A putative new peptide synthase operon in *Bacillus subtilis*: partial characterization

Angelo Tognoni, Elisabetta Franchi, Claudio Magistrelli,
Emanuela Colombo, Paola Cosmina and Guido Grandi

Author for correspondence: Guido Grandi. Tel: +39 2 5205970. Fax: +39 2 52022974.

ENIRICERCHE S.p.A. –
Genetic Engineering and
Microbiology Laboratories,
Via F. Maritano, 26 San
Donato Milanese, 20097
Milan, Italy

A large operon-type structure has been located between the *gltA* and *citB* loci on the *Bacillus subtilis* chromosome. On the basis of the analysis of the 25 kb sequenced so far, it potentially encodes at least three large proteins which contain structural motifs associated with the subunits of all characterized peptide synthases. The amino acid recognition specificity of this new peptide synthase is discussed in the light of sequence homology with other synthases.

Keywords: *Bacillus subtilis*, peptide synthase operon, racemase

Micro-organisms produce a large number of peptides, generally as secondary metabolites, which have many important activities varying from antibacterial to antiviral or antifungal, and from anticarcinogenic to immuno-suppressive. Regardless of their function, these peptides are synthesized either using the usual ribosomal system or through the action of complex multi-enzyme systems known as peptide synthases.

It is now clear that although their products can vary considerably in length and amino acid composition, all peptide synthases studied so far share remarkable similarity in their structural organization and mechanism of action (Kleinkauf & von Döhren, 1990; Cosmina *et al.*, 1993). They are in fact organized into structural domains, each of which is responsible for the recognition and binding of a specific amino acid. Peptide synthesis takes place through subsequent reactions of thioester cleavage and amide bond formation.

Bacillus subtilis produces a number of peptides and lipopeptides, both linear and cyclic (Zuber *et al.*, 1993). Of these, only surfactin has been thoroughly characterized at the genetic level (Cosmina *et al.*, 1993; van Sinderen *et al.*, 1993). From gene sequence analysis and a variety of biochemical studies, it was found that surfactin synthase is a multi-subunit enzyme complex in which there are seven structural domains, corresponding to the number of amino acids present in surfactin.

In this communication we report the identification in the *B. subtilis* chromosome of a large operon encoding a new peptide synthase.

The 15.4 kb fragment present in λ clone λ B21 (Fig. 1), derived from a *B. subtilis* 168 chromosomal library (Ferrari

et al., 1981), was mapped by PBS1 transduction (Haldenwang *et al.*, 1980) to position 175°, approximately, on the *B. subtilis* genetic map. Using a 1.2 kb *Hind*III fragment of λ B21 to screen the *B. subtilis* library in λ FIXII (Kunst & Devine, 1991), two additional λ clones (λ 1A and λ 4B) were identified by DNA:DNA hybridization, the inserts of which overlap with the chromosomal fragment carried by λ B21 (Fig. 1). Finally, an additional 4.3 kb overlapping fragment was isolated by chromosome walking (Glaser *et al.*, 1993) after cloning the λ B21 *Hind*III–*Acy*I fragment into the suicide vector pDIA5304. Altogether, the three λ clones and the pDIA5304 derivative contained an uninterrupted chromosomal region of approximately 25 kb.

The identity of the 25 kb region carried by the λ and plasmid vectors with the corresponding chromosomal area was confirmed by both restriction enzyme and Southern blot analyses. In particular, when λ B21, λ 4B and λ 1A were digested with *Eco*RI and *Not*I, fragments of identical size were identified on an agarose gel. Furthermore, the large 3.0 kb *Eco*RI fragment of the three clones (Fig. 1) hybridized with an internal 32 P-labelled probe in a Southern blot experiment (data not shown). Finally, the Southern blot analyses of the chromosomal DNA digested with a variety of restriction enzymes and hybridized with a duplicated region (see below) located within the 7270 bp *Not*I–*Eco*RI fragment of λ B21 and within the 4149 bp *Eco*RI–*Not*I fragment of λ 1A, respectively, gave restriction patterns consistent with the data deduced from sequence analysis.

To confirm the map location of this 25 kb segment, two fragments, the 1.3 kb *Eco*RI–*Not*I fragment located at the 5' end of the region and the 1.2 kb *Xba*I–*Apa*I fragment located at the 3' end were used as probes to screen the YAC collection of ordered *B. subtilis* DNA fragments (Azevedo *et al.*, 1993), kindly provided by Dr D. Ehrlich.

The EMBL accession number for the nucleotide sequence data reported in this paper is Z34883.

