JOURNAL OF BACTERIOLOGY, July 1987, p. 3088–3093
0021-9193/87/073088-06$02.00/0

Vol. 169, No. 7

Cloning and Nucleotide Sequencing of Genes for a Second Type of Small, Acid-Soluble Spore Proteins of *Bacillus cereus*, *Bacillus stearothermophilus*, and "*Thermoactinomyces thalpophilus*"

DONGXU SUN AND PETER SETLOW*

Department of Biochemistry, University of Connecticut Health Center, Farmington, Connecticut 06032

Received 24 February 1987/Accepted 22 April 1987

The nucleotide sequences of the single genes coding for the B-type small, acid-soluble spore proteins (SASP) of *Bacillus cereus*, *B. stearothermophilus*, and "*Thermoactinomyces thalpophilus*" were determined, and the amino acid sequences of all B-type SASP were compared. While this type of SASP showed significant sequence conservation around the two spore protease cleavage sites, alignment of these sequences required the introduction of gaps, and even then only 19 of the residues were conserved exactly in all five proteins. However, all five B-type SASP did contain a large (27 to 35-residue), rather well-conserved amino acid sequence repeat, and four of the five proteins had well-conserved regions of 14 to 17 amino acids which appeared three times.

Up to 20% of the protein of spores of *Bacillus* species is degraded in the first few minutes of spore germination, thus providing amino acids for both new protein synthesis and metabolism (3, 11). The proteins degraded in this process are a group of small, acid-soluble spore proteins (SASP) which are synthesized only during sporulation under transcriptional control (3, 11). In *Bacillus megaterium*, three types of SASP (termed A, B, and C) make up ~85% of the protein degraded, with A- and C-type SASP being very similar in primary sequence and B-type SASP being more different (3, 4). Studies of this system at the gene level in a number of *Bacillus* species as well as in "*Thermoactinomyces thalpophilus*" have revealed that these organisms contain at least seven different genes coding for SASP which are extremely similar to A- and C-type SASP (2, 3, 8). Strikingly, the amino acid sequences of the proteins coded for by the A- and C-type SASP genes from different species exhibit a very high degree of homology, since alignment of these sequences requires the introduction of no gaps (2, 3, 8) and since 27 of the residues in these 15 proteins (61 to 72 residues each) are conserved exactly across species (8). While the reason(s) for the high degree of sequence conservation in the A- and C-type SASP is not completely clear, it is possible that it is related to the crucial role that this group of proteins appears to play in the resistance of spores to UV light (9).

In contrast to the A- and C-type SASP, there appears to be only a single gene which codes for B-type SASP (6, 7). In addition, a comparison of primary sequences of B-type SASP from *B. megaterium* and *B. subtilis* has indicated that these proteins show much less sequence conservation across species than do the A- and C-type SASP (7). Consequently, we felt it might be valuable to examine B-type SASP genes from organisms more distantly related than are *B. megaterium* and *B. subtilis*, in that a comparison of primary sequences of B-type SASP might indicate which regions or features of these proteins or both are nonessential for their function. In this communication we report the cloning and nucleotide sequencing of B-type SASP genes from *B. cereus*, *B. stearothermophilus*, and "*T. thalpophilus*."

MATERIALS AND METHODS

Organisms used and isolation of DNA. The organisms used were *B. cereus* T (originally obtained from H. O. Halvorson) and "*T. thalpophilus*" HA-01 (5). DNA was isolated and purified from *B. cereus* and "*T. thalpophilus*" as described previously (5, 8); DNA from *B. stearothermophilus* ATCC 7953 was a gift from H. Cudny. The sources of the *Escherichia coli* strains and plasmids were described previously (2, 4, 8). Plasmid-carrying strains were grown in 2 × YT medium (8) with appropriate antibiotics (ampicillin, 50 μg/ml; tetracycline, 10 μg/ml; and chloramphenicol, 10 μg/ml), and plasmid DNA was isolated by the method of Birnboim and Doly (1). If necessary, plasmid DNA was purified by two cycles of CsCl gradient centrifugation.

