NOTES

Cloning and Nucleotide Sequence of a Plasmid-Carried Gene Coding for a Minor Small, Acid-Soluble Protein from *Bacillus megaterium* Spores

YARERY CARRILLO-MARTINEZ AND PETER SETLOW*

Department of Biochemistry, University of Connecticut Health Center,
Farmington, Connecticut 06030-3305

Received 24 May 1993/Accepted 27 July 1993

The gene (termed *sspG*) coding for a small, acid-soluble protein (SASP) from spores of *Bacillus megaterium* QMB1551, termed SASP-G, has been cloned, and its nucleotide sequence has been determined. SASP-G is a 42-residue protein containing 2 tryptophan and 11 lysine residues, including a hexalysine sequence, and is not homologous to any previously described SASP. The *sspG* gene appears to be an additional member of the σ^G regulon. No gene homologous to *sspG* is present in *B. cereus* T or *B. subtilis* 168. The reason for the absence of *sspG* from other *Bacillus* species appears to be that in *B. megaterium*, *sspG* is present only on a 111-kb plasmid; this plasmid is not present in *B. cereus* T or *B. subtilis* 168. The *sspG* gene is the first forespore-expressed gene found to be on a plasmid.

Approximately 10 to 20% of the protein in spores of *Bacillus* species is a group of small, acid-soluble proteins (SASP) (10). These proteins are synthesized late in sporulation, almost certainly only in the developing forespore; are located in the spore core; and are degraded in the first minutes of spore germination, with the resultant amino acids supporting protein synthesis at that time (9, 10). About 50% of the SASP pool in dormant spores is composed of an extremely homologous group of α/β -type SASP which are coded for by a multigene family. The α/β -type SASP are DNA-binding proteins and are essential for spore UV resistance and long-term survival (11). An additional 30 to 40% of the SASP pool is composed of a single γ -type SASP whose only known function is generation of amino acids early in spore germination (10). The remainder of the SASP pool is a group of minor SASP apparently unrelated to either the α/β - or γ -type SASP (9). While genes for both α/β - and γ -type SASP from a number of *Bacillus* species have been cloned, sequenced, and analyzed, no genes for the unrelated minor SASP have been isolated (10). However, a number of these minor SASP have been purified from spores of *B. megaterium* QMB1551 (9). One of these proteins, termed SASP-G, is unique among known SASP in that it contains one or more tryptophan residues; SASP-G also contains ~30% lysine residues. To further analyze this potentially novel small (~5 kDa) protein (9), we undertook to clone and analyze the gene coding for SASP-G, which we have termed *sspG*.

The strategy we used to clone *sspG* from *B. megaterium* QMB1551 began with determination of a significant amount of the protein's amino acid sequence. Using SASP-G purified as previously described (9), we were able to determine the first 40 residues of the protein with a MilliGen automated protein sequencer; the sequence determined was GYKEY WDDEKHSEESSKKKKKSEHKDSGATVRIISLWK. On the basis of this amino acid sequence, we synthesized two

degenerate oligonucleotides: oligonucleotide A [5'-GGGA (T/C)GA(T/C)GA(A/G)AA(A/G)CA(T/C)TC-3'] and oligonucleotide B [3'-CC(A/G/T)CG(A/T/C)TG(A/T/C)CA(A/C/G/T) (G/T)C(A/G/T)TA. Oligonucleotide A is homologous to all or part of the region coding for amino acids 6 to 12; oligonucleotide B is homologous to all or part of the region coding for amino acids 30 to 35. Use of the two oligonucleotides in a polymerase chain reaction (PCR) with *B. megaterium* DNA as the template yielded a number of PCR products (Fig. 1, lane A), a prominent one of which was ~90 bp long, close to the size predicted from SASP-G's amino acid sequence (88 bp). Use of this ~90-bp band in a second PCR yielded primarily the 90-bp product (Fig. 1, lane B). This 90-bp PCR product was cloned in *HincII*-cut plasmid pUC19, and determination of the nucleotide sequence of the insert confirmed that it coded for amino acids 6 to 35 of SASP-G (data not shown). Use of a DNA probe prepared from the cloned PCR fragment in hybridizations with Southern blots of restriction enzyme-digested *B. megaterium* DNA yielded good hybridization to *EcoRI*, *HincII*, and *PstI* fragments of 4.7, 2.0, and 15 kb, respectively (Fig. 1, lane C, and data not shown).