Fig. 1. Restriction map and functional organization of the 25 kb sequenced region. The relative positions of the chromosomal inserts of the λ clones (1A, 4B and B21) and the pDIA5304 derivative (B22) are shown. Boxes indicate the restriction fragments used in Southern blot experiments for the screening of the λ library (fragment 2), the isolation of the pDIA5304 derivative B22 (fragment 3) and the analysis of the YAC library (fragments 1, 4 and 5). A, *Apal*; B, *Bam*HI; E, *Eco*RI; N, *Not*I; S, *Sma*I; Sl, *Sal*I; Xb, *Xba*I. ■, *Eco*RI–*Not*I fragment; ▨, *Hind*III fragment; ▩, *Hind*III–*Acyl* fragment; ▪, *Xho*I–*Apal* fragment; ▤, 1.2 kb repeated fragment.

Hybridization experiments showed that the region maps between *gltA* (177°) and *citB* (173°) and its orientation is such that the *Eco*RI–*Not*I fragment is proximal to *gltA* and the *Xho*I–*Apal* fragment maps next to *citB*. In fact, the *Eco*RI–*Not*I fragment hybridized with YAC15-19 which carries *gltA* whereas the other fragment hybridized with YAC15-37 which contains *citB* (the chromosomal fragments of the two YACs have overlapping regions).

The entire region was sequenced using a combination of two strategies. Restriction fragments covering approximately 30 % of the region were subcloned in pUC vectors and their nucleotide sequence was determined using a variety of sequencing kits, including Sequenase (USB), TaqTrack, fmol (Promega) and Circumvent (BioLabs). The remaining 70 % of the region was sequenced by the cycle-sequencing technique using the recombinant λ DNAs as templates (Saluz & Jost, 1989). The entire region was sequenced manually (radioactive labelling) on both strands with a mean of 1.5 readings per base.

The restriction map and the functional organization of the region as deduced by sequence analysis are shown in Fig. 1.

Six ORFs were identified, the first three of them, potentially encoding proteins of 463, 325 and 491 amino acids are located within the first 4 kb of the region. The presence of two putative transcription terminators downstream from *orf1* and *orf3* suggests that *orf2* and *orf3* might be in a common single transcription unit. Homology studies show that ORF3 shares 27 % identity and 15 % similarity with *E. coli* penicillin binding protein 4 (Korath *et al.*, 1991), but no significant homology was found between ORF1 and ORF2 and any of the protein

sequences of the Swiss-Prot data bank (release 28) using the FSTPSCAN (PCGENE) program.

Analysis of the three remaining ORFs, the last of which is only partially contained in the sequenced region, suggested that they are embedded in a large operon whose structure resembles the bacterial operons encoding peptide synthases (Turgay *et al.*, 1992; Cosmina *et al.*, 1993). The 2561 amino acid product of *orf4* is preceded by a putative ribosome binding site (RBS; GGAGG), 7 bases upstream from its TTG start codon. *orf5*, a gene encoding a protein of identical size to that of ORF4, starts after an intergenic region of 27 bp, whereas *orf6* is located only 16 bp downstream from the 3' end of *orf5* and encodes a protein of at least 859 amino acids. Both the intergenic regions between *orf4* and *orf5* and between *orf5* and *orf6* contain putative *B. subtilis* RBSs (AAAGGTGTGTGG and ATAGGAGAGGAG, respectively).

From the analysis of the sequence with the CLUSTAL and MACAW programs, it turns out that ORF4, ORF5 and ORF6 are organized into five structural domains (two in both ORF4 and ORF5, and one in ORF6) which are homologous not only to each other (with more than 20 % identity) but also to the structural domains of all peptide synthases characterized so far. In particular, the 4-phosphopantetheine cofactor binding site LGG(D/H)SH(I/L), the ATP binding sites SGTGKPRG, YGPTE, TGD and NGK, and the sequence HHILDGW delimiting the regions of the structural domains are highly conserved. Interestingly, the carboxy-terminal of both ORF4 and ORF5, spanning about 500 amino acids, shares homology with the carboxy-terminal of the SrfAORF1 and SrfAORF2 subunits of surfactin synthase, the TycA subunit of tyrocidine syn-

Fig. 2. Homology study among the structural domains of all peptide synthases characterized so far. The dendrogram has been drawn using the CLUSTAL program and taking into account the most conserved central region (approximately 180 amino acids in length spanning from the conserved sequences PKG and GELC/Y) of each domain which is supposed to be involved in amino acid recognition and binding (Cosmina *et al.*, 1993). The domains of peptide x synthase are indicated as ppsDOM(1–5).

thase, the GrsA subunit of gramicidin synthase, and ACV synthase. It has been suggested that this conserved sequence might be responsible for the racemase activity which changes the configuration of the amino acid bound to the upstream domain. If this is the case, it is reasonable to assume that the new peptide synthetase can potentially produce a peptide having D-amino acids at positions 2 and 4.