Identification, cloning, and sequencing of SASP genes. Restriction enzyme digests of genomic DNA (2 μg) were run in agarose gel electrophoresis, and fragments were transferred to nitrocellulose (12). These blots were hybridized with a variety of SASP gene probes: (i) a *B. megaterium* B-type SASP gene probe, a 0.65-kilobase (kb) *Mbo*I fragment containing the coding sequence for *B. megaterium* B-type SASP (6); (ii) a partial *B. cereus* B-type SASP gene probe, a 0.4-kb *Hin*dIII-*Xba*I fragment containing the carboxyl-terminal portion of the coding sequence for *B. cereus* B-type SASP (see Fig. 2); (iii) a complete *B. cereus* B-type SASP gene probe, a 0.3-kb *Dde*I fragment containing almost the complete coding sequence for *B. cereus* B-type SASP (see Fig. 2); (iv) a *B. stearothermophilus* B-type SASP gene probe, a 0.55-kb *Dra*I-*Hin*dIII fragment containing the complete coding sequence for *B. stearothermophilus* B-type SASP (see Fig. 2); and (v) a "*T. thalpophilus*" B-type SASP gene probe, a 0.3-kb *Hin*dIII-*Taq*I fragment containing almost the complete coding sequence for "*T. thalpophilus*" B-type SASP (see Fig. 2). Hybridization of these blots was done at 48, 55, or 70°C under the conditions described previously (2, 7).

For cloning of B-type SASP genes, genomic DNA (50 μg)

* Corresponding author.

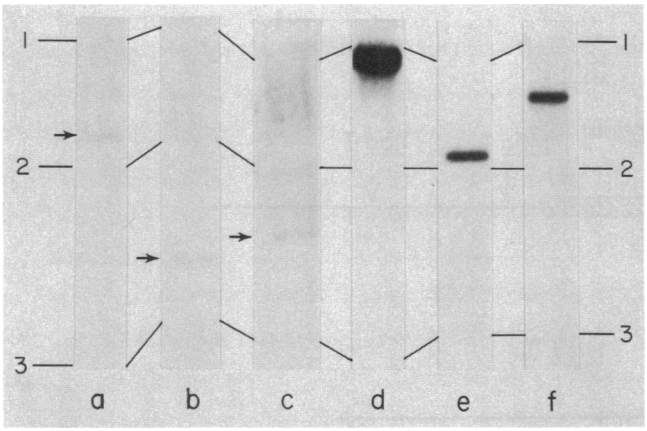

FIG. 1. Southern blots of genomic DNA. Genomic DNA was cut with various restriction enzymes, samples (2 μg) were run on a 1% agarose gel, the DNA was transferred to nitrocellulose, and the blots were hybridized as described in Materials and Methods. Lanes: a, *B. cereus* DNA cut with *Hin*dIII and hybridized to the *B. megaterium* B-type SASP gene probe at 55°C; b, *B. stearothermophilus* DNA cut with *Hin*dIII and hybridized to the *B. megaterium* B-type SASP gene probe at 48°C; c, "*T. thalpophilus*" DNA cut with *Hin*dIII and hybridized to the *B. megaterium* B-type SASP gene probe at 48°C; d, *B. cereus* DNA cut with *Bcl*I and hybridized to the complete *B. cereus* B-type SASP gene probe at 55°C; e, *B. stearothermophilus* DNA cut with *Pst*I and hybridized to the *B. stearothermophilus* B-type SASP gene probe at 55°C; and f, "*T. thalpophilus*" DNA cut with *Pvu*II and hybridized to the "*T. thalpophilus*" B-type SASP gene probe at 55°C. The lines numbered 1 to 3 show the positions of marker DNAs of 6.7, 2.3, and 0.6 kb, respectively. The bands marked with arrows in lanes a, b, and c were those found to contain B-type SASP genes.

was digested with *Hin*dIII, and fragments of appropriate size were isolated as previously described (4). Fragments were initially cloned in plasmid pBR325, and clones containing B-type SASP genes were identified by Southern blot analysis of plasmids from ~500 clones containing insert fragments of appropriate size (8). Clones were grown individually in 1-ml cultures, pooled in groups of five, subjected to plasmid isolation, cleaved with *Hin*dIII, and run in agarose gel electrophoresis. Fragments were then transferred to nitrocellulose, and the blots were hybridized to the *B. megaterium* B-type SASP gene probe at 48°C ("*T. thalpophilus*" and *B. stearothermophilus* genes) or 55°C (*B. cereus* gene). The complete B-type SASP gene of *B. cereus* was cloned in plasmid pUC12 and identified by colony hybridization at 70°C with the partial *B. cereus* B-type SASP gene probe.

