JOURNAL OF BACTERIOLOGY, Aug. 1989, p. 4178–4188
0021-9193/89/084178-11$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 171, No. 8

Cloning and Sequencing of the Gene Encoding a 125-Kilodalton Surface-Layer Protein from *Bacillus sphaericus* 2362 and of a Related Cryptic Gene

RON D. BOWDITCH,[1]† PAUL BAUMANN,[1]* AND ALLAN A. YOUSTEN[2]

Department of Microbiology, University of California, Davis, California 95616,[1] and Microbiology and Immunology Section, Biology Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0406[2]

Received 17 January 1989/Accepted 2 May 1989

Using the vector pGEM-4-blue, a 4,251-base-pair DNA fragment containing the gene for the surface (S)-layer protein of *Bacillus sphaericus* 2362 was cloned into *Escherichia coli*. Determination of the nucleotide sequence indicated an open reading frame (ORF) coding for a protein of 1,176 amino acids with a molecular size of 125 kilodaltons (kDa). A protein of this size which reacted with antibody to the 122-kDa S-layer protein of *B. sphaericus* was detected in cells of *E. coli* containing the recombinant plasmid. Analysis of the deduced amino acid sequence indicated a highly hydrophobic N-terminal region which had the characteristics of a leader peptide. The first amino acid of the N-terminal sequence of the 122-kDa S-layer protein followed the predicted cleavage site of the leader peptide in the 125-kDa protein. A sequence characteristic of promoters expressed during vegetative growth was found within a 177-base-pair region upstream from the ORF coding for the 125-kDa protein. This putative promoter may account for the expression of this gene during the vegetative growth of *B. sphaericus* and *E. coli*. The gene for the 125-kDa protein was followed by an inverted repeat characteristic of teminators. Downstream from this gene (11.2 kilobases) was an ORF coding for a putative 80-kDa protein having a high sequence similarity to the 125-kDa protein. Evidence was presented indicating that this gene is cryptic.

Bacillus sphaericus is a promising agent for the biological control of mosquitos which are vectors of important human and animal diseases (47). Toxicity for mosquito larvae has been associated with the formation, during sporulation, of a parasporal crystal and with proteins of 100 to 125 kilodaltons (kDa) which may be made during vegetative growth (3, 4, 7, 9, 12, 32, 33). In the case of *B. sphaericus* 2362, proteins with a molecular size of 42, 51, and 110 kDa have been shown to play a role in toxicity for mosquito larvae (2–4, 9). The genes for the 42- and 51-kDa proteins have recently been cloned and sequenced (2, 3, 21).

A number of gram-positive and gram-negative bacteria possess a protein or glycoprotein surface (S) layer that forms a barrier between the cell and the environment (39, 40). These proteins may constitute between 5 and 10% of the total protein of cells in exponential growth (39, 40). The high energy expenditure required for the synthesis of such a large amount of protein suggests that it has a vital function requiring its maintenance on the cell surface. Evidence has been presented indicating that the S layer acts as a protective barrier and plays a role in bacterial pathogenesis (22, 40). In the larvicidal strains of *B. sphaericus*, the S layer consists of a linear array of glycoproteins (29), the monomer having a molecular size of 127 to 129 kDa (45). In this species, it may serve as a site for bacteriophage attachment (28). On the basis of susceptibility to different bacteriophages, the mosquito-pathogenic strains have been subdivided into groups (47) which are in agreement with the subdivisions established by serological studies of the S-layer protein (29, 45).

In the present study, we cloned and sequenced a gene which codes for a 125-kDa precursor of the 122-kDa *B. sphaericus* 2362 S-layer protein (gene 125). Evidence is also presented indicating that the 122-kDa protein is the precursor of the 110-kDa larvicide and that both of these proteins are absent from the parasporal crystal of this species. The latter finding indicates that our previous conclusion that the 122- and 110-kDa proteins were constituents of the crystal is invalid since our "crystal" preparation (4) appears to have been contaminated with cell wall material containing S-layer proteins. The cryptic gene also cloned and sequenced in this study which codes for a putative 80-kDa protein has been designated as gene 80.

* Corresponding author.

† Present address: Department of Immunology, Research Institute of Scripps Clinic, LaJolla, CA 92037.

MATERIALS AND METHODS

Bacterial strains and vectors. *B. sphaericus* 2362 was used in our previous investigations (2, 4, 9). The λ vector EMBL3 (15) with the host strains *Escherichia coli* NM538 and NM539, the λ vector GEM2 (34) with the host strain *E. coli* LE392, and pGEM-4-blue were obtained from Promega Biotec (Madison, Wis.). The sources of *E. coli* TB1, the host for pGEM-4-blue, as well as *E. coli* JM101 and JM107 (31), used as the host strains for M13mp18/19, have been previously indicated (2, 3).

General procedures. Methods for restriction endonuclease analysis of DNA, agarose gel electrophoresis, large-scale λ DNA purification, plasmid DNA purification by $CsCl_2$-ethidium bromide density gradient centrifugation, and minipreparations of plasmid and bacteriophage DNA have been described by Maniatis et al. (30). Transformation of competent cells by plasmid DNA was performed by the procedure of Hanahan (18). Enzymes and substrates purchased from Boehringer Mannheim Biochemicals (Indianapolis, Ind.), New England BioLabs, Inc. (Beverly, Mass.), and Promega Biotec were used according to the directions of the manufacturers. The following procedures were performed as previously described (2, 4): sodium dodecyl sulfate-poly-

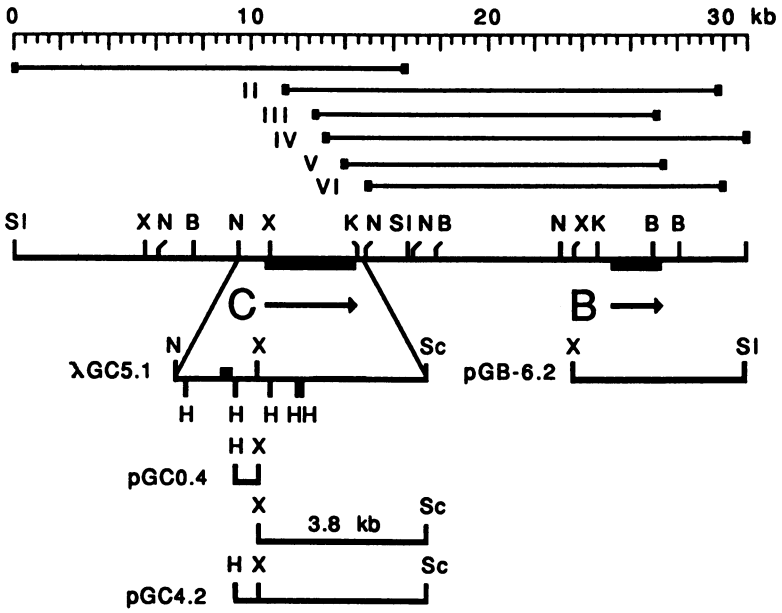


FIG. 1. Composite restriction map of a 32-kb fragment of *B. sphaericus* DNA containing gene 125 and gene 80 and the strategy used for subcloning these genes. Stippled bars indicate the regions coding for proteins. Arrows indicate direction of ORF. Regions spanned by bars and designated by Roman numerals represent the classes of λEMBL3 recombinants. Black bar on λGC5.1 indicates the region of the DNA which could not be subcloned into *E. coli*. B, *Bgl*II; H, *Hin*dIII; K, *Kpn*I; N, *Nhe*I; Sc, *Sac*I; Sl, *Sal*I; X, *Xba*I.

acrylamide gel electrophoresis (SDS-PAGE), electroblotting of proteins onto nitrocellulose and detection with antiserum to the crystal proteins (Western immunoblots), preparation of cell extracts of *E. coli*, Ouchterlony immunodiffusion experiments (antiserum dilution of 1:4), and bioassays involving larvae of the mosquito *Culex pipiens*. Protein content was determined by the BCA Protein Assay (Pierce Chemical Co., Rockford, Ill.), with bovine serum albumin as the standard.

