotide sequence	Position[b]	Amino acid sequence[c]	Position[b]
DR1	C TTT GCT GGT G	318	(S) F A G (V)	18
		648	(A) F A G (G)	128
DR2	T GGT GTT GCT	324	(A) G V A	20
		798	(Y) G V A	178
DR3	TAC GCT ACT G	364	Y A T (A)	33
		1216	Y A T (V)	317
DR4	GCT ACT GGT AC	424	A T G T	53
		940	A T G T	225
DR5	CT GGT ACT AA	428	(T) G T (N)	55
		1376	(A) G T (K)	371
DR6	GCT TCT TCA AA	511	A S S (K)	82
		1552	A S S (N)	429
DR7	CT TCA AAG AA	515	(S) S K (K)	84
		731	(T) S K (N)	156
DR8	CT GCT ACT AA	836	(A) A T (K)	191
		878	(T) A T (N)	205
DR9	TA ACT GCT AC	875	(V) T A T	204
		1601	(L) T A T	446
DR10	CT ACT GCT ACT G	935	(T) T A T (G)	224
		1628	(S) T A T (A)	455

[a] DR, direct repeat.
[b] Positions as shown in Fig. 4.
[c] In-frame amino acids corresponding to the nucleotide repeats. Amino acids in parentheses represent the uncomplete codons of the repeats.

(Table 5). According to the Chargpro program of PC/GENE, the isoelectric point of the protein would be 9.88. A hydropathy plot of the amino acid sequence shows no highly hydrophobic or hydrophilic regions in the molecule. According to secondary structure prediction analysis (Garnier program of PC/GENE), the S-layer protein is irregularly folded, containing α-helix (27.7%), turn (6.6%), random coil (9.0%), and extended (56.5%) structures in mixed order. An aperiodic structure has also been shown for other S-layer proteins by infrared spectroscopy and circular dichroism (32). The β-structure content of the *L. brevis* S-layer protein seems to be somewhat higher than that of other S-layer proteins, for which contents of around 30% have been found (24).

Protein and nucleic acid homology searches of the NBRF, SWISSPROT, and GenBank data bases revealed no genuinely related sequences. Interestingly, weak similarity was found between the S-layer protein and the serine proteinases of *L. lactis* SK11 and Wg2 (11, 38). The amino acids of the S-layer protein from 95 to 357 and from 159 to 386 showed 21.5 and 19.6% identities to the PIII- and PI-type proteinases of SK11 and Wg2, respectively. The overlapping S-layer sequences matched to two different regions in the amino acid

TABLE 4. Codon usage of the *L. brevis* S-layer protein

Amino acid and codon	No.	Amino acid and codon	No.	Amino acid and codon	No.	Amino acid and codon	No.
Ala		Gln		Leu		Ser	
GCT	46	CAA	9	CTC	0	AGT	9
GCC	8	CAG	1	CTA	0	AGC	4
GCA	13	Glu		CTG	1	Thr	
GCG	0	GAA	6	Lys		ACT	58
Arg		GAG	0	AAA	3	ACC	7
CGT	8	Gly		AAG	35	ACA	8
CGC	0	GGT	34	Met, ATG	5	ACG	10
CGA	0	GGC	6	Phe		Trp, TGG	4
CGG	0	GGA	1	TTT	4	Tyr	
AGA	0	GGG	2	TTC	5	TAT	7
AGG	0	His		Pro		TAC	23
Asn		CAT	0	CCT	5	Val	
AAT	6	CAC	2	CCC	0	GTT	21
AAC	15	Ile		CCA	1	GTC	7
Asp		ATT	3	CCG	0	GTA	7
GAT	16	ATC	3	Ser		GTG	0
GAC	8	ATA	0	TCT	11	Stop	
Cys		Leu		TCC	1	TAA	0
TGT	0	TTA	9	TCA	18	TGA	0
TGC	0	TTG	7	TCG	0	TAG	0
		CTT	8				

TABLE 5. Deduced amino acid composition of the *L. brevis* S-layer protein

Amino acid	No.	%	Amino acid	No.	%
Ala	67	14.4	Leu	25	5.3
Arg	8	1.7	Lys	38	8.1
Asn	21	4.5	Met	5	1.0
Asp	24	5.1	Phe	9	1.9
Cys	0	0.0	Pro	6	1.2
Gln	10	2.1	Ser	43	9.2
Glu	6	1.2	Thr	83	17.8
Gly	43	9.2	Trp	4	0.8
His	2	0.4	Tyr	30	6.4
Ile	6	1.2	Val	35	7.5

sequences of these serine proteinases, indicating intramolecular homology in these enzymes (38). Even if these similarities between the S-layer protein and the *Lactococcus* proteinases were found, their biological significance is not obvious and more knowledge about the functions and evolution of the *L. brevis* S layer would be required to evaluate this result. It is also notable that no amino acid sequence similarity to other S-layer proteins was found. Only a short, 68-bp overlap with 66.2% identity with the DNA sequence of *A. kivui* was found. At the protein level, however, only seven identical amino acid residues were discontinuously present. The result is not entirely unexpected, since the S-layer sequence information available comes from organisms which are phylogenetically very distant. Previous homology comparisons have also revealed that no or only partial sequence similarity can be found among the S layers from *D. radiodurans*, *B. brevis*, and *A. kivui* (6, 24, 25).

