Cloning and nucleotide sequence of the gene encoding the γ-D-glutamyl-L-diamino acid endopeptidase II of *Bacillus sphaericus*

Marie-Laure Hourdou [1], Colette Duez [1], Bernard Joris [1], Marie-Jeanne Vacheron [2], Micheline Guinand [2], Georges Michel [2] and Jean-Marie Ghuysen [1]

[1] Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie, Sart Tilman, Belgium, and [2] Laboratoire de Biochimie microbienne, Université Claude Bernard Lyon I, Villeurbanne, France

1. SUMMARY

The gene encoding the *Bacillus sphaericus* γ-D-glutamyl-L-diamino acid endopeptidase II, a cytoplasmic enzyme involved in cell sporulation [1], contains the information for a 271-amino acid protein devoid of a signal peptide. The endopeptidase lacks sequence relatedness with other proteins of known primary structure except that its C-terminal region has significant similarity with the C-terminal region of the 54-kDa P54 protein of *Enterococcus faecium*, of unknown function [2].

2. INTRODUCTION

One essential step of sporulation in the bacilli is the synthesis of a cortex peptidoglycan [3] whose particular structure requires the sequential involvement of several peptidoglycan-degrading enzymes. Among the hydrolases that appear to be related to sporulation in *Bacillus sphaericus*, the γ-D-glutamyl-L-diamino acid endopeptidase II has been purified to protein homogeneity and its 38-amino acid N-terminal region has been sequenced [4]. It hydrolyses L-Ala-γ-D-Glu-L-Xaa-Y peptidoglycan peptide units in which L-Xaa is L-Lys, *meso*-A$_2$pm or ω-substituted *meso*-A$_2$pm; and Y is H, D-Ala or D-Ala-D-Ala. A strict requirement is the occurrence of a free N-terminal L-Ala [1,5]. In order to shed light on the functioning of the endopeptidase and its role in the cell cycle, the relevant gene has been cloned and sequenced, giving access to the primary structure.

Correspondence to: J.-M. Ghuysen, Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie, B6, B-4000 Sart Tilman (Liège 1), Belgium.

3. MATERIALS AND METHODS

The genomic DNA of *B. sphaericus* NCTC9602 was prepared from vegetative cells as described [6]. Plasmid pSP73 (Promega, Madison, WI) was used for cloning experiments and *Escherichia coli* HB101, grown at 37°C in Luria-Bertani medium, was used as host of the recombinant plasmids. Transformants were selected on ampicillin (50 μg ml^{-1}) containing agar plates. Genomic libraries were screened by hybridization using a modified Southern-blot procedure [7,8]. Other DNA-manipulation experiments were carried out as described [9].

Polymerase chain reaction (PCR) was performed according to [10,11] on a 100-μl sample containing the *B. sphaericus* chromosomal DNA (2 μg), the primers (0.5 μM, each), the dNTPs (200 μM, each), gelatine (10 μg) and the *Taq* DNA polymerase (2.5 U, Perkin Elmer-Cetus, Norwolk, CT). The sample, covered with mineral oil, was submitted to 30 amplification cycles on a New Brunswick automated thermal cycler: 1 min denaturation at 94°C, 1.5 min annealing at 50°C (Tm = 6°C), 1 min polymerization at 72°C.

DNA segments cloned into bacteriophage M13tg130 or tg131 (Amersham, U.K.) and double-stranded DNA segments were sequenced by the dideoxynucleotide chain termination method [12] using the T7 sequencing kit (Sequenase USB, Cleveland, OH or Pharmacia LKB Biotechnology AB, Uppsala, Sweden). Denaturation of double-stranded DNA was performed as described [13] and the sequencing reactions were initiated with synthetic oligonucleotides (Eurogentec, Liège, Belgium).

Homology searches (PIR data bank) were carried out by compairing pairs of proteins on a matrix in a graphical form [14], thus giving a rough but immediate estimate of the significance of any similarity. In this process, two sequences under comparison are divided into all possible segments of a given length and each segment of one protein is compared with each segment of the other protein. Each pair of amino acids being assigned a score, only those segments having a total score such that the probability that similarity occurs by chance is lower than 1 in 1000, are graphically represented. Amino acid alignments were made as described [15,16].