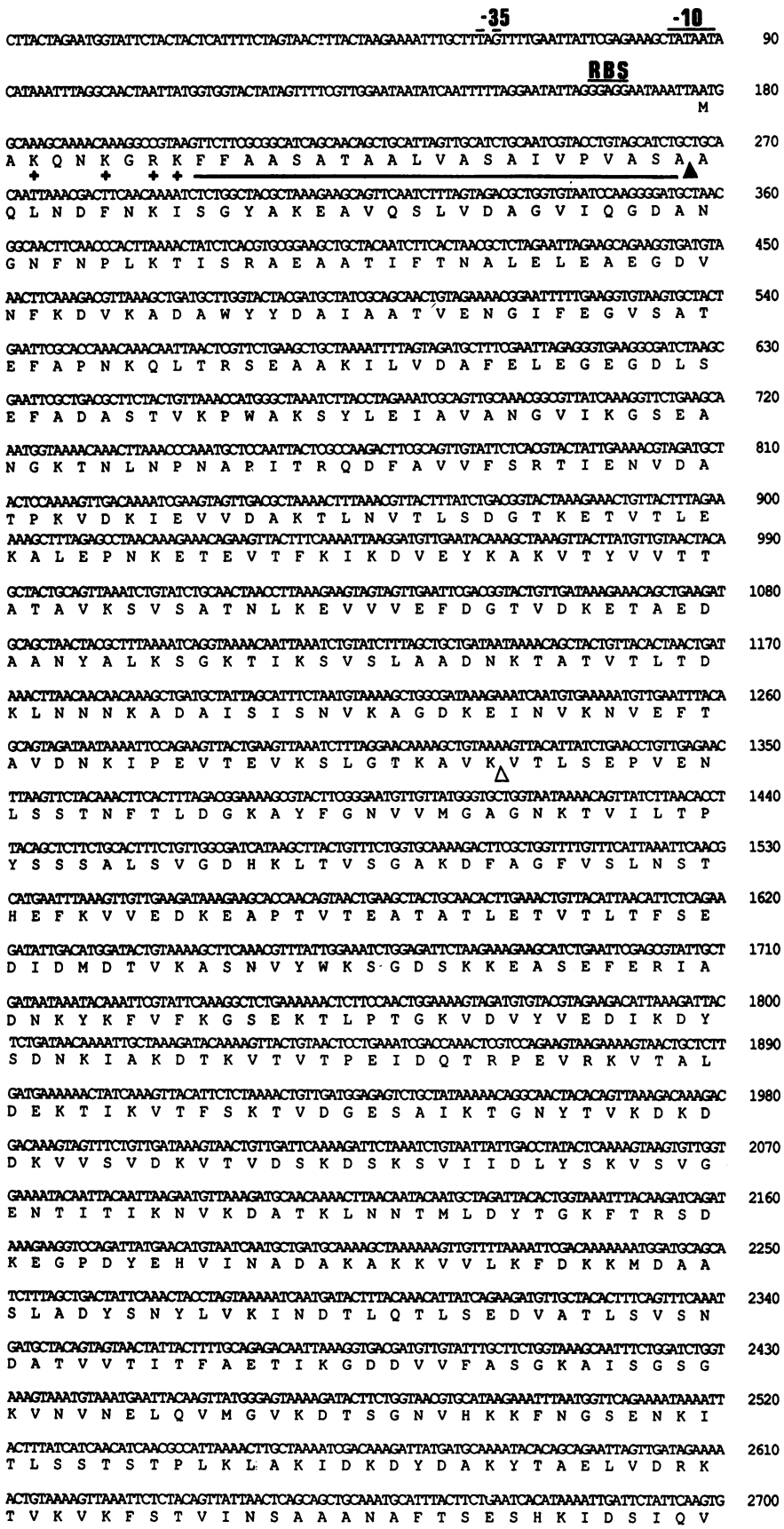
Cloning of gene 80. *B. sphaericus* 2362 DNA was isolated (2), partially digested with the restriction enzyme *Sau*3A, and size fractionated on sucrose gradients (30). The 10- to 20-kilobase (kb) *Sau*3A-digested DNA fragments were ligated into the *Bam*HI-digested λEMBL3 vector (15, 23), packaged by means of the Packagene Lambda DNA System (Promega Biotec), and plated on *E. coli* NM538 (Spi^+, permissive strain) and NM539 (Spi^-, nonpermissive strain) (23). A 0.5-ml portion of the packaging mix contained a total of 2.4×10^5 PFU; 41% of the phage particles were recombinants.

The library was screened by plating approximately 1,000 recombinant PFU per plate on *E. coli* NM539 (23). Plaque lifts were made with nitrocellulose filter disks (6). ^{35}S-labeled DNA probes of gene 80 and gene 125 (B and C genes, respectively, in reference 2) were obtained as described previously (2). A mixture of the radioactive gene 80 and gene 125 probes was hybridized to nitrocellulose filters under conditions of high stringency (5°C less than the temperature at which 50% of the DNA double helix is dissociated). At this temperature there was no cross hybridization between gene 80 and gene 125 (30). X-ray film (AR; Eastman Kodak Co., Rochester, N.Y.) was exposed to the filters as directed by the manufacturer. Recombinant phage which hybridized with the mixture of gene 80 and gene 125 probes were plaque purified three times. A total of 2.3×10^4 recombinant λEMBL3 phage yielded 24 positive plaques. These were characterized by restriction enzyme digestion followed by Southern blotting and hybridization with probes to either gene 80 or gene 125 as previously described (2) (Fig. 1). Six classes of overlapping recombinants were obtained (Fig. 1), with each class represented by the following number of identical recombinants; I = 6, II = 1, III = 6, IV = 4, V = 6, and VI = 1.

Using the *Sal*I site located in the polylinker of λEMBL3, we subcloned the 6.2-kb *Xba*I-*Sal*I fragment from the λEMBL3 class VI recombinant (Fig. 1), which contained the DNA hybridizing with the gene 80 probe, into pGEM-4-blue and designated it pGB6.2 (Fig. 1).

Cloning of gene 125. Difficulty was encountered in routinely obtaining sufficient DNA from the λEMBL3 class I recombinants for the subcloning of gene 125 since the titers of the recombinant phage were generally below 5×10^8 PFU/ml of broth culture as compared with 10^{10} to 10^{11} for the remaining λEMBL3 recombinants. Based on the restriction enzyme analysis (Fig. 1), a different strategy was devised for obtaining gene 125. The DNA from *B. sphaericus* 2362 was cut with *Nhe*I-*Sac*I (Fig. 1), ligated into λGEM2, packaged as described above, and plated on *E. coli* LE392. The titer of the resulting preparation was 3×10^6 PFU in 0.5 ml, and approximately 80% of the phage contained recombinants. The resulting library was screened for hybridization with the gene 125 probe as described above; 28 positive plaques were obtained from 3.6×10^4 recombinants. All were characterized by restriction analysis and found to be identical. This λGEM2 recombinant was designated λGC5.1 (Fig. 1).

Subcloning of a smaller DNA fragment containing gene 125 from λGC5.1 proved to be a problem since it was not possible to obtain any recombinants in pGEM-4-blue which contained DNA 150 base pairs upstream of the second *Hin*dIII site (Fig. 1). A similar result was obtained by Yamagata et al. (46), and Belland and Trust (5). Yamagata et al. (46) were unable to clone the 5′ region of a related cell wall-associated protein gene (see Discussion) from *Bacillus*