The 4.7-kb *EcoRI* fragment from *B. megaterium* which contained *sspG* was cloned in *EcoRI*-cut bacteriophage λ gt10 DNA by methods described previously (5). *EcoRI* digestion of DNA from a phage hybridizing to a probe from the initial cloned PCR fragment yielded the expected 4.7-kb hybridizing *EcoRI* fragment; digestion of the latter fragment with *HincII* yielded a 0.5-kb *EcoRI*-*HincII* hybridizing fragment. This fragment was cloned in *EcoRI*-*HincII*-cut plasmid pUC19, and the nucleotide sequence of the insert was determined completely in both directions. Examination of the DNA sequence of this fragment (Fig. 2) showed that it contains an open reading frame coding for a 43-residue protein whose sequence is identical to that determined directly for SASP-G with the exception of an amino-terminal methionine (presumably removed posttranslationally) and

* Corresponding author.

FIG. 1. Identification of a PCR product containing a part of the *sspG* coding sequence as well as the complete *sspG* gene. In lane A, a PCR was run with oligonucleotides A and B and *B. megaterium* DNA as described previously (14) with a 50°C annealing temperature in the first cycle, a 52°C annealing temperature in the second cycle, and a 55°C annealing temperature in the subsequent 26 cycles. An aliquot of the final reaction mixture was run on a 3% low-melting-temperature agarose gel and stained with ethidium bromide. The arrowheads numbered 1, 2, and 3 denote the positions of DNA markers of 310, 194, and 118 bp, respectively. The arrowhead labeled G denotes a band seen only with a PCR run with both oligonucleotides A and B. All other bands were seen in reactions with oligonucleotide B alone. In lane B, the band labeled G was cut from the gel shown in lane A and used as a template in a second PCR. An aliquot was run on a gel as described above and stained with ethidium bromide. The arrowheads labeled 1, 2, and 3 give the migration positions of DNA size markers of 310, 194, and 118 bp, respectively. In lane C, the ~90-bp fragment from the second PCR shown in lane B was cloned in the *HincII* site of pUC19. This fragment was cut from the plasmid by digestion with *PstI* and *XbaI*, isolated by electrophoresis on a 3% low-melting-temperature agarose gel, and labeled with [α - 32 P]dATP by random priming. This probe was then used in a hybridization against *EcoRI*-digested chromosomal DNA from *B. megaterium* QMB1551. Hybridization was carried out as described previously (5) in a mixture of 0.75 M NaCl and 75 mM sodium citrate (pH 7.0) containing 0.1% sodium dodecyl sulfate, 0.02% Ficoll, 0.02% bovine serum albumin, and 10 μ g of denatured salmon sperm DNA per ml. Hybridization was done at 52°C for 16 to 24 h; filters were washed at 52°C, once with 0.30 M NaCl–30 mM sodium citrate (pH 7.0)–0.1% sodium dodecyl sulfate and three times with 0.75 M NaCl–75 mM sodium citrate (pH 7.0)–0.1% sodium dodecyl sulfate, and then dried and autoradiographed. The arrowheads labeled 1, 2, 3, and 4 give the migration positions of DNA size markers of 6.7, 4.5, 2.0, and 0.6 kb, respectively.