4. RESULTS

The degenerated oligonucleotide primers shown in Fig. 1, upper part, were synthesized on the basis of the known sequence of the 38-amino acid N-terminal region of the endopeptidase II
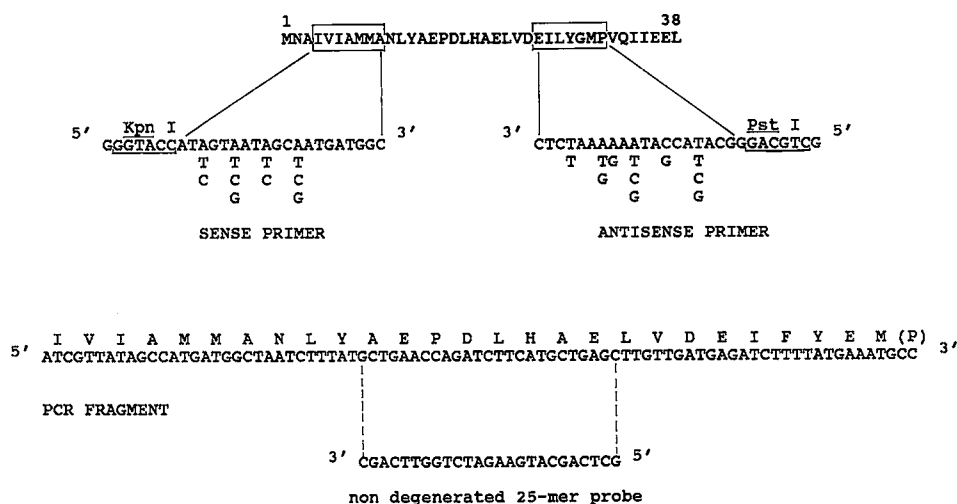

Fig. 1. DNA probe construction. The M1-L38 amino acid sequence is that determined by chemical sequencing of the purified endopeptidase II [4].

[4]. The sense oligonucleotide 1 had a *Kpn*I site at the 5'OH and encoded the I4VIAMM(A)10 sequence of the protein. The antisense oligonucleotide 2 had a *Pst*I site at the 5'OH and was complementary of the DNA segment encoding the E25ILYGM(P)31 sequence. The 93-bp DNA product obtained by PCR had the expected nucleotide sequence. On the basis of these data, the non-degenerated 25-mer probe encoding the A14EPDLHAE(L)22 amino acid sequence shown in Fig. 1, lower part, was synthesized, labelled with $\gamma[^{32}P]$-ATP and used to screen *Bam*HI, *Bgl*II and *Hin*dIII libraries of the *B. sphaericus* genomic DNA. The fragments were cloned in pSP73 (a 2464-bp plasmid that contains multiple cloning sites and the R-TEM β-lactamase-encoding gene). Among the 3500 ampicillin-resistant *E. coli* transformants thus obtained, one clone, from the *Hin*dIII library, gave a strong hybridization signal with the radioactive probe after washing the filters at 55°C (Tm = 1°C). That clone had a plasmid, called pBSII, which contained a 5-kb insert.

pBSII, prepared in large quantities, was cleaved with restriction enzymes and the DNA fragments, separated by agarose gel electrophoresis, were transferred to nitrocellulose membrane and submitted to hybridization with the radioactive DNA probe. The 1-kb *Hin*dIII fragment (i.e. the smallest fragment which gave a positive signal) and its two 350-bp and 650-bp *Sph*I–*Hin*dIII subfragments were cloned in M13tg130 and tg131. Sequencing, using the strategy shown in Fig. 2, led to the conclusion that the 1-kb *Hin*dIII fragment had the information for a major part (M1-L270) of the endopeptidase II. But it lacked that portion encoding the C-terminal end of the protein.

Given that the 1-kb *Hin*dIII DNA fragment contained an internal *Sph*I site, a *Sph*I genomic DNA library was made and the recombinant *E. coli* strains were screened using, as probe, the radioactive nucleotide 5'CGCCAAGATTTAGC-AACGAC3' that codes for the R257QDLAT (T)263 protein sequence (Fig. 3, below). One clone gave a hybridization signal. Sequencing of the double-stranded DNA revealed that immediately downstream from the nucleotide sequence coding for the I267GSL270 tetrapeptide, there was a F271 coding triplet, itself followed by a TAA stop codon.