-35
-10
RBS
M
A K Q N K G R K F F A A S A T A A L V A S A I V P V A S A A
Q L N D F N K I S G Y A K E A V Q S L V D A G V I Q G D A N
G N F N P L K T I S R A E A A T I F T N A L E L E A E G D V
N F K D V K A D A W Y Y D A I A A T V E N G I F E G V S A T
E F A P N K Q L T R S E A A K I L V D A F E L E G E G D L S
E F A D A S T V K P W A K S Y L E I A V A N G V I K G S E A
N G K T N L N P N A P I T R Q D F A V V F S R T I E N V D A
T P K V D K I E V V D A K T L N V T L S D G T K E T V T L E
K A L E P N K E T E V T F K I K D V E Y K A K V T Y V V T T
A T A V K S V S A T N L K E V V V E F D G T V D K E T A E D
A A N Y A L K S G K T I K S V S L A A D N K T A T V T L T D
K L N N N K A D A I S I S N V K A G D K E I N V K N V E F T
A V D N K I P E V T E V K S L G T K A V K V T L S E P V E N
L S S T N F T L D G K A Y F G N V V M G A G N K T V I L T P
Y S S S A L S V G D H K L T V S G A K D F A G F V S L N S T
H E F K V V E D K E A P T V T E A T A T L E T V T L T F S E
D I D M D T V K A S N V Y W K S G D S K K E A S E F E R I A
D N K Y K F V F K G S E K T L P T G K V D V Y V E D I K D Y
S D N K I A K D T K V T V T P E I D Q T R P E V R K V T A L
D E K T I K V T F S K T V D G E S A I K T G N Y T V K D K D
D K V V S V D K V T V D S K D S K S V I I D L Y S K V S V G
E N T I T I K N V K D A T K L N N T M L D Y T G K F T R S D
K E G P D Y E H V I N A D A K A K K V V L K F D K K M D A A
S L A D Y S N Y L V K I N D T L Q T L S E D V A T L S V S N
D A T V V T I T F A E T I K G D D V V F A S G K A I S G S G
K V N V N E L Q V M G V K D T S G N V H K K F N G S E N K I
T L S S T S T P L K L A K I D K D Y D A K Y T A E L V D R K
T V K V K F S T V I N S A A A N A F T S E S H K I D S I Q V

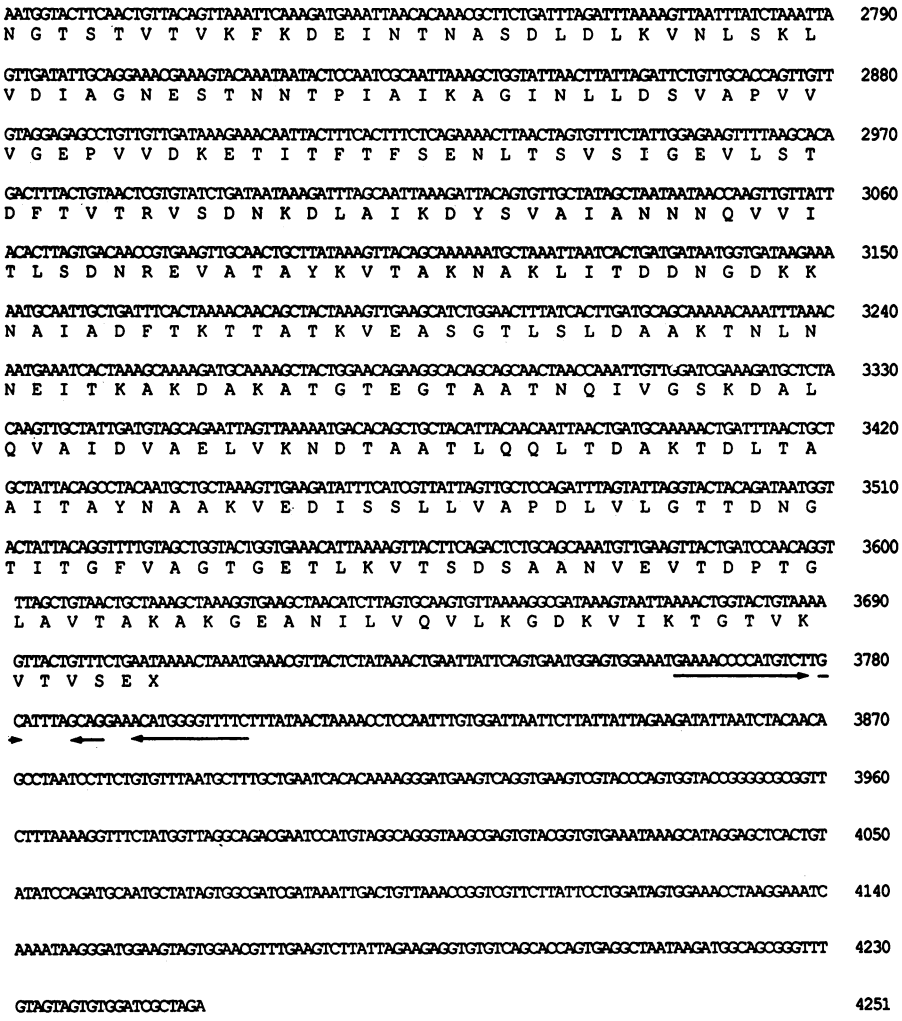


FIG. 2. Nucleotide sequence of the 4,251-base-pair pGC4.2 *B. sphaericus* DNA insert containing gene 125. The predicted amino acid sequence is given by the single-letter code. Lines labeled −35 and −10 designate the putative regions recognized by primary vegetative RNA polymerase. RBS, Putative ribosome-binding site; +, charged amino acid residues of the leader peptide; underlined amino acids, hydrophobic region of the leader peptide; ▲, the site of cleavage of the signal peptide; △, the beginning of the sequence of the related gene 80 ORF; arrows, inverted repeat.

brevis 47 into *E. coli*. Belland and Trust (5) were unable to subclone the 5′ region of the S-layer protein gene from *Aeromonas salmonicida* from a λgt11 clone into pBR322 or pUC18. Gene 125 was therefore reconstructed by first subcloning a 0.4-kb *Hin*dIII-*Xba*I fragment into pGEM-4-blue to give the recombinant pGC0.4 (Fig. 1). This plasmid was cut with *Xba*I and *Sac*I and ligated to the 3.8-kb *Xba*I-*Sac*I fragment from λGC5.1 (Fig. 1) to give the recombinant pGC4.2.

DNA sequencing. A more detailed restriction enzyme analysis was performed on the inserts of plasmids pGC4.2 (gene 125) and pGB6.2 (gene 80) (data not shown). Overlapping 1 to 2 kb DNA fragments were subcloned into M13mp18 and -19 by standard methods (31). A series of overlapping deletions were obtained with the IBI Cyclone System (International Biotechnologies, Inc., New Haven, Conn.) (11). In all cases, both strands of the DNA were sequenced by the dideoxy-chain termination method (37) with [^{35}S]dATP (Amersham Corp., Arlington Heights, Ill.) as the radioactive label.

Purification of 122- and 110-kDa proteins. The 122-kDa S-layer protein used for N-terminal amino acid sequence determination and immunological studies was purified as follows. Exponentially growing cells of *B. sphaericus* 2362 were harvested at an optical density at 620 nm of 0.2, and the cell walls were purified as described by Lewis et al. (29). The cell wall preparation was layered on a 48% NaBr solution

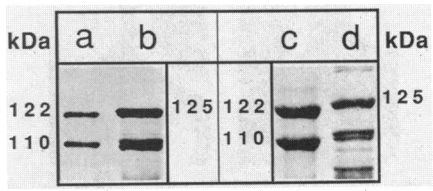


FIG. 3. Comparison of the gene 125 products made in *E. coli* by pGC4.2 with the 122- and 110-kDa S-layer proteins in the crystal preparation obtained from sporulated cells of *B. sphaericus* 2362. Lanes a and b, Western immunoblots, SDS-PAGE (7%); lanes c and d, SDS-PAGE (6%) stained with Coomassie blue. Lane a, 4 μg of crystal preparation; lane b, 30 μg of *E. coli* cell extract; lane c, 50 μg of crystal preparation; lane d, 100 μg of *E. coli* cell extract.

