

# Comparison of Various Properties of Low-Molecular-Weight Proteins from Dormant Spores of Several *Bacillus* Species

KATHERINE YUAN,<sup>1</sup> W. CHARLES JOHNSON,<sup>2</sup> DONALD J. TIPPER,<sup>2</sup>  
AND PETER SETLOW<sup>1\*</sup>

*Department of Biochemistry, University of Connecticut Health Center, Farmington, Connecticut 06032<sup>1</sup>, and  
Department of Microbiology, University of Massachusetts Medical School, Worcester, Massachusetts 06105<sup>2</sup>*

Received 15 January 1981/Accepted 16 March 1981

Several properties of the major proteins degraded during germination of spores of *Bacillus cereus*, *Bacillus megaterium*, and *Bacillus subtilis* have been compared. All of the proteins had low molecular weights (6,000 to 13,000) and lacked cysteine, cystine, and tryptophan. The proteins could be subdivided into two groups: group I (*B. megaterium* A and C proteins, *B. cereus* A protein, and *B. subtilis*  $\alpha$  and  $\beta$  proteins) and group II (*B. cereus* and *B. megaterium* B proteins and *B. subtilis*  $\gamma$  protein). Species in group II had lower levels of (or lacked) the amino acids isoleucine, leucine, methionine, and proline. Similarly, proteins in each group were more closely related immunologically. However, antisera against a *B. megaterium* group I protein cross-reacted more strongly with the *B. megaterium* group II protein than with group I proteins from other spore species, whereas antisera against the *B. megaterium* group II protein cross-reacted most strongly with *B. megaterium* group I proteins. Analysis of the primary sequences at the amino termini and in the regions of the *B. cereus* and *B. subtilis* proteins cleaved by the *B. megaterium* spore protease revealed that the *B. cereus* A protein was most similar to the *B. megaterium* A and C proteins, and the *B. cereus* B protein and the *B. subtilis*  $\gamma$  protein were most similar to the *B. megaterium* B protein. However, amino terminal sequences within one group of proteins varied considerably, whereas the spore protease cleavage sites were more highly conserved.

Approximately 20% of the protein in dormant spores of several *Bacillus* and at least one *Clostridium* species is degraded during the first minutes of spore germination (2, 8, 9, 16). This massive proteolysis generates free amino acids which are reutilized by the germinating and outgrowing spore for energy production and new protein synthesis (15). The proteins degraded in this process are a group of low-molecular-weight, acid-soluble proteins which are unique to the developing and dormant spore (8). In *Bacillus megaterium*, three distinct species constitute the majority of the protein degraded; these proteins (termed A, B, and C in *B. megaterium*) have been purified and characterized from *B. megaterium* spores (8, 9). The complete primary sequence of all three proteins has been reported, as well as the sites where they are cleaved by a highly specific protease purified from *B. megaterium* spores (11, 13, 14, 14a). *Bacillus subtilis* spores also contain three major acid-soluble proteins (termed  $\alpha$ ,  $\beta$ , and  $\gamma$ ) which are degraded during spore germination (2). The characterization of these proteins has recently been reported (2). *Bacillus cereus* spores also contain several major acid-soluble proteins presumably analo-

gous to the *B. subtilis* and *B. megaterium* proteins (8). However, the purification of the *B. cereus* proteins has not been reported.

One function of this unique group of proteins which has been established in *B. megaterium* is to serve as a repository for amino acids in a polymeric form (8). These proteins are degraded rapidly during spore germination, thus regenerating free amino acids. However, it has also been suggested that these proteins may play some role in the resistance of spores to UV light (6, 9). Consequently, it might be fruitful to compare proteins from different spore species to determine whether they have a highly conserved sequence. In this work, we report a comparison of various physical, chemical, immunological, and primary structure parameters of some of the major unique low-molecular-weight proteins from dormant spores of *B. cereus*, *B. megaterium*, and *B. subtilis*.

## MATERIALS AND METHODS

**Proteins, protease, and reagents.** The A, B, and C proteins and the spore protease were purified from dormant spores of *B. megaterium* QM B1551 (originally obtained from H. S. Levinson, U.S. Army Labo-

laboratories, Natick, Mass.) as previously described (9, 10). The mixture of proteins  $\alpha$  and  $\beta$  and the homogenous protein  $\gamma$  were purified from spores of *B. subtilis* 168 as described by Johnson and Tipper (2). The sources of reagents for primary sequence analysis were described previously (13).