The 1276-bp DNA segment had a 813-nucleotide open-reading frame (Fig. 3). The ORF starts with a ATG codon at position 223, terminates by a TAA codon at position 1036 and translates into a 271-amino acid protein that is devoid of signal peptide. The 38-amino acid N-terminal region had exactly the same sequence as that deter-
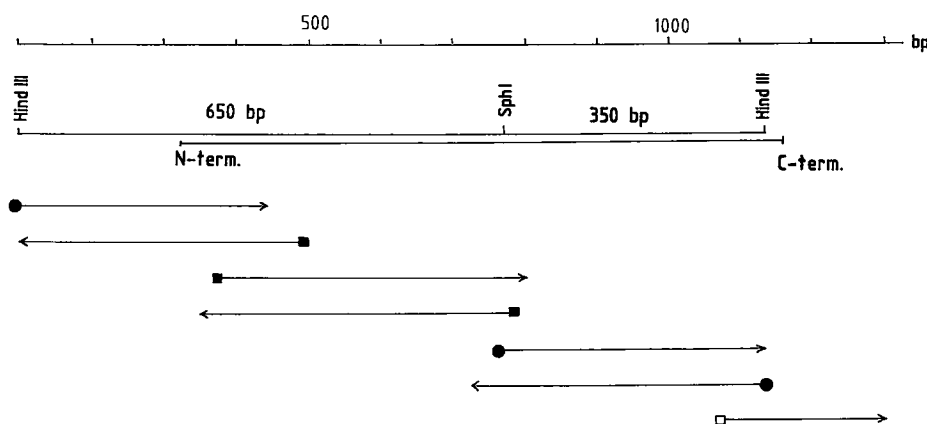

Fig. 2. Strategy of nucleotide sequencing. M13tg130 and tg131 were used to clone the 1-kb *Hin*dIII fragment and its 350-bp and 650-bp *Sph*I– *Hin*dIII subfragments. Nucleotide sequences initiated with the M13 universal primer are marked by ●; those initiated with synthetic oligonucleotides are marked by ■. The arrows indicate the orientation and length of the sequenced segments. □ symbolizes the primer used for screening the *Sph*I genomic library and for sequencing the C-terminal encoding part of the gene.

HindIII
AAGCTTGCGCCGTTATATCGTTGATGATGTGACAATCGCCCATAAAACAGGTGGCCTTGA 60

TACCGTTGACCACGATGTTGGCATTGTTTTTAGTCACGTACAAGATTATGGTATTGGTGT 120

CTTTATCTCAGAGGTCAAGCAAAATGATGTCGCCAGACAATTAATCGGTCGTCTTTCAAA 180

```
                                                1
                                         M  N  A  I  V  I
GGTAGTCTATGACAAGATGATAGGAGTGAAAGGAGAGACAACATGAATGCGATTGTCATT 240
```

```
               10                    20
  A  M  M  A  N  L  Y [A  E  P  D  L  H  A  E] L  V  D  E  I
GCTATGATGGCTAATCTTTATGCTGAACCAGATCTTCATGCTGAGCTTGTTGATGAAATA  300
```

```
               30                    40
  L  Y  G  M  P  V  Q  I  I  E  E  L  E  N  D  W  L  Y  V  R
CTTTACGGTATGCCTGTTCAAATTATAGAAGAATTAGAAAATGATTGGCTGTATGTTCGC  360
```

```
               50                    60
  T  A  Y  R  Y  E  G  Y  C  Q  R  N  D  V  L  F  D  D  A  I
ACTGCCTATCGCTATGAAGGTTATTGTCAGCGAAACGATGTCCTTTTCGATGATGCGATT  420
```

```
               70                    80
  T  N  T  W  I  Q  K  A  Q  H  V  I  G  Q  R  F  A  D  V  L
ACAAATACTTGGATTCAAAAGGCTCAACATGTTATTGGACAGCGTTTCGCAGATGTTTTA  480
```

```
               90                    100
  Q  E  P  K  I  Q  S  T  K  I  I  T  L  V  K  G  S  I  L  Y
CAAGAACCGAAAATACAAAGCACCAAAATCATCACATTGGTAAAAGGTTCTATTTTATAC  540
```

```
               110                   120
  N  V  D  S  D  T  T  S  N  T  P  W  T  A  V  Q  L  A  T  G
AATGTAGATTCCGACACGACATCAAACACGCCATGGACTGCCGTTCAATTAGCGACGGGC  600
```

```
               130                   140
  E  I  G  Y  L  R  S  Q  W  L  H  P  K  I  A  E  H  T  F  E
GAAATAGGCTATTTACGTTCACAATGGTTGCATCCTAAAATTGCTGAACATACTTTTGAA  660
```

```
               150                   160
  E  H  A  F  R  E  N  V  V  Q  T  A  L  S  Y  I  A  T  P  Y
GAGCATGCGTTTCGAGAAAATGTTGTACAAACAGCACTAAGCTATATAGCCACACCGTAT  720
  SphI
```

```
               170                   180
  R  W  G  G  K  S  P  L  G  I  D  C  S  G  L  C  S  M  A  Y
CGTTGGGGAGGTAAGTCACCGCTAGGCATTGATTGTTCGGGCTTATGTTCGATGGCTTAT  780
```

```
               190                   200
  L  L  N  G  V  I  I  F  R  D  A  R  I  V  E  G  F  P  I  K
CTGCTAAATGGCGTAATTATTTTCCGAGATGCTAGAATAGTAGAAGGTTTTCCAATTAAG  840
```

```
               210                   220
  E  I  T  I  D  R  M  Q  K  G  D  L  L  F  F  P  G  H  V  A
GAGATAACAATTGATCGTATGCAAAAAGGAGATTTGCTATTTTTCCCAGGGCATGTAGCA  900
```

```
  L  Y  L  G  Q  T  L  Y  V  H  A  S  L  G  G  N  E  V  N  V
CTGTATCTAGGGCAGACTCTTTATGTTCATGCCTCACTTGGTGGCAATGAGGTCAATGTC  960
```

```
               250               260
  N  S  L  D  E  Q  H  P  L  Y [R  Q  D  L  A  T  T] I  T  A
AATAGCCTAGATGAACAACACCCGTTATATCGCCAAGATTTAGCAACGACTATTACCGCA  1020
```

```
               270
  I  G  S  L  F  *
ATCGGAAGCTTGTTTAAATGTATGTAATGCTTGCAGAAAAATGGATGCAAAAAAACGAG  1080
  HindIII
GCATTTGACGGTAGTTGCATCGTCAAATACCTCGCCTTGAGTATGTTTACGCTAATATTG 2040

ATAAATATTTAACTGTTGTCCAGCTTCAATATTATCTGAACGCAATTGATTCCATTGTTT 2100

AATGGATGTTACTTTCACATTATGCGTCTGTGCGATGTGCATCAGCGAGTCACCATCCGC 2160

AACTGTAATGGTTTTT
```