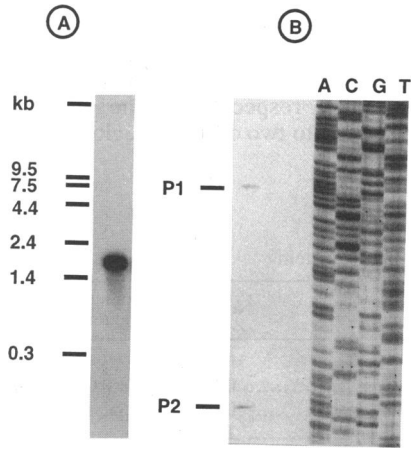

FIG. 5. Analysis of the S-layer mRNA. (A) Northern blot analysis of transcripts. Total *L. brevis* GRL1 RNA denatured with glyoxal and dimethyl sulfoxide was run in an 0.8% agarose gel with 10 mM phosphate buffer, pH 6.5, and then blotted to a Zetaprobe membrane and hybridized. [α-^{32}P]dCTP-labelled PCR1 was used as probe. The filter was washed in 0.5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at 50°C. RNA molecular weight markers (Bethesda Research Laboratories) were used to determine transcript size. (B) Analysis of mRNA transcripts after the 5′-end-mapping experiment in a 6% sequencing gel. Approximately 5 μg of total RNA of *L. brevis* was hybridized to 200 pmol of a 19-mer primer (5′-CTTAGCCATATGAGCCTTA-3′ [Fig. 4, positions 507 to 489]). After ethanol precipitation, primer extension of the hybrids was performed in the presence of [α-^{32}P]dCTP, nonlabelled nucleotides, avian myeloblastosis virus reverse transcriptase, actinomycin C1, and RNasin. For determination of the lengths of the extended products, sequencing reactions of PCR2 (see Fig. 3) were performed with the same primer.

The encompassing S layer on an average-size cell has been estimated to consist of approximately 5×10^5 protein monomers (32). Synthesis, transport, and assembly of such a large amount of a single protein require highly efficient expression and secretion signals, synthetic capacity, and translocation machinery. The putative expression and secretion signals characterized here from *L. brevis* show high sequence similarity within the evolutionarily conserved regions to those of other prokaryotes, especially of gram-positive organisms. A distinctive feature of the *L. brevis* S-layer operon is the presence of two functional RNA polymerase binding sites, P1 and P2 (Fig. 4 and 5). A similar multiple and tandem promoter region has been found, among the S-layer operons, so far only from *B. brevis* strains (1, 6, 40). In *B. brevis* 47, there are several promoters upstream of the *cwp* (cell wall protein gene) operon, of which the promoters P2 and P3 play the most important role (1). P2 is a constitutive promoter, whereas P3 is used only at the exponential growth phase (1). In *L. brevis*, both promoters are used, at least during the exponential phase of growth, with similar efficiency (Fig. 5B). Whether there is selective use of these promoters at other growth phases remains to be seen. Inspection of secretion signals revealed that in *L. brevis*, the S-layer protein is transported on the cell surface by the aid of the putative 30-amino-acid signal peptide (Fig. 4). This is the first signal peptide described for *Lactobacillus* species in the literature. The S-layer protein is not released from the cell surface to culture medium (Fig. 1), indicating that its synthesis and transport are tightly regulated. In some other microbes, the transport and assembly of S layer are not efficiently coordinated, resulting in a partial loss of the protein to the surrounding medium (32, 33). The regulatory mechanisms involved in the biosynthesis, transport, and assembly of this S-layer proteins will be further elucidated.

ACKNOWLEDGMENTS

Lars Paulin and Christophe Roos are gratefully acknowledged for preparing the synthetic oligonucleotides and for performing the homology searches, respectively. We thank Eila Kujamäki and Mervi Lindman for excellent technical assistance.

This research was supported by the Finnish National Fund for Research and Development (SITRA) and the Academy of Finland.

