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.

A large operon-type structure has been located between the gitA 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 λB21 (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 HindIII fragment of λB21 to screen the B. subtilis library in λFIXII (Kunst & Devine, 1991), two additional λ clones (λ1A and λ4B) were identified by DNA:DNA hybridization, the inserts of which overlap with the chromosomal fragment carried by λB21 (Fig. 1). Finally, an additional 4.3 kb overlapping fragment was isolated by chromosome walking (Glaser et al., 1993) after cloning the λB21 HindIII-AcyI 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 λB21, λ4B and λ1A were digested with EcoRI and NotI, fragments of identical size were identified on an agarose gel. Furthermore, the large 3.0 kb EcoRI fragment of the three clones (Fig. 1) hybridized with an internal 32P-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 7270bp NotI–EcoRI fragment of λB21 and within the 4149bp EcoRI–NotI fragment of λ1A, 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 EcoRI–NotI fragment located at the 5' end of the region and the 1.2 kb XhoI–ApaI 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.
Hybridization experiments showed that the region maps between \(gltA\) (177') and \(citB\) (173') and its orientation is such that the \(EcoRI-NotI\) fragment is proximal to \(gltA\) and the \(XbaI-ApaI\) fragment maps next to \(citB\). In fact, the \(EcoRI-NotI\) 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 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 (Korat 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 FSTSCAN (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 ATAGGAGAGG, 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 SGTTGKPRG, 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-
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 $\sigma^A$ 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 ENIRICERCHE S.p.A.

**REFERENCES**


van Sinderen, D., Galli, G., Cosmina, P., de Ferra, F., Withoff, S.,


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