The hybridizing regions in cloned DNA fragments were localized by restriction enzyme digestion and Southern blotting as previously described (2, 4), and subfragments were cloned in pUC plasmid vectors. DNA sequence analysis was carried out by the method of Maxam and Gilbert (10); all sequences were determined completely in both directions, and all restriction sites used in sequencing were overlapped.

TABLE 1. B-type SASP genes cloned from *B. cereus*, *B. stearothermophilus*, and "*T. thalpophilus*"[a]

Organism	Size (kb) of fragment cloned (restriction enzyme used)	Plasmid that fragment was cloned in
B. cereus	2.8 (*Hin*dIII)[b]	pBR325
B. cereus	1.8 (*Xba*I)[c]	pUC12
B. stearothermophilus	0.9 (*Hin*dIII)	pBR325
"*T. thalpophilus*"	1.3 (*Hin*dIII)	pBR325

[a] Fragments were identified and cloned as described in Materials and Methods.
[b] This fragment contained only the carboxyl-terminal coding region of the B-type SASP gene.
[c] This fragment contained a complete B-type SASP gene coding sequence.

RESULTS AND DISCUSSION

Cloning of B-type SASP genes from *B. cereus*, *B. stearothermophilus*, and "*T. thalpophilus*." Previous work has demonstrated that spores of *B. cereus*, *B. stearothermophilus*, and "*T. thalpophilus*" contain a predominant SASP which cross-reacts with antiserum to *B. megaterium* B-type SASP (8). This result suggested that the B-type SASP genes from these different organisms might be sufficiently similar to allow their detection by using a heterologous B-type SASP gene as a hybridization probe. Indeed, hybridization of Southern blots of *Hin*dIII digests of genomic DNAs from *B. cereus*, *B. stearothermophilus*, and "*T. thalpophilus*" with a *B. megaterium* B-type SASP gene probe revealed only a single major hybridizing fragment (Fig. 1, lanes a, b and c, bands marked with arrows). These three fragments were cloned, and the regions hybridizing to the *B. megaterium* B-type SASP gene probe were localized (Table 1 and Fig. 2). Subsequent nucleotide sequence analyses showed that the *Hin*dIII fragment from *B. cereus* contained only a part of the B-type SASP gene. Consequently, the 0.4-kb *Hin*dIII-*Xba*I fragment (Fig. 2) was used as a probe to isolate the complete *B. cereus* B-type SASP gene on an *Xba*I fragment (Table 1 and Fig. 2).

Nucleotide sequences of *B. cereus*, *B. stearothermophilus*, and "*T. thalpophilus*" genes. DNA sequence analyses of the regions of the cloned fragments which hybridized to the *B. megaterium* B-type SASP gene probe confirmed that they did encode B-type SASP genes (Fig. 3, 4, and 5). This result was indicated by the homology between the amino acid sequences of the proteins coded for and the sequences of the B-type SASP from *B. megaterium* and *B. subtilis* (see below), as well as the presence of two spore protease cleavage sites in these proteins (arrows in Fig. 3). In addition, the sequence of the first 14 amino acids coded for by the *B. cereus* B-type SASP gene is identical to that previously determined for a B-type SASP purified from *B. cereus* spores (Fig. 3, underlined residues), with the exception of the

TABLE 2. Comparison of various properties of different B-type SASP[a]

Source of B-type SASP	No. of residues	Net charge	No. of large hydrophobic residues[b]	No. of amides
B. cereus	92	+4	10	18
B. megaterium	96	+6	9	31
B. stearothermophilus	81	+1	8	32
B. subtilis	83	+1	8	32
"*T. thalpophilus*"	95	+3	11	35

[a] Data are taken from primary sequences given in Fig. 3, 4, and 5 and references 6 and 7. All values are for the proteins without the amino-terminal methionine, since this appears to be removed posttranslationally.
[b] Includes isoleucine, leucine, methionine, phenylalanine, proline, tyrosine, and valine.