REFERENCES

1. **Adachi, T., H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1989. Multiple and tandemly arranged promoters of the cell wall protein gene operon in *Bacillus brevis* 47. J. Bacteriol. **171:** 1010–1016.
2. **Argos, P.** 1990. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J. Mol. Biol. **211:**943–958.
3. **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.
4. **Chu, S., S. Cavaignac, J. Feutrier, B. M. Phipps, M. Kostrzynska, W. W. Kay, and T. J. Trust.** 1991. Structure of the tetragonal surface virulence array protein and gene of *Aeromonas salmonicida*. J. Biol. Chem. **266:**15258–15265.
5. **Crowe, J. S., H. J. Cooper, M. A. Smith, M. J. Sims, D. Parker, and D. Gewert.** 1991. Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucleic Acids Res. **19:**184.
6. **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.
7. **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.
8. **Graves, M. C., and J. C. Rabinowitz.** 1986. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for extended promoter elements in Gram-positive organisms. J. Biol. Chem. **261**:11409–11415.
9. **Kalkkinen, N., and C. Tilgmann.** 1988. A gas-pulsed-liquid-phase sequencer constructed from a Beckman 890D instrument by using Applied Biosystems delivery and cartridge blocks. J. Protein Chem. **7**:242–243.
10. **Koivula, T., M. Sibakov, and I. Palva.** 1991. Isolation and characterization of *Lactococcus lactis* subsp. *lactis* promoters. Appl. Environ. Microbiol. **57**:333–340.
11. **Kok, J., K. J. Leenhouts, A. J. Haandrikman, A. M. Ledeboer, and G. Venema.** 1988. Nucleotide sequence of the cell wall proteinase gene of *Streptococcus cremoris* Wg2. Appl. Environ. Microbiol. **54**:231–238.
12. **Kunkel, T. A., R. D. Sabatino, and R. A. Bambara.** 1987. Exonucleolytic proofreading by calf thymus DNA polymerase δ. Proc. Natl. Acad. Sci. USA **84**:4865–4869.
13. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
14. **Lechner, J., and M. Sumper.** 1987. The primary structure of a procaryotic glycoprotein. J. Biol. Chem. **262**:9724–9729.
15. **Ludwig, W., E. Seewaldt, R. Kilpper-Bälz, and E. Schleifer.** 1985. The phylogenetic position of streptococcus and enterococcus. J. Gen. Microbiol. **131**:543–551.
16. **Masuda, K., and T. Kawata.** 1979. Ultrastructure and partial characterization of a regular array in the cell wall of *Lactobacillus brevis*. Microbiol. Immunol. **23**:941–953.
17. **Masuda, K., and T. Kawata.** 1980. Reassembly of the regularly arranged subunits in the cell wall of *Lactobacillus brevis* and their reattachment to cell walls. Microbiol. Immunol. **24**:299–308.
18. **Mattila, P., J. Korpela, T. Tenkanen, and K. Pitkänen.** 1991. Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase—an extremely heat stable enzyme with proofreading activity. Nucleic Acids Res. **19**:4967–4973.
19. **Mozdzanowski, J., and D. W. Speicher.** 1990. Quantitative electrotransfer of proteins from polyacrylamide gels onto PVDF membranes, p. 87–93. *In* J. Villafranca (ed.), Current research in protein chemistry. Academic Press, Inc., San Diego.
20. **Newman, G. R.** 1989. LR White embedding medium for colloidal gold methods, p. 47–73. *In* M. A. Hayat (ed.), Colloidal gold: principles, methods, and applications, vol. 2. Academic Press, Inc., San Diego, Calif.
21. **Palva, A., K. Nyberg, and I. Palva.** 1988. Quantitation of α-amylase mRNA in *Bacillus subtilis* by nucleic acid sandwich hybridization. DNA **7**:135–142.
22. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
23. **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.
24. **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.
25. **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.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, second ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Shields, D. C., and P. M. Sharp.** 1987. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. Nucleic Acids Res. **15**:8023–8040.
29. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16 S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
30. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
31. **Sleytr, U. B., and P. Messner.** 1983. Crystalline surface layers on bacteria. Annu. Rev. Microbiol. **37**:311–339.
32. **Sleytr, U. B., and P. Messner.** 1988. Crystalline surface layers in procaryotes. J. Bacteriol. **170**:2891–2897.
33. **Smit, J.** 1986. Protein surface layers of bacteria, p. 343–376. *In* M. Inouye (ed.), Bacterial outer membranes as model systems. John Wiley & Sons, Inc., New York.
34. **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.
35. **Tsuboi, A., R. Uchihi, R. Tabata, Y. Takahashi, H. Hashiba, T. 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 gene. J. Bacteriol. **168**:365–373.
36. **Von Heijne, G.** 1985. Signal sequences: the limits of variation. J. Mol. Biol. **184**:99–105.
37. **Von Heijne, G.** 1987. Sequence analysis in molecular biology: treasure trove or trivial pursuit? Academic Press, Inc., San Diego.
38. **Vos, P., G. Simons, R. J. Siezen, and W. M. de Vos.** 1989. Primary structure and organization of the gene for a procaryotic cell envelope-located serine proteinase. J. Biol. Chem. **264**: 13579–13585.
39. **Yamada, H., N. Tsukagoshi, and S. Udaka.** 1981. Morphological alterations of cell wall concomitant with protein release in a protein-producing bacterium, *Bacillus brevis* 47. J. Bacteriol. **148**:322–332.
40. **Yamagata, H., T. Adachi, A. Tsuboi, M. Takao, T. Sasaki, N. Tsukagoshi, and S. Udaka.** 1987. Cloning and characterization of the 5′ region of the cell wall protein gene operon in *Bacillus brevis* 47. J. Bacteriol. **169**:1239–1245.
41. **Yamagata, H., K. Nakahama, Y. Suzuki, A. Kakinuma, N. Tsukagoshi, and S. Udaka.** 1989. Use of *Bacillus brevis* for efficient synthesis and secretion of human epidermal growth factor. Proc. Natl. Acad. Sci. USA **86**:3589–3593.