***B. cereus* proteins A and B.** Proteins A and B were purified from dormant spores of *B. cereus* T (originally obtained from H. O. Halvorson, Brandeis University, Waltham, Mass.) by a modification of the procedure for purification of the *B. megaterium* proteins. Spores of *B. cereus* were prepared by growth in supplemented nutrient broth, harvested, and lyophilized as previously described (12). A 3-g amount of dry spores was suspended in 30 ml of cold water, and 30 ml of 9.5 N HNO<sub>3</sub> was added with rapid mixing. After 40 min at 4°C, the mixture was diluted threefold with cold water and centrifuged (15 min at 10,000 × *g*). The supernatant fluid was dialyzed exhaustively in Spectrapor dialysis tubing (molecular weight cutoff, 3,500) against 5 mM Tris-maleate (pH 6.0) (buffer A). The dialyzed material was centrifuged (10 min at 15,000 × *g*), and the supernatant fluid was applied to a column of carboxymethyl-cellulose (2.4 by 18 cm) equilibrated at 4°C in buffer A. The column was washed with 50 ml of buffer A, and proteins A and B were eluted with a linear gradient of 0 to 0.4 M NaCl in buffer A (250 ml of each). Two major protein peaks were eluted from the column (Fig. 1), and these were pooled, dialyzed against 1% acetic acid, lyophilized, dissolved in a small volume of water, and stored frozen. The yield of protein from 3 g of spores was 34 mg of protein A and 26 mg of protein B. Both proteins were homogeneous as shown by electrophoresis on either sodium dodecyl sulfate-polyacrylamide gels (17) or polyacrylamide gels run at low pH (5).

**Amino acid analysis and primary sequence determinations.** Amino acid compositions were determined on acid hydrolysates, using a Beckman 121

automatic amino acid analyzer. Determinations of cysteine/cystine and tryptophan were carried out as described previously (9).

Primary sequence determinations were carried out on a Beckman 890 C sequencer, using the 0.1 M Quadrol program 102474. The sequencer was also coupled with a Sequemat P-6 Auto-converter. The phenylthiohydantoin derivatives were identified and quantitated by high-performance liquid chromatography as described previously (13). In some cases, the identification or quantitation or both was also carried out by amino acid analysis after back-hydrolysis of the phenylthiohydantoin derivatives with hydroiodic acid (13).

Six cycles of primary sequence analysis of *B. megaterium* spore protease digests of various proteins (see below) was also carried out as described above. The new amino terminal sequences generated by spore protease cleavage were determined from these analyses by subtraction of residues which arose from the original amino terminus of the protein as described previously (11). The number of protease cleavage sites was determined by the yield of new amino acids in each cycle relative to the yield of amino acids derived from the original amino terminus of the protein, by identification of more than one new amino acid in any given cycle, or by both methods.

**Other methods.** The radioimmunoassays for the A or B proteins from *B. megaterium* were carried out as previously described (5). The molecular weights of various proteins were determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (17) as previously described (9). *B. megaterium* spore protease cleavage of various proteins (50 to 150 nmol) was carried out in 0.2 to 0.3 ml of 50 mM Tris-hydrochloride (pH 7.4) and 5 mM CaCl<sub>2</sub>. Spore protease (10 to 30 U) was added to start the digestion, and after 4 h at 37°C, the mix was lyophilized and desalted on a column (1.3 by 50 cm) of Sephadex G-10 in 50% acetic acid.