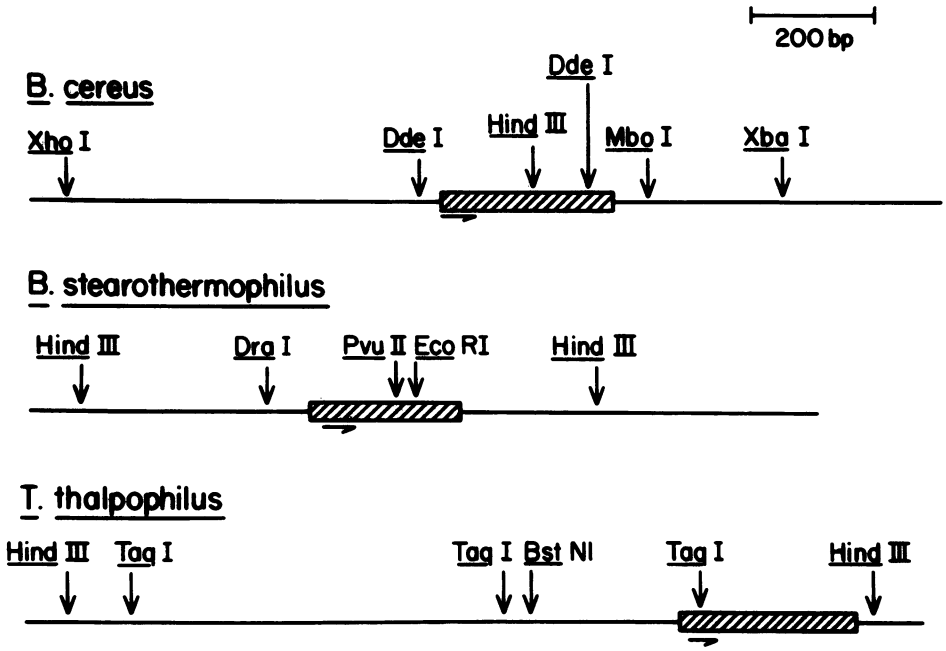

FIG. 2. Restriction maps of cloned DNA. Restriction maps were determined, and the regions hybridizing to B-type SASP gene probes were localized as described in Materials and Methods and from DNA sequence analyses. The boxed regions represent the B-type SASP gene coding sequences, and the arrows below them show the direction of translation. bp, Base pairs.

amino-terminal methionine, which is presumably removed posttranslationally. Previous work had suggested that the amino acid sequence downstream from both spore protease cleavage sites in the *B. cereus* B-type SASP (originally called SASP-II) was Phe-Ala-Thr-Glu-Thr-Asp (13). However, reexamination of the protein sequencing experiment which produced the data on which this sequence assignment was made showed that the yield of alanine in cycle 2 was 50% that of phenylalanine in cycle 1 and only 65% that of glutamic acid in cycle 4 (data not shown). Consequently, it appears that only one of the two spore protease cleavage sites has alanine in position 2 and that the other has serine, which was not detected by the high-pressure liquid chromatography system in use at that time.

All of the genes had a strong gram-positive-type ribosome-binding site 6 base pairs upstream from the site of initiation of translation. The two genes for which sufficient DNA sequence data were obtained also had regions of dyad symmetry shortly after the translational stop codon which may be a transcription stop signal (2, 4, 8) (Fig. 3 and 4).

When coding sequence probes from the *B. cereus*, *B. stearothermophilus*, and "*T. thalpophilus*" B-type SASP genes were hybridized to restriction enzyme digests of homologous genomic DNA, only a single major hybridizing