two additional carboxy-terminal residues which were below detection limits in the sequencer run with the intact protein. The *sspG* coding sequence is preceded by a sequence showing good homology to the 3' end of *B. megaterium* 16S rRNA (Fig. 2); this is presumably a ribosome binding site. Downstream of the coding sequence there is a sequence with the potential to form a stem-loop structure (Fig. 2) followed by a T-rich region; this is presumably a transcription termination site (16). Previous work has shown that SASP-G appears during sporulation in parallel with both α/β - and γ -type SASP (9). Since genes coding for α/β -type and γ -type SASP are transcribed exclusively during sporulation by

RNA polymerase that contains σ^G (8, 10), it seemed likely that *sspG* would also be a member of the σ^G regulon. Indeed, upstream of the *sspG* coding sequence there are appropriately spaced regions with homology to –35 and –10 sequences recognized by RNA polymerase containing σ^G (Fig. 2) (4, 8). Unfortunately, primer extension analyses of the site of initiation of *sspG* transcription with two different primers from within the *sspG* coding sequence and multiple RNA preparations isolated at a number of times during growth and sporulation have failed to yield a detectable extension product (data not shown). Consequently, the assignment of the –35 and –10 sequences of *sspG* is only tentative at present.

Examination of the amino acid sequence of SASP-G reveals a highly unusual protein. Two tryptophan residues are present, as suggested previously (9), indicating that SASP-G is one source of the tryptophan for protein synthesis early in spore germination which must be provided by protein degradation, because of the absence of tryptophan synthetase at that time (10). The mature protein also contains 11 lysines out of 42 residues, similar to the percentage (~30%) found previously (9). Strikingly, six of the lysines are in a row; this hexalysine sequence is extremely unusual in proteins. Indeed, SASP-G exhibits no significant similarity to any protein in available data bases. Not surprisingly, given its high level of basic amino acids, SASP-G binds to a variety of nucleic acids in vitro and provides complete DNase protection at a protein/DNA ratio of ~2/1 (wt/wt) (2).

Previous work has shown that SASP-G is rapidly degraded during spore germination, but that the purified protein is not attacked by the sequence-specific protease which initiates degradation of both α/β -type and γ -type SASP during spore germination (9). Analysis of the SASP-G amino acid sequence shows that this protein lacks the pentapeptide sequence –E-I/F-A-S-E–, which is essential for recognition and cleavage by this sequence-specific protease (10); this finding explains its lack of degradation by this enzyme. At present, the identity of the protease that initiates SASP-G degradation during spore germination is not clear.

One of the major reasons for cloning the *sspG* gene from *B. megaterium* was to use this gene as a probe to clone the homologous gene from *B. subtilis*, as this would allow facile genetic manipulation of the gene. Surprisingly, when the 4.7-kb *EcoRI* fragment containing *B. megaterium sspG* was used as a probe against Southern blots of *B. subtilis* 168 (Fig. 3) or *B. cereus* T (data not shown) DNA under nonrestrictive conditions, no obvious hybridizing bands were obtained when the DNA was cut with *EcoRI* (Fig. 3, lane B) or with a number of other restriction enzymes (*BamHI*, *BclI*, *BglII*, *HindIII*, *SalI*, *Sau3AI*, and *TaqI* [data not shown]). In contrast, use of more stringent hybridization and washing conditions (temperature of 55°C and washing with 15 mM NaCl–1.5 mM Na citrate (pH 7.0)–0.1% sodium dodecyl sulfate) still results in strong cross-hybridization between *ssp* genes coding for α/β -type SASP from *B. megaterium* and *B. subtilis* (3). Consequently, the lack of hybridization between the *B. megaterium sspG* probe and *B. cereus* and *B. subtilis* DNA strongly suggested that *sspG* had either diverged rapidly during evolution or was not present in these other two *Bacillus* species. One obvious reason for the latter possibility would be if *sspG* was located on a plasmid. *B. megaterium* QMB1551 carries at least seven different plasmids comprising ~400 kb of potential coding capacity (6); these plasmids are almost certainly absent from *B. cereus* T and *B. subtilis* 168. Some of the *B. megaterium* plasmids are extremely large (>100 kb) (6) and would probably be present in chromosomal DNA preparations. Indeed, isolation and

sspG is the first plasmid-carried gene identified which is expressed only in the forespore during sporulation and which seems likely to code for a protein present in the spore core. Unfortunately, given the absence of *sspG* from other *Bacillus* species, it seems unlikely that SASP-G plays an important role in any highly conserved spore property.