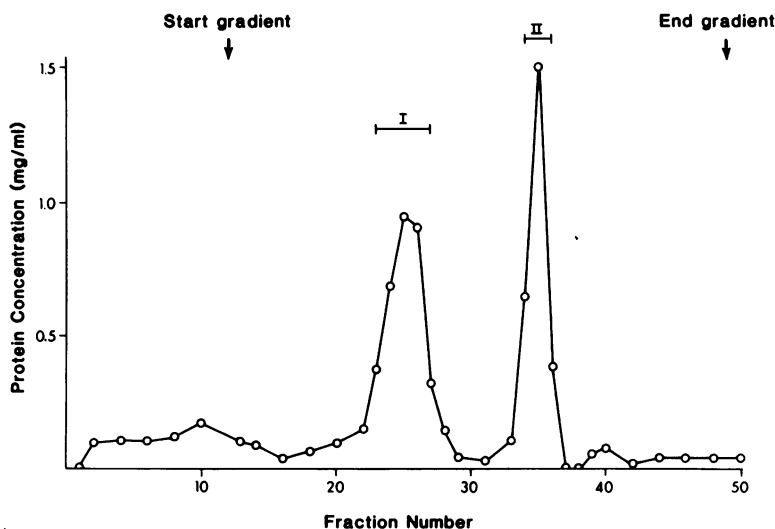


FIG. 1. Carboxymethyl-cellulose chromatography of the *B. cereus* A and B proteins. The column was run and analyzed as described in the text.

Protein was routinely determined by the method of Lowry et al. (3).

## RESULTS

**Molecular weight and amino acid composition.** All of the purified proteins, as well as the mixture of *B. subtilis*  $\alpha$  and  $\beta$  proteins, had low molecular weights, as determined by their migration on sodium dodecyl sulfate-polyacrylamide gels (Table 1). The values for the *B. megaterium* proteins were only slightly higher than the values calculated from primary sequence data (Table 1).

All of the proteins analyzed lacked the amino acids cysteine, cystine, and tryptophan, and histidine was present in only two proteins (Table 2). *B. cereus* protein A, *B. megaterium* proteins A and C, and the mixture of *B. subtilis* proteins  $\alpha$  plus  $\beta$  contained a significant amount of isoleucine, leucine, and methionine (12 to 18%), and all but the *B. subtilis* proteins contained proline. In contrast, *B. cereus* protein B, *B. megaterium* protein B, and *B. subtilis* protein  $\gamma$  lacked or had low levels (0 to 3%) of these four amino acids.

**Immunological cross-reaction.** The groupings of the proteins by their amino acid composition also held true for their immunological relationship. Within a given species, the protein(s) which was more like *B. megaterium* protein A in amino acid composition cross-reacted best with the antiserum against *B. megaterium* protein A (Table 3). Similarly, proteins which were more like *B. megaterium* protein B in amino acid composition cross-reacted best with the antiserum against *B. megaterium* protein B. However, the proteins from *B. cereus* or *B. subtilis* spores cross-reacted poorly with the antisera against the *B. megaterium* proteins (Table 3). Similarly, the antisera against the *B. subtilis*

TABLE 1. Molecular weights of low-molecular-weight spore proteins

| Protein                                 | Molecular wt <sup>a</sup>    |
|---|------------------------------|
| <i>B. megaterium</i> A                  | 7,000 <sup>b</sup> (6,250)   |
| <i>B. megaterium</i> B                  | 11,000 <sup>b</sup> (10,300) |
| <i>B. megaterium</i> C                  | 10,000 <sup>b</sup> (7,250)  |
| <i>B. cereus</i> A                      | 9,000                        |
| <i>B. cereus</i> B                      | 12,000                       |
| <i>B. subtilis</i> $\alpha$ and $\beta$ | 5,900 <sup>c</sup>           |
| <i>B. subtilis</i> $\gamma$             | 11,000 <sup>c</sup>          |

<sup>a</sup> Molecular weights were determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis as described in the text. Values in parentheses were calculated from the primary sequence (13 to 15).

<sup>b</sup> Values taken from reference 9.

<sup>c</sup> Values taken from reference 2.

TABLE 2. Amino acid composition of low-molecular-weight spore proteins<sup>a</sup>