The 400 amino acid central region of the first structural domain in ORF4 and the last 400 amino acids of the partially sequenced ORF6 domain are almost identical, differing at only 19 positions. Since the domain of the central region contains most of the structural information for amino acid specificity (Cosmina *et al.*, 1993), we predict that domains 1 and 5 recognize and bind the same amino acid.

As far as the nature of the amino acids constituting the hypothetical peptide is concerned, Turgay *et al.* (1992) and Cosmina *et al.* (1993) have shown that in bacterial peptide synthases the domains which bind the same amino acid have specific and characteristic patterns of amino acid conservation. The comparative analysis of the central region of the five domains of the new peptide synthase with the same regions of the peptide synthases from *Bacilli* (Fig. 2) shows that domains 1 and 5 are mostly related to

the glutamic acid binding domain of surfactin synthase, domain 2 is remarkably similar to the ornithine binding domain of gramicidin synthase and domains 3 and 4 resemble the leucine binding domain of gramicidin synthase.

Preliminary data indicate that the operon is not functional in *B. subtilis* 168 under normal growth conditions although two putative σ^A promoters are present 38 bp (TTGTCA-17 bp-TACAAT) and 140 bp (TTCAAA-17 bp-TTTAAT) upstream from the *orf4* start codon, respectively. When active promoters are inserted upstream from *orf4*, the expression of protein species having the expected molecular mass can be detected on SDS-PAGE (data not shown). Purification and characterization of these proteins is in progress to help elucidate the activity of this peptide synthase.

We propose to name the new enzyme peptide x synthase and the coding genes *ppsorf1*, *ppsorf2* and *ppsorf3*.

ACKNOWLEDGEMENTS

We would like to thank Dr Anna Crispino for her contribution to the sequencing experiments. This work was financially supported partially by the EEC Bacillus Genome Sequencing Project and partially by ENIRICERCHIE S.p.A.

REFERENCES

- Azevedo, V., Alvarez, E., Zumstein, E., Damiani, G., Sgarbetta, V., Ehrlich, S. D. & Serror, P. (1993). An ordered collection of *Bacillus subtilis* DNA segments cloned in yeast artificial chromosomes. *Proc Natl Acad Sci USA* **90**, 6047–6051.
- Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G. & van Sinderen, D. (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol Microbiol* **8**, 821–831.
- Ferrari, E., Henner, D. & Hoch, J. A. (1981). Isolation of *B. subtilis* genes from a Charon 4A library. *J Bacteriol* **146**, 430–432.
- Glaser, P., Kunst, F., Arnaud, M., Coudart, M.-P., Gonzales, W., Hullo, M.-F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schweizer, J., Vertès, A., Rapoport, G. & Danchin, A. (1993). *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol Microbiol* **10**, 371–384.
- Haldenwang, G. W., Banner, C. D. B., Ollington, J. F., Losick, R., Hoch, J. A., O'Connor, M. B. & Sonenshein, A. L. (1980). Mapping a cloned gene under sporulation control by insertion of a drug resistance marker into the *Bacillus subtilis* chromosome. *J Bacteriol* **142**, 90–98.
- Kleinkauf, H. & von Döhren, H. (1990). Non-ribosomal biosynthesis of peptide antibiotics. *Eur J Biochem* **192**, 1–25.
- Korat, B., Mottl, H. & Keck, W. (1991). Penicillin binding protein 4 of *Escherichia coli*: molecular cloning of the *dacB* gene, controlled overexpression and alterations in murein composition. *Mol Microbiol* **5**, 675–684.
- Kunst, F. & Devine, K. (1991). The project of sequencing the entire *Bacillus subtilis* genome. *Res Microbiol* **142**, 905–912.
- Saluz, H. & Jost, J.-P. (1989). A simple high-resolution procedure to study DNA methylation and *in vivo* DNA–protein interactions on a single copy gene level in higher eukaryotes. *Proc Natl Acad Sci USA* **86**, 2602–2606.
- van Sinderen, D., Galli, G., Cosmina, P., de Ferra, F., Withoff, S.,

Venema, G. & Grandi, G. (1993). Characterization of the *srfA* locus of *Bacillus subtilis*: only the valine-activating domain of SrfA is involved in the establishment of genetic competence. *Mol Microbiol* **8**, 833–841.

Turgay, K., Krause, M. & Marahiel, M. A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* **6**, 529–546.

Zuber, P., Nakano, M. M. & Marahiel, M. A. (1993). Peptide antibiotics. In *Bacillus subtilis and Other Gram-positive Bacteria*, pp. 897–916. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Received 27 September 1994; revised 14 November 1994; accepted 22 November 1994.