Fig. 3. Nucleotide sequence of the gene encoding the *B. sphaericus* endopeptidase II and deduced amino acid sequence of the protein. The *Hind*III and *Sph*I restriction sites are underlined. The inverted repeats of the putative transcription signals are marked by horizontal arrows. The oligonucleotide boxed in full line is complementary to that used to screen and sequence the *Hind*III genomic library. The oligonucleotide boxed in dotted line was used to screen and sequence the *Sph*I genomic library.

```
          390       400       410       420       430
E.faecium P54   SGSVNGAAIVAEAYKYIGTPYVWGGKDPSGFDCSGFTRYVVYLQVT
                   |   | || ||| |   |||||  | |||  ||      ||
B.sph. endo II  EEHAFRENVVQTALSYIATPYRWGGKSPLGIDCSGLCSMAYLLNG
                150       160       170       180       190
```

```
          440       450       460       470
E.faecium P54   GRDIGGWTVPQESAGTKISVSQAKAGDLLFWGSAGGTYHVAISLG
                         ||||   |       | ||  |||     ||| ||
B.sph. endo II  VIIFRDARIVEGFPIKEITIDRMQKGDLLFFPG-----HVALYLG
                  200       210       220       230
```

```
          480       490       500
E.faecium P54   GGQYIHAPQPGENVKVGSVQWYTPDFAVSM
                   ||| |   ||| |
B.sph. endo II  QTLYVHASLGGNEVNVNSLDEQHPLYRQDLATTITAIGSLF
                  240       250       260       270
```

Fig. 4. Amino acid alignment of the E146-F271 polypeptide segment of the *B. sphaericus* endopeptidase II and the S388-M507 polypeptide segment of the *E. faecium* P54 protein. Identities are shown by vertical bars.

mined experimentally on the purified protein [4] and the calculated relative molecular mass 30 604 compared well with the 28-kDa value derived from SDS gel electrophoresis [4]. As a result of homology searches in the PIR data bank, the only protein which showed similarity, at least in part, with the *B. sphaericus* endopeptidase II was a 54-kDa protein of *Enterococcus faecium* of unknown function [2]. Similarity occurs at the level of a 98-amino acid segment that extends from V155 to S248 in the *B. sphaericus* endopeptidase and from V397 to 5495 in the *E. faecium* protein. The alignment shown in Fig. 4 reveals 36 identities for approx. 100 aligned amino acids. The unique cysteine, C420, present in the *E. faecium* protein aligns with C178 of the *B. sphaericus* enzyme.