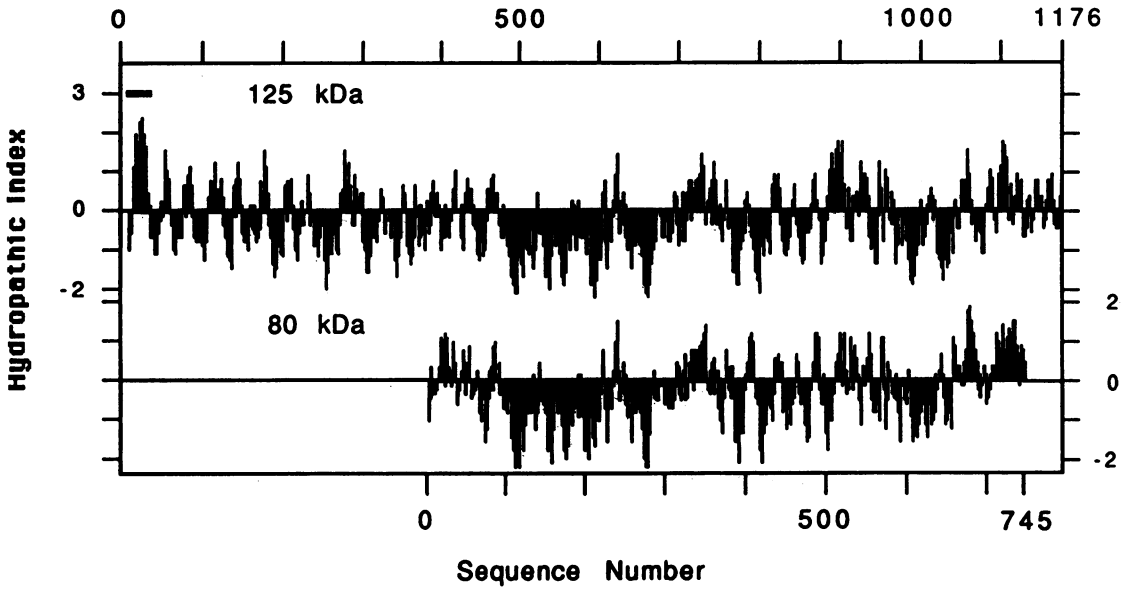


FIG. 4. Hydropathic analysis of the deduced amino acid sequence of the 125-kDa protein and the putative 80-kDa protein. The plots are aligned by the identical sequences. Solid bar indicates the hydrophobic region of the leader peptide.

and centrifuged at 17,000 × *g* for 3 h at 22°C in an HB-4 swinging-bucket rotor (Sorvall-DuPont, Newton, Conn.). The single top band contained the S-layer protein associated with the cell wall fragments. The purified cell wall suspension was dialyzed against water, and the S-layer protein was solubilized by treatment with 6 M urea (29) followed by dialysis against 10 mM NaH_2-Na_2HPO_4 buffer (pH 7.0). About 2 mg of the S-layer protein was obtained from a 1-liter culture.

The 122-kDa S-layer protein used for antiserum preparation was purified as described above with the sole difference that it was eluted from a nondenaturing polyacrylamide gel and dialyzed against 0.7% (wt/vol) NaCl.

The 122- and 110-kDa proteins from the crystal preparation (4) which were used for N-terminal amino acid sequence determination and immunological studies were purified by separation on SDS-PAGE (5% acrylamide) (4). The two regions of the gel containing the proteins were excised, and the proteins were electroeluted from the gel slice by use of the Elutrap Chamber (Schleicher & Schuell, Inc., Keene, N.H.) according to the directions of the manufacturer. SDS was removed by acetone precipitation (16), and the protein was dialyzed against 10 mM NaH_2-Na_2HPO_4 buffer (pH 7.0).

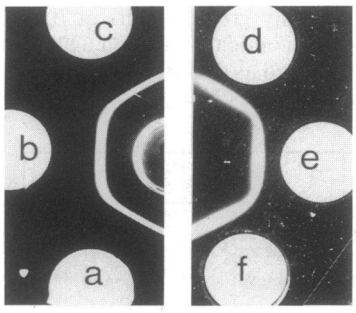


FIG. 5. Relationships of the 122- and 110-kDa proteins obtained from the crystal preparation, recombinant-produced 125-kDa protein, and the 122-kDa S-layer protein established by Ouchterlony immunodiffusion experiments. (a and d) cell-extract (200 μg) of *E. coli* containing pGC4.2 (125 kDa); (b and f) 20 μg of 122-kDa protein from crystal preparation; (c) 20 μg of 122-kDa S-layer protein; and (e) 20 μg of previously purified (9) 110-kDa protein. Center well, Anti-SC.

N-terminal sequence determination. With samples of 200 to 500 μg, the N-terminal sequences of the 122- and 110-kDa proteins from the crystal preparation and of the 122-kDa S-layer protein were determined by automated Edman degradation at the Protein Structure Laboratory of the University of California, Davis, School of Medicine (4).

Antisera. In this study, we used two antisera, of which the first was against a mixture of *B. sphaericus* 2362 S-layer and crystal proteins (122, 110, 51, and 42 kDa) (4, 9) (anti-SC), while the second was only against the 122-kDa S-layer protein (anti-S) from the same strain. For the latter, the antiserum was obtained from rabbits immunized as described by Lewis et al. (29). In the initial experiments involving 125-, 122-, and 110-kDa proteins, anti-SC was used. Subsequently selected experiments were confirmed by use of anti-S.

Electron microscopy. *B. sphaericus* 2362 was harvested during the exponential phase of growth at 30°C in NYSM broth (29) as well as from an overnight culture. The cells in 1.5 ml of the culture were suspended in 2% glutaraldehyde in 0.1 M NaH_2-Na_2HPO_4 (pH 7.0), stored at 4°C for 5 h, and washed twice in the buffer without glutaraldehyde. The preparation was suspended in 1% agar at 45°C and immediately centrifuged as the agar solidified. The agar containing the cells was cut into cubes and dehydrated in successive 10-min incubations in 30, 50, 80, 95, and 100% ethanol (the last performed three times). The samples were infiltrated overnight with a 1:1 Lowicryl K4M-ethanol mixture, for 8 h with a 3:1 Lowicryl-ethanol mixture, and for 4 days in 100% Lowicryl (all at 4°C). The infiltrated samples were polymerized under UV light for 3 days at room temperature, and sections were collected on nickel grids. Immunostaining was done by treatment with buffer A (20 mM Tris hydrochloride [pH 7.4], 500 mM NaCl, 0.3% Tween 80, 1% [wt/vol] bovine

serum albumin) for 10 min at room temperature followed by 2 h of incubation in anti-S (29) diluted 1/100 in buffer A. Grids were washed by immersion in buffer A lacking the albumin. They were subsequently incubated for 1.5 h in a solution containing *Staphylococcus aureus* protein A conjugated with 5 nM colloidal gold which had been diluted to a pale pink color in buffer A. The grids were washed twice in buffer A lacking albumin and twice in distilled water. Sections were poststained for 10 min in 1% (wt/vol) aqueous uranyl acetate and for 10 s in lead citrate.

RESULTS

Restriction analysis of DNA fragment containing gene 80 and gene 125. The restriction enzyme analysis of the λEMBL3 recombinants is presented in Fig. 1. The positions of gene 125 and gene 80 are indicated together with the DNA fragments which were subcloned into the plasmid vector pGEM-4-blue. The two genes were found to be within 11.2 kb of each other and were localized in a 32-kb DNA fragment.

Properties of gene 125. The sequence of a 4,251-base-pair DNA fragment containing gene 125 was determined (Fig. 2). This fragment contained an open reading frame (ORF) of 3,528 nucleotides (nt), beginning at nt 178 and ending at nt 3,705, which coded for a protein of 1,176 amino acids with a deduced molecular size of 125,085 daltons. The coding region is preceded by the sequence GGAGG (at nt 163) characteristic of a ribosome-binding site. The 12-nt spacing between the middle A of this sequence and the ATG initiation codon, as well as the free energy of binding (41) between this region of *B. sphaericus* mRNA and its counterpart in the *B. subtilis* 16S rRNA (−14.4 kcal/mol) is typical of that found in gram-positive bacteria (17). Beginning at nt 84 (Fig. 2) there is a putative −10 sequence, TATAAT, corresponding to the consensus −10 sequence utilized by the primary vegetative sigma factor of *Bacillus subtilis* and *E. coli* (13). The putative −35 sequence at nt 59 has the highly conserved TntG of the consensus TTGACA sequence and maintains the optimal 18-base-pair spacing (19). Following the ORF at nt 3,763 to 3,806 is an inverted repeat which is complementary in 17 of 19 nt having −29.2 kcal/mol as the free energy of binding (41). This G+C-rich hairpin loop is suggestive of a transcription termination signal.

Relation of 125-kDa protein to 122- and 110-kDa proteins. Figure 3 compares the migration of the 122- and 110-kDa proteins from the crystal preparation in SDS-PAGE and Western immunoblots with the proteins synthesized in *E. coli* TB1 containing the recombinant plasmid pGC4.2. The largest recombinant-produced protein (125 kDa) (lanes b and d) had a larger molecular size than the 122-kDa protein from the crystal preparation (lanes a and c). In *E. coli*, some degradation products (lane b) primarily consisting of proteins of 112 to 114 kDa and 95 kDa were present. No detectable difference was observed in the amount of the 125-kDa protein or degradation products in *E. coli* cells harvested from the exponential or stationary phase of growth (data not shown).