| Amino acid     | Composition of protein |                |                  |                |                    |                               |                       |
|----------------|------------------------|----------------|------------------|----------------|--------------------|-------------------------------|-----------------------|
|                | <i>B. megaterium</i>   |                | <i>B. cereus</i> |                | <i>B. subtilis</i> |                               |                       |
|                | A <sup>b</sup>         | B <sup>c</sup> | C <sup>d</sup>   | A <sup>e</sup> | B <sup>e</sup>     | $\alpha + \beta$ <sup>f</sup> | $\gamma$ <sup>f</sup> |
| Lys            | 6.6                    | 10.4           | 5.6              | 9.2            | 11.5               | 4                             | 7                     |
| His            | 0                      | 1.0            | 0                | <0.3           | 1.1                | 0                             | 0                     |
| Arg            | 3.3                    | 1.0            | 4.3              | 3.5            | <0.3               | 3                             | 4                     |
| Asp            | 8.2                    | 13.5           | 14.1             | 6.9            | 9.1                | 14                            | 14                    |
| Thr            | 4.9                    | 8.3            | 2.8              | 4.3            | 7.4                | 6                             | 6                     |
| Ser            | 4.9                    | 9.4            | 7.0              | 6.7            | 14.7               | 8                             | 8                     |
| Glu            | 14.8                   | 25.0           | 12.7             | 18.7           | 19.8               | 15                            | 28                    |
| Gly            | 13.1                   | 6.3            | 11.3             | 13.2           | 5.6                | 12                            | 9                     |
| Ala            | 16.4                   | 15.6           | 15.5             | 15.4           | 17.8               | 15                            | 11                    |
| Val            | 6.6                    | 4.2            | 5.6              | 6.7            | 4.5                | 7                             | 5                     |
| Phe            | 1.7                    | 4.2            | 2.8              | 1.4            | 1.8                | 3                             | 5                     |
| Tyr            | 1.7                    | 1.0            | 2.8              | 1.4            | 2.1                | 0                             | 0                     |
| Pro            | 3.3                    | 0              | 2.9              | 1.3            | <0.3               | 0                             | 0                     |
| Met            | 3.3                    | 0              | 1.4              | 1.3            | <0.3               | 1                             | 0                     |
| Ile            | 4.9                    | 0              | 4.2              | 2.1            | 1.0                | 7                             | 1                     |
| Leu            | 6.6                    | 0              | 7.0              | 7.4            | <0.3               | 6                             | 2                     |
| Cys/Cys (half) | 0                      | 0              | 0                | <0.2           | <0.2               | 0                             | 0                     |
| Trp            | 0                      | 0              | 0                | <0.2           | <0.2               | 0                             | 0                     |

<sup>a</sup> Data are given as the number of residues per 100 residues.

<sup>b</sup> Data calculated from the primary sequence (13).

<sup>c</sup> Data calculated from the primary sequence (14a).

<sup>d</sup> Data calculated from the primary sequence (14).

<sup>e</sup> Values determined in this study.

<sup>f</sup> Data taken from reference 2.

TABLE 3. Cross-reaction of low-molecular-weight spore proteins with antisera to *B. megaterium* A or B proteins

| Protein                                 | Cross-reaction <sup>a</sup> with antiserum to: |                        |
|---|--|------------------------|
|   | <i>B. megaterium</i> A                         | <i>B. megaterium</i> B |
| <i>B. megaterium</i> A                  | 100  | 6                      |
| <i>B. megaterium</i> B                  | 9  | 100                    |
| <i>B. megaterium</i> C                  | 39   | 4                      |
| <i>B. cereus</i> A                      | 2  | 0.06                   |
| <i>B. cereus</i> B                      | 0.2  | 0.7                    |
| <i>B. subtilis</i> $\alpha$ and $\beta$ | 1  | 0.1                    |
| <i>B. subtilis</i> $\gamma$             | 0.1  | 1                      |

<sup>a</sup> All values are given as: (nanograms of homologous protein tested to give 50% inhibition in the immunoassay/nanograms of heterologous protein to give 50% inhibition)  $\times$  100%.

$\alpha$  plus  $\beta$  proteins or the *B. subtilis*  $\gamma$  protein cross-reacted quite strongly with each other but very poorly with the *B. megaterium* proteins (data not shown).

**Primary sequence determinations.** The low degree of immunological cross-reaction be-

tween the spore proteins from different species indicated significant differences in primary sequence. Indeed, analysis of the amino terminal primary sequences of the proteins from *B. cereus* and *B. subtilis* spores (Table 4) showed that all three sequences were very different. In addition, these sequences were different from those of the analogous *B. megaterium* proteins, although proteins grouped on the basis of overall amino acid composition and immunological cross-reaction did show significantly greater sequence homology (see below).