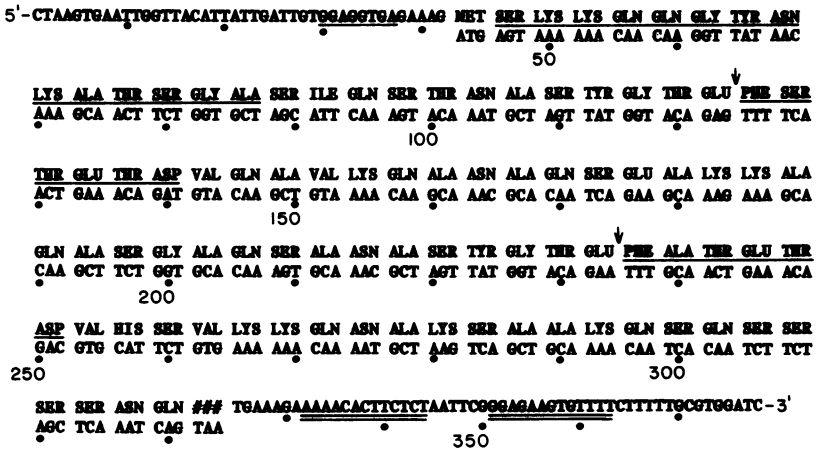

FIG. 3. Nucleotide sequence of the *B. cereus* B-type SASP gene coding and flanking sequences and the predicted amino acid sequence. The singly underlined bases from positions 30 to 37 show good complementarity to the 3' end of the 16S rRNA of *B. cereus*. The two regions that are doubly underlined can base pair with each other and may be a transcription stop signal. Dots below the nucleotides are positioned every 10 base pairs. Underlined amino acid residues are identical to those determined directly for the B-type SASP of *B. cereus* (13). ###, Translation stop signal.

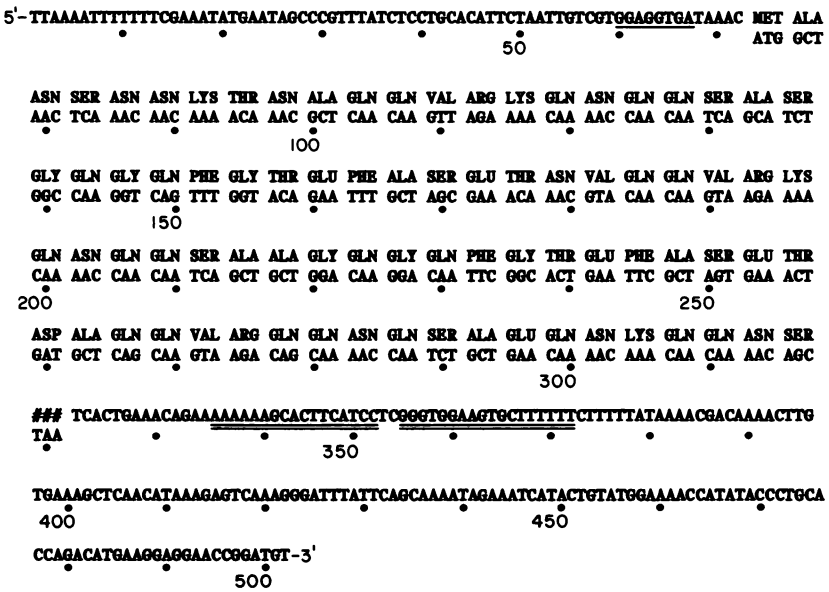

FIG. 4. Nucleotide sequence of the *B. stearothermophilus* B-type SASP gene coding and flanking sequences and the predicted amino acid sequence. The singly underlined bases from positions 60 to 67 show good complementarity to the 3′ end of the 16S rRNA of *B. stearothermophilus*. The two regions that are doubly underlined can base pair with each other and may be a transcription stop signal. Dots below the nucleotides are positioned every 10 base pairs. ###, Translation stop signal.

band was observed (Fig. 1, lanes d, e, and f). This result suggests that these organisms contain only a single B-type SASP gene, as has been found previously in *B. megaterium* and *B. subtilis* (6, 7).

Comparative analysis of properties and primary sequences of B-type SASP. A comparison of the properties and amino acid compositions of the B-type SASP from five different organisms revealed a number of conserved characteristics (Table 2). All five of the B-type SASP had a net positive charge at a neutral pH, even though the magnitude of the positive charge varied (Table 2). This result is in contrast to that for A- and C-type SASP, which have net charges which