The results of a hydropathic analysis by the computerized method of Kyte and Doolittle (26) are presented in Fig. 4. The N-terminus of the 125-kDa protein is highly hydrophobic. Examination of the primary sequence indicates that the first 30 residues (Fig. 2) have characteristics of a leader peptide found in secreted proteins of gram-positive organisms (10). Among these properties are four positively charged residues followed by a hydrophobic stretch of 21 amino acids which at its end contains the potential cleavage signal sequence ASA. The 12 N-terminal amino acids of SDS-PAGE-purified 122-kDa protein from the S layer and 122- and 110-kDa proteins from the crystal preparation were determined and found to be identical (AQLNDFNKISGY). This is the sequence of the 125-kDa protein following the signal sequence ASA (Fig. 2, nt 267). The molecular size of the 125-kDa protein with the leader peptide removed is 122,129 daltons.

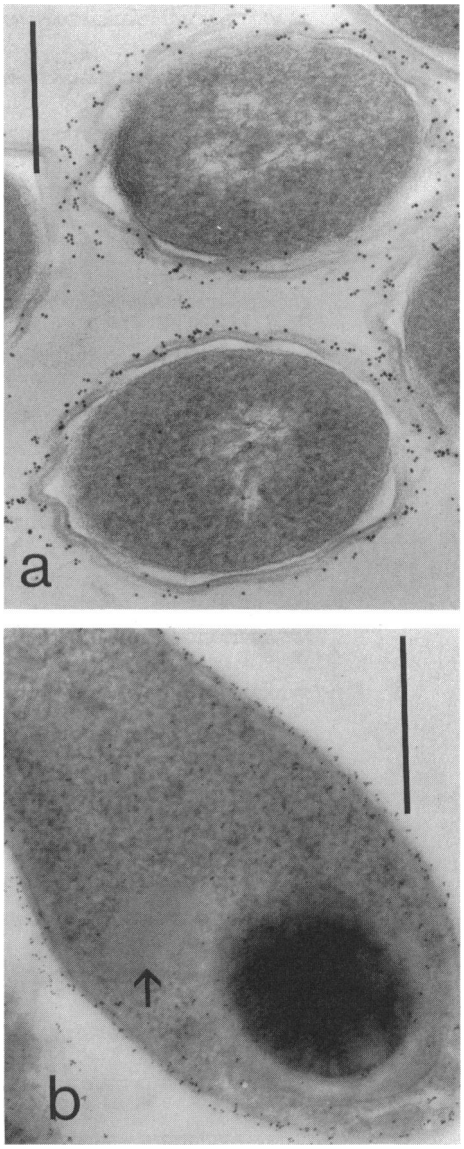


FIG. 6. Immunogold staining of thin sections of *B. sphaericus* 2362 with anti-S. (a) Cells in the exponential phase of growth. (b) Cells in the process of sporulation. Arrow in panel b designates the parasporal crystal. Bar, 0.5 μm.

The detection of a 125-kDa protein in *E. coli* cells containing the recombinant plasmid pGC4.2 (Fig. 3, lanes b and d) suggests that the leader peptide is not functional in this species. This is consistent with our inability to detect protein reacting with anti-SC in the supernatant fraction of the culture (tested by Western immunoblots, data not shown).

Ouchterlony immunodiffusion experiments (Fig. 5) showed a reaction of identity between the 125-kDa protein made in

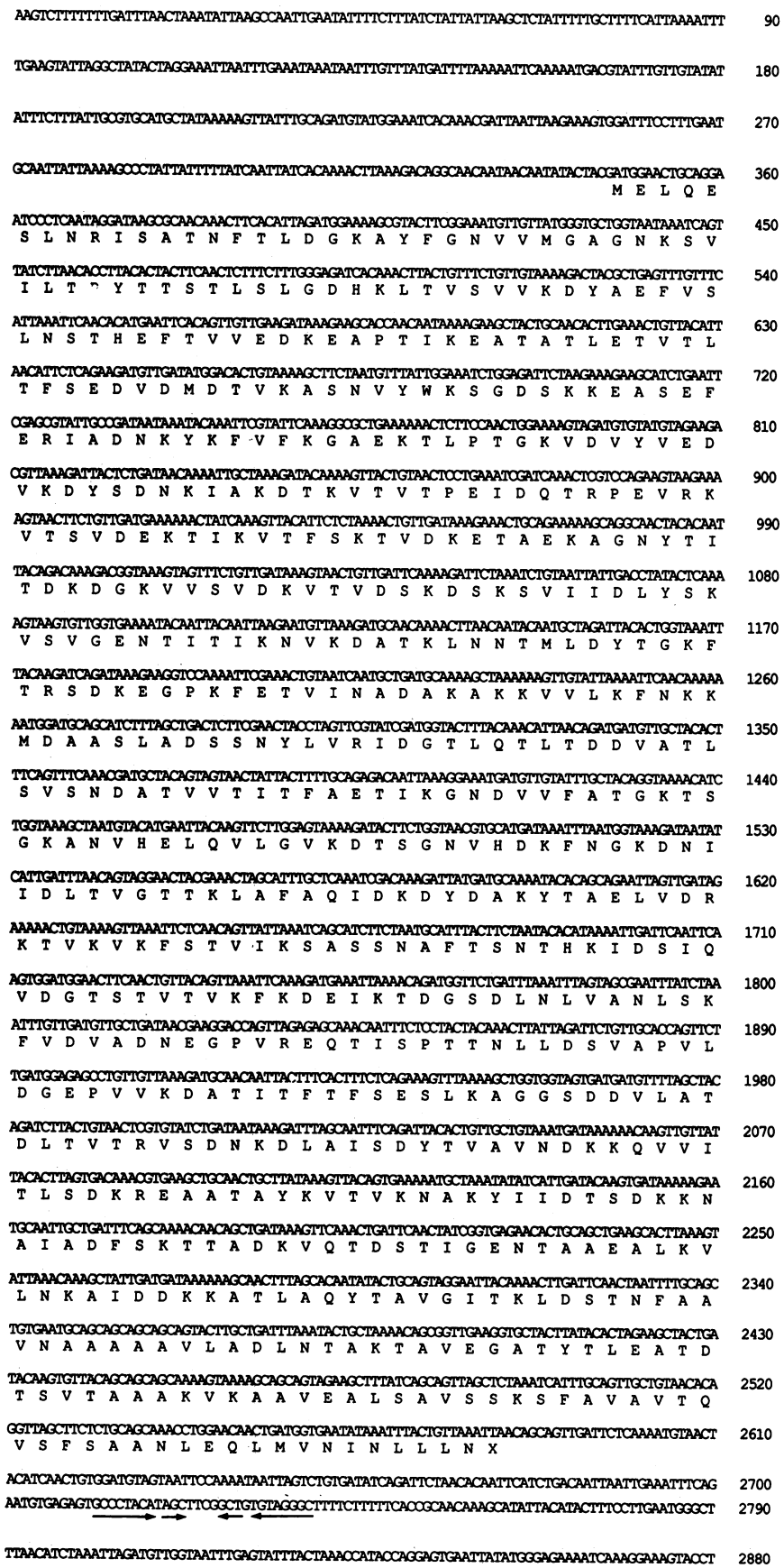


FIG. 7. Nucleotide sequence of the 2,880-base-pair region of *B. sphaericus* DNA containing gene 80. The predicted amino acid sequence is given in the single-letter code. Arrows designate an inverted repeat.

E. coli, the 122-kDa protein purified from the crystal preparation, the 122-kDa protein purified from the S layer, and the 110-kDa protein previously purified from the crystal preparation and shown to be toxic for larvae of *C. pipiens* (9). The results of these experiments are consistent with sequence identity or a high sequence similarity among these proteins. In addition, the identity of the N-terminal sequences of the 122- and 110-kDa proteins indicates that the former is the precursor of the latter and that the reduction in the molecular size is due to removal of amino acids at the C terminus.