**Determination of primary sequences at *B. megaterium* spore protease cleavage sites.** Although the amino terminal sequences of the proteins from different species showed significant differences, it was possible that other regions of the proteins—in particular those with a specific function—might be more highly conserved. The spore protease of *B. megaterium* cleaves the *B. megaterium* proteins only at highly conserved sites whose primary sequences are essential for protease recognition (1, 11). These initial cleavages by the spore protease then allow further degradation of the fragments by spore peptidases (10). In all protease cleavage sites in the *B. megaterium* A, B, and C proteins, the sequences cleaved are X-Glu<sup>1</sup>-Phe/Ile-Gly/Ala-Ser-Glu-X, with cleavage at the first glutamyl bond (arrow). Automated sequenator analysis of unfractionated spore protease digests of the *B. cereus* and *B. subtilis* proteins, plus knowledge of the amino terminal sequence of the proteins, allowed determination of the number of cleavage sites and of their primary sequence distal to the *B. megaterium* spore protease cleavage sites (Fig. 2). The primary sequences in the cleavage sites were highly conserved, with only two conservative substitutions, both at position 4 of the *B. megaterium* pentapeptide cleavage site sequence (Fig. 2). The residues following the cleavage site pentapeptide were also conserved in the *B. cereus* and *B. subtilis* proteins (Fig. 2).

## DISCUSSION

The data presented in this communication support previous contentions that dormant spores of various *Bacillus* species contain large amounts of a few similar acid-soluble proteins. These proteins appear to be of two types: group I (*B. megaterium* A and C; *B. cereus* A, and *B. subtilis*  $\alpha$  and  $\beta$ ) and group II (*B. megaterium* B, *B. cereus* B, and *B. subtilis*  $\gamma$ ). These groupings are based on amino acid composition, immunological cross-reaction, and primary sequence comparison. All *Bacillus* species tested had at least one protein in each group, suggesting that possession of these different types of proteins is advantageous in some as yet unknown way. Despite the existence of these two distinct protein groups, it is clear that both groups are related within a given organism. Thus, there is cross-reaction between a group II protein and antisera to a group I protein (and vice versa), and there is significant sequence homology between the *B. megaterium* proteins in the two classes (14a). The existence of a common sequence in the spore protease cleavage site(s) in the proteins from different species also points out their relatedness.

Nevertheless, one of the striking findings in this work was that, despite the gross similarities between proteins from different species, detailed comparisons revealed large differences. Thus, proteins from the two groups from the same spore species were more closely related immunologically than were proteins from the same group but from different species. Other such striking differences were seen upon comparison of the amino terminal sequences of proteins from the same group but from different species (Fig. 3 and 4). Protein A from *B. cereus* had an amino terminal sequence which was most similar to that of the *B. megaterium* A and C proteins (Fig. 3). However, there were mismatches at both the amino terminus and the internal resi-

TABLE 4. Automated sequenator analysis of low-molecular-weight spore proteins

| Cycle | Analysis of protein <sup>a</sup> |                                  |   |
|-------|----------------------------------|----------------------------------|---|
|       | <i>B. cereus</i> A<br>(60 nmol)  | <i>B. cereus</i> B<br>(120 nmol) | <i>B. subtilis</i> $\gamma$<br>(150 nmol) |
| 1     | Ser (25) <sup>b</sup>            | Ser <sup>c</sup>                 | Ala (80)                                  |
| 2     | Arg (27) <sup>b</sup>            | Lys (33)                         | Asn (67)                                  |
| 3     | Ser (25) <sup>b</sup>            | Lys (38)                         | Ser <sup>c</sup>                          |
| 4     | Thr (6)                          | Gln (8)                          | Asn (52)                                  |
| 5     | Asn (27)                         | Gln (11)                         | Asn (48)                                  |
| 6     | Lys (23)                         | Gly (3)                          | Phe (31) <sup>b</sup>                     |
| 7     | Leu (18)                         | Tyr (13)                         | Ser <sup>c</sup>                          |
| 8     | Ala (13)                         | Asn (9)                          | Lys (36)                                  |
| 9     | Val (5) <sup>b</sup>             | Lys (12)                         | Thr (16)                                  |
| 10    | Pro (8)                          | Ala (11)                         | Asn (36)                                  |
| 11    | Gly (7)                          | Thr (3)                          | Ala (32)                                  |
| 12    | Ala (9)                          | Ser <sup>c</sup>                 | Gln (26)                                  |
| 13    | Glu (7)                          | Gly (6)                          | Gln (36)                                  |
| 14    | Ser <sup>c</sup>                 | Ala (12)                         | Val (24) <sup>b</sup>                     |
| 15    | Ala (6)                          |                                  | Arg (13)                                  |

<sup>a</sup> Phenylthiohydantoin derivatives were identified by high-performance liquid chromatography. Values in parentheses are the yield in nanomoles as determined by high-performance liquid chromatography.