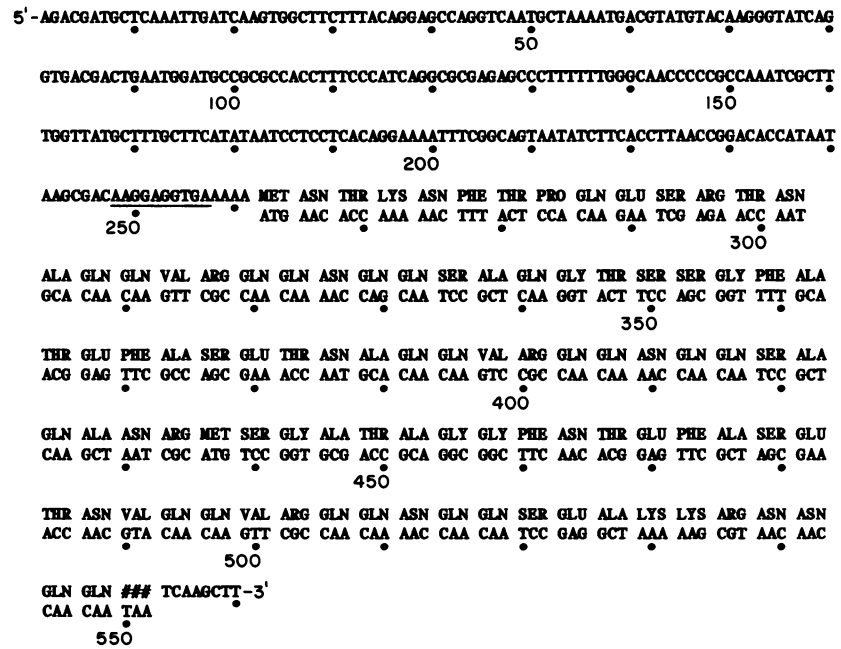

FIG. 5. Nucleotide sequence of the "*T. thalpophilus*" B-type SASP gene coding and flanking sequences and the predicted amino acid sequence. The underlined bases from positions 248 to 257 show good complementarity to the 3′ end of the 16S rRNA of *Bacillus* species and, by analogy, with *Thermoactinomyces* species as well. Dots below the nucleotides are positioned every 10 base pairs. ###, Translation stop signal.

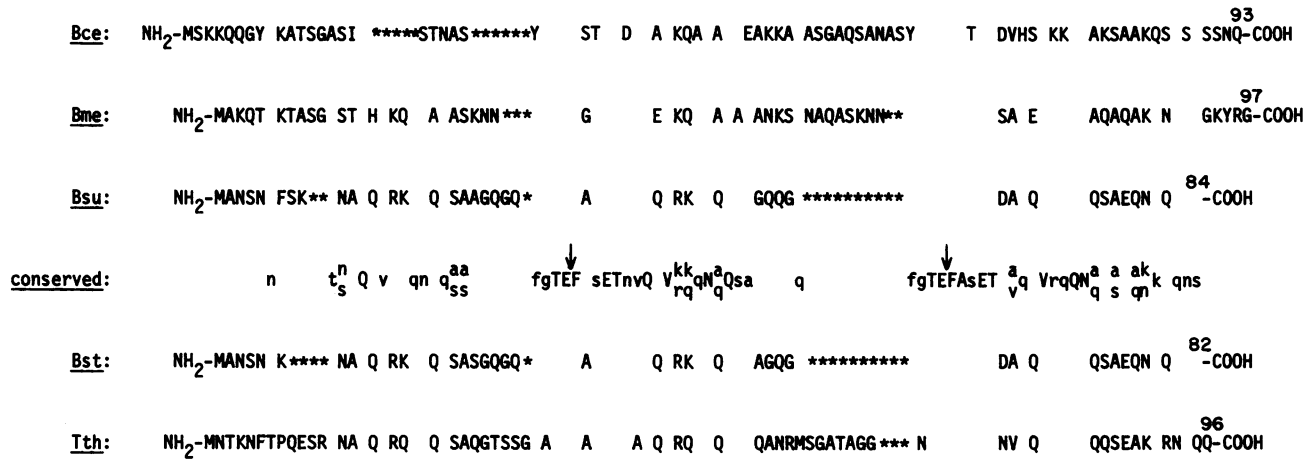

FIG. 6. Comparison of primary sequences of B-type SASP. Data are given in the one-letter amino acid code. Amino acid sequences were aligned to give maximum homology across species by introduction of gaps (spaces marked with asterisks) as necessary. The arrows denote the sites of spore protease cleavage. Residues conserved exactly in all five B-type SASP are shown in the middle of the figure in uppercase letters; residues identical in four of five B-type SASP are shown in lowercase letters. If only two residues were found at a position in all five B-type SASP, but no single residue in four B-type SASP, the two conserved residues are also shown in lowercase letters above and below the line. Bce, *B. cereus*; Bme, *B. megaterium*; Bsu, *B. subtilis*; Bst, *B. stearothermophilus*; Tth, "*T. thalpophilus*." Data are taken from Fig. 3, 4, and 5 and from references 6 and 7.

can be either positive or negative (2, 4, 8). Similarly, all five B-type SASP had very few large hydrophobic residues (only ~10% of total residues) (Table 2), in contrast to A- and C-type SASP, which are much more hydrophobic (~30% of total residues) (2, 8). B-type SASP also had a very high percentage of asparagine and glutamine residues—up to 40% of total residues (Table 2), although the *B. cereus* B-type SASP had significantly fewer amides than did the other four.