Electron microscopy. Thin sections of *B. sphaericus* 2362 in the exponential phase of growth (Fig. 6a) or during the process of sporulation (Fig. 6b) reacted with anti-S predominantly at the region of the cell corresponding to the location of the S-layer protein. Anti-S did not react with the parasporal crystal (Fig. 6b), indicating that the 122- and 110-kDa proteins are not associated with the crystal.

Toxicity of 125- and 122-kDa proteins. Using second- to third-instar larvae of *C. pipiens*, we assayed the toxicity of *E. coli* cells containing pGC4.2 which accumulate the 125-kDa protein and proteins of 110 to 113 kDa (Fig. 3, lanes b and d). No toxicity was found up to a concentration of 67 μg (dry weight) of cells per ml. The purified cell wall fraction from exponentially growing cells of *B. sphaericus* 2362, which by SDS-PAGE was shown to contain only the 122-kDa S-layer protein, was also assayed. No toxicity was detected at a concentration of up to 20 μg of protein per ml.

Properties of gene 80. The sequence of a 2,880-base-pair DNA fragment containing gene 80 (Fig. 1) was determined (Fig. 7). One major ORF was detected (nt 346 to 2,580) which could code for a protein of 745 amino acids with a deduced molecular size of 80,024 daltons. The putative protein was not preceded by a recognizable ribosome-binding sequence. Following the ORF at nt 2,712 to 2,739 is an inverted repeat in which 11 of the 12 nt are complementary. The free energy of binding between these nucleotides was −22 kcal/mol (41). This hairpin structure together with the following row of Ts is suggestive of a transcription termination signal.

The amino acid sequence of the putative 80-kDa protein has extensive homology to that of the 122-kDa protein as indicated by the dot-matrix comparison presented in Fig. 8 (MacGene Plus; Applied Genetic Technology, Inc., Fairview Park, Ohio). Except for a portion of 127 amino acids at the C terminus, the similarity extends throughout the molecule. The region between amino acids 13 and 618 of the putative 80-kDa protein (Fig. 7) has an 81% sequence similarity to the region between amino acids 365 and 974 of the 122-kDa protein (Fig. 2). The sole changes required for this comparison were the introduction into the putative 80-kDa protein sequence of a 3-amino-acid gap at residue 364 and a 1-amino-acid gap at residue 505 (Fig. 7). Comparison of the nucleotide sequence within the same region indicated a similarity of 88%. A hydropathic analysis of the 80-kDa protein is presented in Fig. 4. The results showed a major similarity in the hydrophobicity profiles of the 80- and 125-kDa proteins, with differences residing at the C termini of these two molecules.

Western immunoblots indicate that the gene coding for the putative 80-kDa protein, contained on pGB6.2 (Fig. 1), is not expressed in *E. coli* (data not shown). Similarly, this fragment (Fig. 1) when cloned into *B. subtilis* DB104 or *B. sphaericus* SSII-1 did not result in the production of a detectable protein in Western immunoblots, in contrast to the DNA fragment contained on pGC4.2 (Fig. 1), which when cloned into these species produced the 122-kDa gene

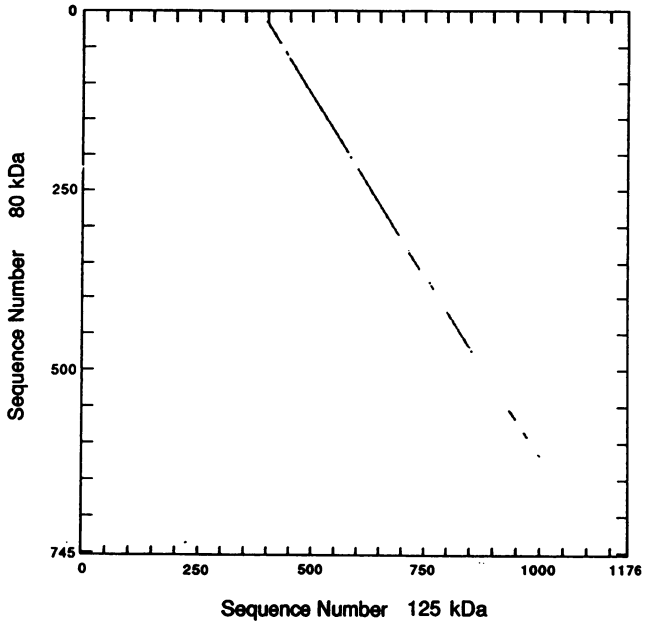


FIG. 8. Comparison of the 125-kDa protein and the putative 80-kDa protein by a dot-matrix plot. The window was 10 amino acids, and the stringency was 90%.

product (R. D. Bowditch, Ph.D. thesis, University of California, Davis, 1989). These results and the absence of a discrete 80-kDa protein band in Western immunoblots of vegetative and sporulating cells of *B. sphaericus* 2362 leads to the conclusion that the gene coding for the 80-kDa protein is cryptic.

Codon usage. Table 1 compiles the codon usage for the genes coding for the 42- and 51-kDa crystal proteins, as well as the 125-kDa S-layer protein precursor of *B. sphaericus* 2362. The codon usages for these three genes are compared with those of *E. coli* and *B. subtilis* (36). As would be expected from the low guanine-plus-cytosine content of the DNA of *B. sphaericus* (25) (35 mol%), there is generally a strong bias for A or U in the third position.

DISCUSSION

The results of this study indicate that gene 125 codes for a 125-kDa protein which contains a leader peptide (Fig. 2) and is the precursor of the 122-kDa S-layer protein. The latter is the precursor of the 110-kDa larvicidal protein (9) which appears during sporulation of *B. sphaericus* 2362. This conclusion is based on the reaction of identity among these proteins in Ouchterlony immunodiffusion experiments (Fig. 5), the sequence identity of the N-terminal portion of these molecules, and the reaction of anti-S with the cell wall but not the parasporal crystal (Fig. 6) of *B. sphaericus* 2362.

In *B. sphaericus*, the S-layer protein is synthesized during vegetative growth (29), a fact consistent with the presence of a putative promoter having a −10 sequence identical to the consensus sequence recognized by the primary vegetative RNA polymerase 88 nt upstream of the ORF for the 125-kDa protein (Fig. 2). This promoter may also be recognized by the primary *E. coli* RNA polymerase, allowing its expression in this organism. In *E. coli*, the 125-kDa protein accumulated in the cell (Fig. 3) and was not exported into the medium, indicating that the leader peptide is not functional in this species. In *B. sphaericus*, the 122-kDa protein is associated