<sup>b</sup> The quantitation (as well as the identity) of these residues was determined by amino acid analysis after back-hydrolysis of the phenylthiohydantoin derivatives with hydroiodic acid.

<sup>c</sup> Serine could not be quantitated accurately by high-performance liquid chromatography.

dues (Fig. 3). Examination of the sequences of the *B. cereus* B and *B. subtilis*  $\gamma$  proteins and comparison with the complete sequence of the *B. megaterium* B protein revealed that these species were related, but that there were large differences (Fig. 4). Thus, there was essentially no homology between the *B. cereus* and *B. subtilis* proteins (Table 4). Similarly, only residues 3 to 14 of the *B. cereus* B protein were present in the amino terminal region of the *B. megaterium* B protein (residues 2 to 11), and even within this region there were three changes which could be due to single base changes, as well as one change requiring more than a single

base change and two extra residues in the *B. cereus* protein (Fig. 4). Residues 1 to 5 of the *B. cereus* B protein actually best matched a region towards the carboxyl terminus of the *B. megaterium* B protein (residues 85 to 89) (Fig. 4). The *B. subtilis*  $\gamma$  protein sequence was even more different than that of the *B. megaterium* B protein, since the amino terminal sequences did not match up at all. Rather, residues 1 to 6 of the *B. subtilis*  $\gamma$  protein matched up best with residues 23 to 28 (or 58 to 63) of the *B. megaterium* B protein, and residues 7 to 15 of the *B. subtilis* protein matched up well with residues 69 to 77 of the *B. megaterium* B protein (Fig. 4).

Although comparison of the complete primary sequences for these proteins would be desirable, some significant conclusions can be drawn from the data available. In particular, it seems clear that there are few constraints on the primary sequence at the amino terminal regions of these proteins. Consequently, any function of this region of the proteins probably does not require a specific primary sequence, and if these proteins are involved in the spore's UV light resistance, they must do so by virtue of their general physical and chemical properties, not by means of sequence-specific interaction with spore DNA. In this respect, it is worth noting that the pI's of the *B. subtilis*  $\alpha$ ,  $\beta$ , and  $\gamma$  proteins (6.5 to 7.9) (2) are lower than those of the *B. megaterium* proteins (all >9), so that overall basicity is not a constant factor either.

A second conclusion is that the primary sequences of these related proteins from different species seem to have evolved in part by single amino acid changes, but more dramatically in some cases by rearrangement of significant blocks of primary sequence (i.e., carboxyl terminal and central regions of *B. megaterium* proteins found at the amino terminal regions of *B. cereus* and *B. subtilis* proteins). Although the significance and/or mechanism of such drastic

|  |   |
|--|---|
|  | Group I proteins  |
| <u><i>B. megaterium</i></u><br>A and C | R-Glu <sup>†</sup> Ile-Ala-Ser-Glu-Phe-Gly                  |
| <u><i>B. cereus</i></u> A (2 sites)    | †Ile-Ala-Gln-Glu-Phe-Gly                                    |
|  | Group II proteins   |
| <u><i>B. megaterium</i></u> B          | R-Glu <sup>†</sup> Phe-Gly-Ser-Glu-Thr-Asn                  |
|  | and   |
|  | R-Glu <sup>†</sup> Phe-Ala-Ser-Glu-Thr-Ser                  |
| <u><i>B. cereus</i></u> B (2 sites)    | †Phe-Ala-Thr-Glu-Thr-Asp                                    |
| <u><i>B. subtilis</i></u> $\gamma$     | †Phe-Ala-Ser-Glu-Thr-Asn<br>and<br>†Phe-Ala-Ser-Glu-Thr-Asp |

FIG. 2. Primary sequences around the *B. megaterium* spore protease cleavage sites in proteins from different species. The sequences for the *B. megaterium* proteins are from reference 11; the remaining sequences were determined as described in the text. In all cases, the yield of new amino termini after protease cleavage of the *B. cereus* and *B. subtilis* proteins was such that there must have been two protease cleavage sites. This was certainly the case in the *B. subtilis*  $\gamma$  protein, where two different new amino acids were seen in cycle 6.