5. DISCUSSION

At stage IV of sporulation, two peptidoglycans are deposited between the two membranes of opposite affinity which surround the forespore, within the mother cell [3]. A peptidoglycan of the vegetative type is synthesized by the forespore and assembled on the outer face of the inner membrane (thus forming the germ-cell wall which is retained on germination). A peptidoglycan of the cortex type is synthesized by the mother cell and assembled on the inner face of the outer membrane of the forespore. The cortex peptido-

glycan of *B. sphaericus* and other bacilli differs from the vegetative peptidoglycan by the nature of the diamino acid in the peptide units (*meso*-A$_2$pm vs. L-Lys) and by the presence, in the glycan chains, of an appreciable proportion of unsubstituted muramyl lactams and *N*-acetylmuramic acids bearing a single L-alanine residue [17]. These structural features require the coordinated and sequential action of several peptidoglycan hydrolases.

The *B. sphaericus* endopeptidase II is virtually non-existant in the vegetative cells. The enzyme appears at the onset of the sporulation, remains at a rather constant level throughout the process, and, as shown by cell fractionation, is found in the cytoplasm [1]. Consistently, its encoding gene lacks the information for a signal peptide. Given that the endopeptidase hydrolyses γ-D-Glu-L-(*meso*)A$_2$pm linkages only in those peptide units that have a free N-terminal alanine [1], its activity must follow that of an *N*-acetylmuramyl-L-alanine amidase. The occurrence of amidase activities in *Bacilli* is well documented [18,19]. It may be that in the course of sporulation in *B. sphaericus*, the nucleotide precursors are modified in the cytoplasm of the mother cell by the sequential action of the amidase and the endopeptidase before assembly of the cortex peptidoglycan. Alternatively, the amidase may act on a preformed cortex peptidoglycan, and the released peptidoglycan peptides may be transported in the cytoplasm of the mother cell, where the endopeptidase II is located. It is known that the Opp system of Gram-negative bacteria plays a major role in recycling cell-wall peptides as they are released from the growing peptidoglycan and that the *B. subtilis* oligopeptide transport system is involved in the initiation of sporulation [20].

The *B. sphaericus* endopeptidase II is sensitive to thiol-group reagents, suggesting that it might be a cysteine peptidase [4]. The occurrence of an odd number of cysteines at positions 55, 178 and 182, respectively, is compatible with such a mechanism. Yet, search in the PIR databank reveals that the endopeptidase II lacks sequence relatedness with the cysteine peptidases of known primary structure. The only detected similarity occurs between the C-terminal region of the *B.*

sphaericus endopeptidase and the C-terminal region of a 54-kDa protein of *E. faecium* [2]. This may be another example illustrating the concept according to which protein molecules are frequently of modular design and that each module folds into one domain that performs a particular function.

ACKNOWLEDGMENTS

REFERENCES

[1] Vacheron, M.J., Guinand, M., Françon, A. and Michel, G. (1979) Eur. J. Biochem. 100, 189–196.

[2] Fürst, P., Mösch, H.U. and Solioz, M. (1989) Nucleic. Acids Res. 17, 6724.

[3] Moir, A. and Smith, D.A. (1990) Annu. Rev. Microbiol. 44, 531–553.

[4] Bourgogne, T., Vacheron, M.J., Guinand, M. and Michel, G. (1992) Int. J. Biochem. (in press.)

[5] Valentin, C., Vacheron, M.J., Martinez, C., Guinand, M. and Michel, G. (1983) Biochimie 65, 239–245.

[6] Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985) Genetic manipulation of *Streptomyces*: a laboratory manual. The John Innes Foundation, Norwich, U.K.

[7] Wallace, R.B., Johnson, N.J., Hirose, T., Miyake, M., Kawashima, E.H. and Itakura, K. (1981) Nucleic. Acids Res. 9, 879–894.

[8] Woods, D. (1984) Focus 6, 1–3.

[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[10] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487–491.

[11] Lee, C.C., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.M. and Caskey, C.T. (1988) Science 239, 1288–1291.

[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[13] Zhang, H., Scholl, R., Browse, J. and Somerville, C. (1988) Nucleic Acids Res. 16, 1220.

[14] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic. Acids Res. 12, 387–395.

[15] McLachlan, A.D. (1971) J. Mol. Biol. 61, 409–424.

[16] Staden, R. (1982) Nucleic. Acids Res. 10, 2951–2961.

[17] Warth, A.D., and Strominger, J.L. (1969) Proc. Natl. Acad. Sci. USA 64, 528–535.

[18] Hobot, J.A. and Rogers, H.J. (1991) J. Bacteriol. 173, 961–967.

[19] Kuroda, A. and Sekiguchi, J. (1991) J. Bacteriol. 173, 7304–7312.

[20] Perego, M., Higgins, C.F., Pearce, S.R., Gallagher, M.P. and Hoch, J.A. (1991) Mol. Microbiol. 5, 173–185.