TABLE 1. Codon usage for the genes encoding the crystal proteins

Amino acid	Codon	Gene product (10^3 mol wt)			F_{Bsph}[b]	F_{Bsub}[c]	F_{Ecol}[c]
		42[a]	51[a]	125			
Phe	UUU	16	16	11	0.55	0.57	0.37
Phe	UUC	3	5	27	0.45	0.43	0.63
Leu	UUA	11	16	53	0.60	0.24	0.07
Leu	UUG	4	6	0	0.07	0.14	0.09
Leu	CUU	3	7	13	0.17	0.25	0.07
Leu	CUC	1	1	0	0.01	0.10	0.07
Leu	CUA	2	8	8	0.13	0.06	0.02
Leu	CUG	1	0	0	0.01	0.21	0.68
Ile	AUU	12	15	39	0.52	0.50	0.36
Ile	AUC	8	4	20	0.25	0.37	0.61
Ile	AUA	13	14	2	0.23	0.13	0.03
Met	AUG	8	7	6	1.00	1.00	1.00
Val	GUU	8	9	79	0.56	0.31	0.36
Val	GUC	1	0	0	0.01	0.24	0.15
Val	GUA	5	11	49	0.38	0.21	0.22
Val	GUG	1	1	6	0.05	0.24	0.27
Ser	UCU	5	8	47	0.40	0.23	0.23
Ser	UCC	5	3	0	0.05	0.09	0.27
Ser	UCA	8	10	26	0.29	0.20	0.07
Ser	UCG	1	1	2	0.03	0.11	0.11
Ser	AGU	6	9	7	0.15	0.11	0.06
Ser	AGC	6	2	4	0.08	0.26	0.26
Pro	CCU	9	17	6	0.46	0.30	0.12
Pro	CCC	3	1	0	0.06	0.09	0.07
Pro	CCA	9	6	16	0.45	0.21	0.16
Pro	CCG	1	1	0	0.03	0.40	0.65
Thr	ACU	19	10	74	0.52	0.15	0.25
Thr	ACC	5	4	0	0.05	0.16	0.50
Thr	ACA	14	12	54	0.40	0.45	0.07
Thr	ACG	2	4	2	0.04	0.24	0.18
Ala	GCU	2	6	73	0.49	0.27	0.26
Ala	GCC	2	0	1	0.02	0.22	0.21
Ala	GCA	7	10	55	0.44	0.29	0.22
Ala	GCG	3	3	3	0.05	0.22	0.31
Tyr	UAU	18	21	6	0.62	0.65	0.40
Tyr	UAC	4	6	18	0.38	0.35	0.60
His	CAU	6	5	5	0.84	0.67	0.54
His	CAC	1	2	0	0.16	0.33	0.46
Gln	CAA	11	14	16	0.87	0.52	0.24
Gln	CAG	2	4	0	0.13	0.48	0.76
Asn	AAU	20	28	38	0.62	0.56	0.26
Asn	AAC	6	7	40	0.38	0.44	0.74
Lys	AAA	8	24	120	0.89	0.73	0.76
Lys	AAG	6	4	8	0.11	0.27	0.24
Asp	GAU	17	21	69	0.78	0.61	0.46
Asp	GAC	2	3	26	0.22	0.39	0.54
Glu	GAA	14	24	66	0.83	0.70	0.73
Glu	GAG	8	6	7	0.17	0.30	0.27
Cys	UGU	3	3	0	0.67	0.68	0.43
Cys	UGC	0	3	0	0.33	0.32	0.57
Trp	UGG	3	3	3	1.00	1.00	1.00
Arg	CGU	1	1	8	0.21	0.22	0.56
Arg	CGC	2	1	1	0.08	0.19	0.36
Arg	CGA	1	4	0	0.10	0.12	0.03
Arg	CGG	0	2	0	0.04	0.14	0.03
Arg	AGA	9	9	3	0.44	0.22	0.01
Arg	AGG	2	4	0	0.13	0.11	0.01
Gly	GGU	4	7	32	0.42	0.24	0.48
Gly	GGC	2	2	11	0.15	0.31	0.39
Gly	GGA	10	12	14	0.35	0.32	0.05
Gly	GGG	6	1	2	0.09	0.13	0.08
Ter	UAA	1	0	1	0.67	0.71	0.75
Ter	UAG	0	0	0	0.00	0.08	0.08
Ter	UGA	0	1	0	0.33	0.21	0.17

[a] From sequence in reference 3.
[b] Total codons for *B. sphaericus* 2362 = 1,997.
[c] *B. subtilis* and *E. coli* usage from reference 36.

with the cell wall (28, 29, 45; this study). Upon completion of exponential growth, there is a burst of synthesis of the 122-kDa protein by *B. sphaericus* (9), suggesting a complex pattern of regulation for the expression of gene 125. Adachi et al. (1) found that the expression of cell wall proteins of *B. brevis* 47 is regulated by three tandemly arranged promoters. Evidence was presented for differential utilization of these promoters during the growth and sporulation cycle of *B. brevis* 47.

The lack of toxicity for the larvae of *C. pipiens* of the cell wall-associated 122-kDa protein obtained from exponentially growing cells and the toxicity of the 110-kDa protein purified from the crystal preparation raise the possibility that the proteolytic degradation of the 122-kDa protein to the 110-kDa protein observed during sporulation (9) is a conversion of a protoxin to a toxin. Previous studies (33) with the related strain 1593 also indicated that vegetative cells were not toxic for mosquito larvae. The toxicity of cell wall preparations obtained from strain 1593 (32) could be the consequence of the presence of a 110-kDa protein since the cultures from which the preparations were obtained were in the early stationary phase, a time when some 110-kDa protein would be expected to be present (9). A similar interpretation may account for the "cytoplasmic toxin" of Davidson (12), which was found to have a molecular size of about 100 kDa. The absence of toxicity of *E. coli* cells containing proteins of 110 to 113 kDa derived from the 125-kDa protein (Fig. 3) may be due to the absence or difference in the pattern of glycosylation of these proteins compared with *B. sphaericus* or due to the removal of amino acids at the N terminus.

Amino acid sequence comparisons of the 122-kDa protein with the S-layer proteins from *Halobacterium halobium* (27) and *Deinococcus radiodurans* (35) revealed no primary sequence similarity. Comparisons with the sequence of the "outer wall protein" and "middle wall protein" of *B. brevis* 47 (42, 43), which form the S layer in this species, indicated that the 122-kDa protein had a significant sequence similarity to the N-terminal portion of the "outer wall protein" (Fig.

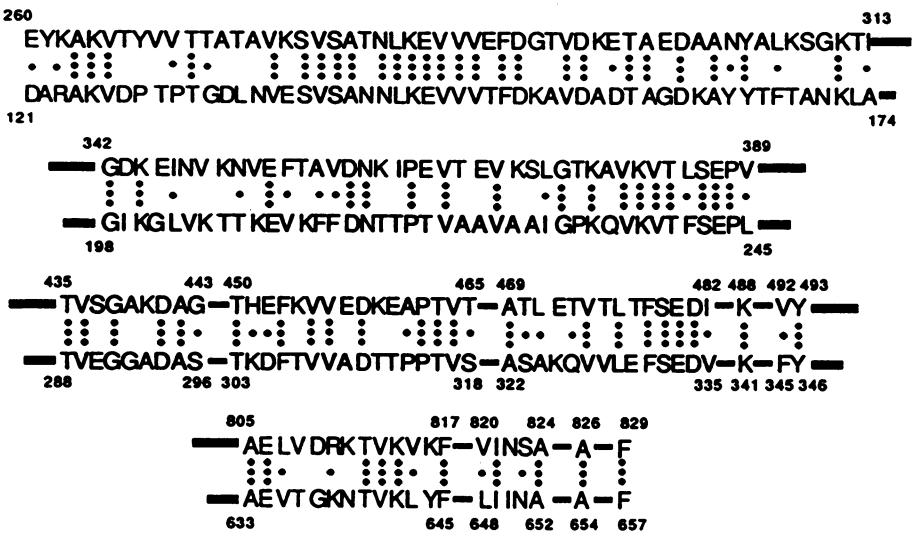


FIG. 9. Amino acid sequence similarity of the 122-kDa protein from *B. sphaericus* and the "outer wall protein" of *B. brevis* (43). The top sequence is from *B. sphaericus* 2362, and the bottom sequence is from *B. brevis* 47. Numbers indicate the position of the amino acids in the sequence. Three dots indicate identical amino acids, and single dots indicate conserved amino acid substitutions. Horizontal bars of unequal length represent adjustments necessary for alignment.

9). As for the 42- and 51-kDa proteins, the 122-kDa protein had no sequence similarity to any of the published Diptera-, Lepidoptera-, or Coleoptera-active toxins of *Bacillus thuringiensis* (8, 14, 20, 38, 44; and references in reference 3) or to the 42- and 51-kDa proteins from *B. sphaericus* 2362. All the published sequences of the *B. thuringiensis* lepidoptera- and diptera-active crystal proteins have some sequence homology, suggesting a common evolutionary origin. In contrast, the 122-kDa protein of *B. sphaericus* is related to at least one S-layer protein, while the related 42- and 51-kDa crystal proteins (3) appear to be distinct. The putative 80-kDa protein coded for by a cryptic gene is highly related to a major portion of the 122-kDa protein, suggesting a previous gene duplication (Fig. 8). Evidence for a cryptic gene having a sequence related to the Lepidoptera-active crystal proteins from *B. thuringiensis* has been previously presented (24).

The present study completes the genetic identification and characterization of the genes of *B. sphaericus* coding for proteins which had been previously found to be toxic for mosquito larvae (4, 7, 9, 12, 47). In our first publication on the crystal proteins of *B. sphaericus* 2362, we obtained antisera to the 42- and 51-kDa proteins and using these antisera detected a cross-reaction with the 122- and 110-kDa proteins in Western immunoblots (4). In view of the lack of sequence similarity between the low- and high-molecular-weight proteins revealed in this study, these results can only be interpreted as indicating that the 42- and 51-kDa proteins originally purified from the crystal preparation and used to raise antisera contained degradation products of the 122- and 110-kDa proteins similar in molecular size to the 42- and 51-kDa proteins.