|                                |   |    |
|--------------------------------|---|----|
| <u><i>B. megaterium</i></u> C: | 13  | 18 |
|                                | Arg-Asn-Ser-Ser-Asn-Lys-Leu-Val-Ala-Pro-Gly-Ala-Gln-Ala-Ala <sup>23</sup>   |    |
| <u><i>B. cereus</i></u> A:     | 5   | 10 |
|                                | Ser <sup>*</sup> -Arg-Ser-Thr-Asn-Lys-Leu-Ala <sup>*</sup> -Val <sup>*</sup> -Pro-Gly-Ala-Glu <sup>†</sup> -Ser <sup>*</sup> -Ala <sup>15</sup> |    |
| <u><i>B. megaterium</i></u> A: | 4   | 9  |
|                                | Ala-Asn-Thr-Asn-Lys-Leu-Val-Ala-Pro-Gly-Ser-Ala-Ala-Ala <sup>14</sup>   |    |

FIG. 3. Amino terminal primary sequence of *B. megaterium* proteins A and C and homologies with *B. cereus* protein A. The sequence of the *B. megaterium* proteins were taken from references 13 and 14. That for the *B. cereus* protein was from Table 4. Underlined residues are identical in the *B. cereus* protein and at least one of the two *B. megaterium* proteins; residues with asterisks can be generated from the amino acid in the comparable position in a *B. megaterium* protein by a single base change in DNA.

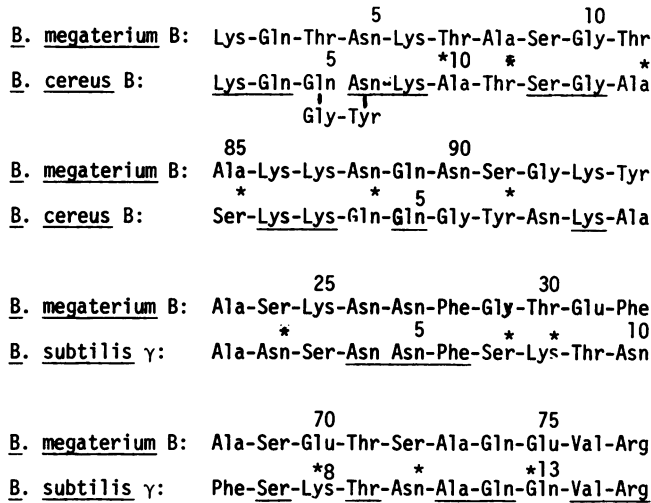


FIG. 4. Comparison of regions of the primary sequence of the *B. megaterium* B protein with those of the *B. cereus* protein B and *B. subtilis* protein  $\gamma$ . The sequence of the *B. megaterium* protein B was taken from reference 14a, and that for the *B. cereus* and *B. subtilis* proteins was from Table 4. Underlined residues are identical in the two sequences compared. Residues with an asterisk can be generated from the amino acid in the comparable position in the *B. megaterium* protein by a single base change in DNA.

evolutionary change is unclear, Ornston and his colleagues have suggested that such types of reorganization of protein sequence (possibly via some type of nonhomologous recombination) play a significant role in the evolution of certain types of genes (4, 18).

In contrast to the amino terminal regions of these proteins, where there is little sequence conservation, there is another region(s) where significant sequence conservation might be expected. This is the pentapeptide region recognized and cleaved by the *B. megaterium* spore protease which, in the *B. megaterium* proteins, is known to be involved in a highly specific interaction with the protease (1, 10). The cleaved site has a sequence of -Glu-Phe/Ile-Gly/Ala-Ser-Glu- in the *B. megaterium* proteins, with cleavage at the first glutamyl bond (arrow) (11). The *B. subtilis* and *B. cereus* proteins are substrates for the *B. megaterium* enzyme, and strikingly, the new amino terminal sequences generated by spore protease action on the *B. cereus* and *B. subtilis* proteins are very homologous (Fig. 2). Positions 2, 3, and 5 of the pentapeptide regions of the *B. cereus* and *B. subtilis* proteins cleaved by the *B. megaterium* protease exhibit no changes from the *B. megaterium* pattern, and only conservative changes occur in position 4, in only two out of three proteins. Thus, the substrate specificity of the enzyme is confirmed, and it is highly probable that the enzyme itself, with its specific function in germination, is conserved within this group of *Bacillus* species. As a consequence, the evolution of the primary sequence

of the soluble proteins is highly constrained at the site of endopeptidase action. These data further suggest that positions 2, 3, and 5 in the *B. megaterium* spore protease's site of recognition and cleavage are more important (or less flexible) than position 4.