Given the conservation of these features of amino acid composition among different B-type SASP, it is not surprising that these proteins also exhibit significant homologies in overall primary sequences. However, to align these sequences maximally, a number of gaps must be introduced (Fig. 6). This result is in contrast to the situation with the A- and C-type SASP sequences, all 15 of which can be aligned with high homology without the introduction of gaps (2, 8). Furthermore, even when all five B-type SASP sequences are maximally aligned, only 19 residues are conserved in all five

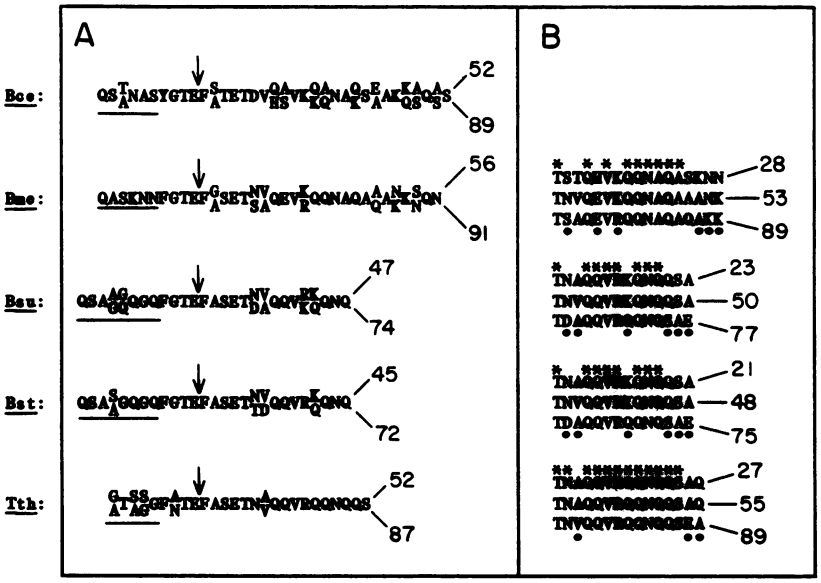

FIG. 7. (A) Primary sequence of the internal duplications in B-type SASP. (B) Primary sequence of regions in B-type SASP which are repeated three times. Data are taken from Fig. 3, 4, and 5 and references 6 and 7 and are given in the one-letter amino acid code; sequences were aligned without the use of gaps. The numbers given are the residue numbers of the carboxy-terminal residues of the repeats when the B-type SASP amino-terminal methionine is assigned as residue 1. In panel A, the single residues are those identical in both repeats. The residues below and above the line are the residues in the more amino-terminal and more carboxyl-terminal repeats, respectively. The arrows denote the spore protease cleavage sites. The underlined residues differ significantly among all five B-type SASP. In panel B, positions with identical residues in all three regions are denoted by an asterisk above the conserved position. Positions at which two of the three residues in all three regions are identical are denoted by a dot below the conserved position. Bce, Bme, Bsu, Bst, and Tth are as in Fig. 6.

proteins. In contrast, 27 residues in the smaller A- and C-type SASP are conserved across species (8). This observation suggests that the function served by B-type SASP is less primary sequence dependent than is that served by A- and C-type SASP.

Despite the fact that B-type SASP show less primary sequence homology among species than do A- and C-type SASP, B-type SASP do have several highly conserved features. All B-type SASP have two sites for cleavage by the spore-specific protease (arrows in Fig. 6), and the primary sequences around these cleavage sites are reasonably well conserved across species. Similarly, all five B-type SASP have large regions of sequences (27 to 35 residues) which appear twice in each protein, with 69 to 89% of the residues being identical in these two repeats (Fig. 7A). The reason and function for this sequence duplication are not clear, but it is possible that the present B-type SASP gene arose from an ancestral gene by gene duplication. This region of sequence duplication includes the spore protease cleavage site, and the sequences from this cleavage site towards the carboxy terminus of the protein are similar in all five B-type SASP, as are the first four residues towards the amino terminus of the protein (Fig. 7A). However, the next five to eight residues towards the amino terminus in the sequence repeats (underlined residues in Fig. 7A) may differ significantly between different B-type SASP, even though they are generally conserved in both sequence repeats in the same organism (e.g.,—compare Bme with Bst). This observation suggests that there is some functional significance to these regions of repeated sequence, such that a change in one of the sequence repeats selects for a similar change in the other. However, at present it is not clear how this selection is effected or what function is served by the sequence repeats.