ACKNOWLEDGMENTS

This investigation was supported by the Vector Biology and Control Components of the United Nations Development Program/WorldBank/World Health Organization Special Programme for Research and Training in Tropical Diseases, the University of California Mosquito Research Program, and the University of California Agricultural Experiment Station.

LITERATURE CITED

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. **Baumann, P., L. Baumann, R. D. Bowditch, and A. H. Broadwell.** 1987. Cloning of the gene for the larvicidal toxin of *Bacillus sphaericus* 2362: evidence for a family of related sequences. J. Bacteriol. **169:**4061–4067.
3. **Baumann, L., A. H. Broadwell, and P. Baumann.** 1988. Sequence analysis of the mosquitocidal toxin genes encoding 51.4- and 41.9-kilodalton proteins from *Bacillus sphaericus* 2362 and 2297. J. Bacteriol. **170:**2045–2050.
4. **Baumann, P., B. M. Unterman, L. Baumann. A. H. Broadwell, S. J. Abbene, and R. D. Bowditch.** 1985. Purification of the larvicidal toxin of *Bacillus sphaericus* and evidence for high-molecular-weight precursors. J. Bacteriol. **163:**738–747.
5. **Belland, R. J., and T. J. Trust.** 1987. Cloning of the gene for the surface array protein of *Aeromonas salmonicida* and evidence linking loss of expression with genetic deletion. J. Bacteriol. **169:**4086–4091.
6. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196:**180–182.
7. **Bourgouin, C., R. Tinelli, J.-P. Bouvet, and R. Pires.** 1984. *Bacillus sphaericus* 1593-4. Purification of fractions toxic for mosquito larvae, p. 387–388. *In* J. E. Alouf, F. J. Fehrenbach, J. F. Freer, and J. Jeljaszewicz (ed.), Bacterial protein toxins. Academic Press, Inc. (London), Ltd., London.
8. **Brizzard, B. L., and H. R. Whiteley.** 1988. Nucleotide sequence of an additional crystal protein from *Bacillus thuringiensis* subsp. *thuringiensis*. Nucleic Acids Res. **16:**2723–2724.
9. **Broadwell, A. H., and P. Baumann.** 1986. Sporulation-associated activation of *Bacillus sphaericus* larvicide. Appl. Environ. Microbiol. **52:**758–764.
10. **Chang, S.** 1987. Engineering for protein secretion in gram-positive bacteria. Methods Enzymol. **153:**507–516.
11. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13:**31–40.
12. **Davidson, E. W.** 1982. Purification and properties of soluble cytoplasmic toxin from the mosquito pathogen *Bacillus sphaeri-*

cus strain 1593. J. Invertebr. Pathol. **39**:6–9.
13. **Doi, R. H., and L.-F. Wang.** 1986. Multiple procaryotic ribonucleic acid polymerase sigma factors. Microbiol. Rev. **50**:227–243.
14. **Donovan, W. P., C. Dankocsik, and M. P. Gilbert.** 1988. Molecular characterization of a gene encoding a 72-kilodalton mosquito-toxic crystal protein from *Bacillus thuringiensis* subsp. *israelensis*. J. Bacteriol. **170**:4732–4738.
15. **Frischauf, A.-M., H. Lehrach, A. Poustka, and N. Murray.** 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**:827–842.
16. **Hagar, D. A., and R. R. Burgess.** 1980. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. Anal. Biochem. **109**:76–86.
17. **Hager, P. W., and J. C. Rabinowitz.** 1985. Translational specificity in *Bacillus subtilis*, p. 1–32. *In* D. A. Dubnau (ed.), The molecular biology of the bacilli, vol. 2. Academic Press, Inc., Orlando, Fla.
18. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
19. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15**:2343–2361.
20. **Herrnstadt, C., T. E. Gilroy, D. A. Sobieski, B. D. Bennett, and F. H. Gaertner.** 1987. Nucleotide sequence and deduced amino acid sequence of a coleopteran-active delta-endotoxin gene from *Bacillus thuringiensis* subsp. *san diego*. Gene **57**:37–46.
21. **Hindley, J., and C. Berry.** 1987. Identification, cloning and sequence analysis of the *Bacillus sphaericus* 1593 41.9 kD larvicidal toxin gene. Mol. Microbiol. **1**:187–194.
22. **Ishiguro, E. E., W. W. Kay, T. Ainsworth, J. B. Chamberlain, J. T. Buckley, and T. J. Trust.** 1981. Loss of virulence during culture of *Aeromonas salmonicida* at high temperature. J. Bacteriol. **148**:333–340.
23. **Kaiser, K., and N. E. Murray.** 1985. The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries, p. 1–47. *In* D. M. Glover (ed.), DNA cloning: a practical approach, vol. 1. IRL Press, Oxford.
24. **Klier, A., F. Fargette, J. Ribier, and G. Rapoport.** 1982. Cloning and expression of the crystal protein genes from *Bacillus thuringiensis* strain *berliner* 1715. EMBO J. **1**:791–799.
25. **Krych, V. K., J. L. Johnson, and A. A. Yousten.** 1980. Deoxyribonucleic acid homologies among strains of *Bacillus sphaericus*. Int. J. Syst. Bacteriol. **30**:476–484.
26. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
27. **Lechner, J., and M. Sumper.** 1987. The primary structure of a procaryotic glycoprotein. J. Biol. Chem. **262**:9724–9729.
28. **Lewis, L. O., and A. A. Yousten.** 1988. Bacteriophage attachment to the S-layer proteins of the mosquito-pathogenic strains of *Bacillus sphaericus*. Curr. Microbiol. **17**:55–60.
29. **Lewis, L. O., A. A. Yousten, and R. G. E. Murray.** 1987. Characterization of surface protein layers of the mosquito-pathogenic strains of *Bacillus sphaericus*. J. Bacteriol. **169**: 72–79.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
32. **Myers, P. S., and A. A. Yousten.** 1980. Localization of a mosquito-larval toxin of *Bacillus sphaericus* 1593. Appl. Environ. Microbiol. **39**:1205–1211.
33. **Myers, P., A. A. Yousten, and E. W. Davidson.** 1979. Comparative studies of the mosquito-larval toxin of *Bacillus sphaericus* SSII-1 and 1593. Can. J. Microbiol. **25**:1227–1231.
34. **Palazzolo, M. J., and E. M. Meyerowitz.** 1987. A family of lambda phage cDNA cloning vectors, λSWAJ, allowing the amplification of RNA sequences. Gene **52**:197–206.
35. **Peters, J., M. Peters, F. Lottspeich, W. Schafer, 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.
36. **Piggot, P. J., and J. A. Hoch.** 1985. Revised genetic linkage map of *Bacillus subtilis*. Microbiol. Rev. **49**:158–179.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Sen, K., G. Honda, N. Koyama, M. Nishida, A. Naki, H. Sakai, M. Himeno, and T. Komano.** 1988. Cloning and nucleotide sequences of two 130 kDa insecticidal protein genes of *Bacillus thuringiensis* var. *israelensis*. Agric. Biol. Chem. **52**:873–878.
39. **Sleytr, U. B., and P. Messner.** 1988. Crystalline surface layers in procaryotes. J. Bacteriol. **170**:2891–2897.
40. **Smit, J.** 1986. Protein surface layers of bacteria, p. 343–376. *In* M. Inouye (ed.), Outer membranes as model systems. John Wiley & Sons, Inc., New York.
41. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
42. **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 protein gene. J. Bacteriol. **170**:935–945.
43. **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.
44. **Ward, E. S., and D. J. Ellar.** 1987. Nucleotide sequence of a *Bacillus thuringiensis* var. *israelensis* gene encoding a 130 kDa delta-endotoxin. Nucleic Acids Res. **15**:7195.
45. **Word, N. S., A. A. Yousten, and L. Howard.** 1983. Regularly structured and non-regularly structured surface layers of *Bacillus sphaericus*. FEMS Microbiol. Lett. **17**:277–282.
46. **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.
47. **Yousten, A. A.** 1984. *Bacillus sphaericus*: microbiological factors related to its potential as a mosquito larvicide. Adv. Biotechnol. Processes **3**:315–343.