#### ACKNOWLEDGMENTS

This work was supported by Public Health Service grants GM 19698 (P.S.) and AI 10806 (D.J.T.) from the National Institutes of Health. K.Y. was a predoctoral fellow of the National Institutes of Health.

We are grateful to Juris Ozols, Robert Strumpf, and Chester Siok for assistance with the protein sequencing and to Susan Goldrick and Michael Sadick for technical assistance.

#### LITERATURE CITED

- Dignam, S. S., and P. Setlow. 1980. *Bacillus megaterium* spore protease: action of the enzyme on peptides containing the amino acid sequence cleaved by the enzyme *in vivo*. *J. Biol. Chem.* **255**:8408-8412.
- Johnson, W. C., and D. J. Tipper. 1981. Acid-soluble spore proteins of *Bacillus subtilis*. *J. Bacteriol.* **146**:972-982.
- Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall. 1951. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**:265-275.
- Ornston, L. N., and W. K. Yeh. 1979. Origins of metabolic diversity: evolutionary divergence by sequence repetition. *Proc. Natl. Acad. Sci. U.S.A.* **76**:3996-4000.
- Reisfield, R. A., V. J. Lewis, and D. E. Williams. 1962. Disk electrophoresis of basic proteins and peptides on polyacrylamide gels. *Nature (London)* **195**:281-283.
- Setlow, B., and P. Setlow. 1979. Localization of low-molecular-weight basic proteins in *Bacillus megaterium* spores by cross-linking with ultraviolet light. *J. Bacteriol.* **139**:486-494.
- Setlow, P. 1975. Protein metabolism during germination of *Bacillus megaterium* spores. II. Degradation of pre-existing and newly synthesized protein. *J. Biol. Chem.*

- 250:631-637.
8. Setlow, P. 1975. Identification and localization of the major proteins degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **250**:8159-8167.
  9. Setlow, P. 1975. Purification and properties of some unique low molecular weight proteins degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **250**:8168-8173.
  10. Setlow, P. 1976. Purification of a highly specific protease from spores of *Bacillus megaterium*. *J. Biol. Chem.* **251**:7853-7862.
  11. Setlow, P., C. Gerard, and J. Ozols. 1980. The amino acid sequence specificity of a protease from spores of *Bacillus megaterium*. *J. Biol. Chem.* **255**:3624-3628.
  12. Setlow, P., and A. Kornberg. 1970. Biochemical studies of bacterial sporulation and germination. XXII. Energy metabolism in early stages of germination of *Bacillus megaterium*. *J. Biol. Chem.* **245**:3637-3644.
  13. Setlow, P., and J. Ozols. 1979. Covalent structure of protein A: a low molecular weight protein degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **254**:11938-11942.
  14. Setlow, P., and J. Ozols. 1980. Covalent structure of protein C: a second major low molecular weight protein degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **255**:8413-8416.
  - 14a. Setlow, P., and J. Ozols. 1980. The complete covalent structure of protein B. The third major protein degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **252**:10445-10450.
  15. Setlow, P., and G. Primus. 1975. Protein metabolism during germination of *Bacillus megaterium* spores. I. Protein synthesis and amino acid metabolism. *J. Biol. Chem.* **250**:623-630.
  16. Setlow, P., and W. M. Waites. 1976. Identification of several unique low-molecular-weight basic proteins in dormant spores of *Clostridium bifementans* and their degradation during spore germination. *J. Bacteriol.* **127**:1015-1017.
  17. Weber, K., and M. Osborn. 1969. The reliability of molecular weight determinations by dodecylsulfate-polyacrylamide gels. *J. Biol. Chem.* **244**:4406-4412.
  18. Yeh, W. K., P. Fletcher, and L. N. Ornston. 1980. Homologies in the NH<sub>2</sub>-terminal amino acid sequences of  $\alpha$ -carboxymuconolactone decarboxylases and muconolactone isomerases. *J. Biol. Chem.* **255**:6347-6354.