In addition to the sequence repeats, four of the five B-type SASP have regions of 14 to 17 residues which appear three times; these sequences are also similar in all four B-type SASP (Fig. 7B). The exception is the *B. cereus* B-type SASP, since the gene coding for these SASP has apparently suffered a deletion in the region coding for the most amino-terminal triplicated domain (Fig. 6 and 7B). In the triplicated sequences, 8 to 12 residues are identical in all three sequences, and 3 to 7 residues are identical in two of the three sequences (Fig. 7B). In only one of the proteins (*B. megaterium* B-type SASP) is there more than one position in which three different residues are found (Fig. 7B). While this region of triplicated sequence is certainly a striking feature of these proteins, again we have no clear understanding of its potential function or significance. Possibly, when the function in vivo of B-type SASP themselves becomes clear, we will be able to understand the significance of the sequence repeats within these proteins.

ACKNOWLEDGMENT

This work was supported by Public Health Service grant GM-19698 from the National Institutes of Health.

LITERATURE CITED

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
2. **Connors, M. J., S. Howard, J. A. Hoch, and P. Setlow.** 1986. Determination of the chromosomal locations of four *Bacillus subtilis* genes which code for a family of small, acid-soluble spore proteins. J. Bacteriol. **166**:412–416.
3. **Fliss, E. R., M. J. Connors, C. A. Loshon, E. Curiel-Quesada, B. Setlow, and P. Setlow.** 1985. Small, acid-soluble spore proteins of *Bacillus*: products of a sporulation-specific, multigene family, p. 60–66. *In* J. A. Hoch and P. Setlow (ed.), Molecular biology of microbial differentiation. American Society for Microbiology, Washington, D.C.
4. **Fliss, E. R., C. A. Loshon, and P. Setlow.** 1986. Genes for *Bacillus megaterium* small, acid-soluble spore proteins: cloning and nucleotide sequence of three additional genes from this multigene family. J. Bacteriol. **165**:467–473.
5. **Foerster, H. F.** 1978. Effects of temperature on the spores of thermophilic actinomycetes. Arch. Microbiol. **118**:257–264.
6. **Hackett, R. H., B. Setlow, and P. Setlow.** 1986. Cloning and nucleotide sequence of the *Bacillus megaterium* gene coding for small, acid-soluble spore protein B. J. Bacteriol. **168**:1023–1025.
7. **Hackett, R. H., and P. Setlow.** 1987. Cloning, nucleotide sequencing, and genetic mapping of the gene for small, acid-soluble spore protein γ of *Bacillus subtilis*. J. Bacteriol. **169**:1985–1992.
8. **Loshon, C. A., E. R. Fliss, B. Setlow, H. F. Foerster, and P. Setlow.** 1986. Cloning and nucleotide sequencing of genes for small, acid-soluble spore proteins of *Bacillus cereus*, *Bacillus stearothermophilus*, and "*Thermoactinomyces thalpophilus*." J. Bacteriol. **167**:168–173.
9. **Mason, J. M., and P. Setlow.** 1986. Essential role of small, acid-soluble spore proteins in resistance of *Bacillus subtilis* spores to UV light. J. Bacteriol. **167**:174–178.
10. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
11. **Setlow, P.** 1981. Biochemistry of bacterial forespore development and spore germination, p. 13–28. *In* H. S. Levinson, A. L. Sonenshein, and D. J. Tipper (ed.), Sporulation and germination. American Society for Microbiology, Washington, D.C.
12. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
13. **Yuan, K., W. C. Johnson, D. J. Tipper, and P. Setlow.** 1981. Comparison of various properties of low-molecular-weight proteins from dormant spores of several *Bacillus* species. J. Bacteriol. **146**:965–971.