Nucleotide sequence accession number. The *B. megaterium* *sspG* sequence determined in this work has been assigned GenBank accession number L20464.

We are grateful to George Korza for the protein sequence analysis and to Pat Vary for *B. megaterium* strains and communication of results.

This work was supported by grant GM-19698 from the National Institutes of Health. Partial support for Yarey Carrillo-Martinez was provided by the Consejo Nacional de Ciencia y Tecnología.

REFERENCES

1. Aronson, A. I. 1993. Insecticidal toxins, p. 953–963. In A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
2. Carrillo-Martinez, Y., and P. Setlow. Unpublished results.
3. Connors, M. J., and P. Setlow. 1985. Cloning of a small, acid-soluble spore protein gene from *Bacillus subtilis* and determination of its complete nucleotide sequence. *J. Bacteriol.* **161**:333–339.
4. Fajardo-Cavazos, P., F. Tovar-Rojas, and P. Setlow. 1991. Effect of promoter mutations and upstream deletions on the expression of genes coding for small, acid-soluble spore proteins of *Bacillus subtilis*. *J. Bacteriol.* **173**:2011–2016.
5. 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.
6. Kieselburg, M. K., M. Weikert, and P. S. Vary. 1984. Analysis of resident and transformant plasmids in *Bacillus megaterium*. *Bio/Technology* **2**:254–259.
7. McCormick, J. R., and R. Santamaria. 1992. Three spore-associated proteins are encoded on linear plasmid SCP1 in *Streptomyces coelicolor*, abstr. 57. Program Abstr. Eleventh Int. Spores Conf.
8. Nicholson, W. L., D. Sun, B. Setlow, and P. Setlow. 1989. Promoter specificity of σ^G -containing RNA polymerase from sporulating cells of *Bacillus subtilis*: identification of a group of forespore-specific promoters. *J. Bacteriol.* **171**:2708–2718.
9. Setlow, P. 1978. Purification and characterization of additional low-molecular-weight basic proteins degraded during germination of *Bacillus megaterium* spores. *J. Bacteriol.* **136**:331–340.
10. Setlow, P. 1988. Small acid-soluble, spore proteins of *Bacillus* species: structure, synthesis, genetics, function and degradation. *Annu. Rev. Microbiol.* **42**:319–338.
11. Setlow, P. 1992. I will survive: protecting and repairing spore DNA. *J. Bacteriol.* **174**:2737–2741.
12. Stevenson, D. M., D. Lach, and P. S. Vary. 1993. A gene required for germination in *Bacillus megaterium* is plasmid-borne. In E. Balla and G. Berencsie (ed.), *Genetic transformation: DNA transformation and gene expression in microorganisms*, in press. Intercept Limited, Andover, Hampshire, United Kingdom.
13. Stevenson, D. M., and P. S. Vary. 1993. Personal communication.
14. Sun, D., and P. Setlow. 1991. Cloning, nucleotide sequence, and expression of the *Bacillus subtilis* *ans* operon which codes for L-asparaginase and L-aspartase. *J. Bacteriol.* **173**:3831–3845.
15. Vary, P. S., and Y.-P. Tao. 1987. Development of genetic methods in *Bacillus megaterium*, p. 403–408. In A. T. Ganesan and J. A. Hoch (ed.), *Genetics and biotechnology of bacilli*, vol. 2. Academic Press, San Diego, Calif.
16. Von Hippel, P. H., D. G. Bear, W. D. Morgan, and J. A. McSwiggen. 1984. Protein-nucleic acid interactions in transcription. *Annu. Rev. Biochem.* **53**:389–446.