A *Bacillus subtilis* chromosome segment at the 100° to 102° position encoding 11 membrane proteins

B. Roche, S. Autret, A. Levine, F. Vannier, N. Medina and S. J. Sror

Author for correspondence: S. J. Sror. Tel: +33 1 69 15 57 14. Fax: +33 1 69 15 78 08.
e-mail: seror@igmors.u-psud.fr

Institut de Gntique et
Microbiologie, URA CNRS
2225, Universit Paris XI
Btiment 409, 91405 Orsay
Cedex, France

The 25.9 kbp region upstream of *nprB* at 100°–102° on the *Bacillus subtilis* chromosome was sequenced. This revealed a known gene, *degA*, which was previously mislocated on the genetic map. A total of 29 putative ORFs were identified including a cluster of three ORFs whose products show clear homology with sulphate adenylyl pathway enzymes and, in addition, 11 ORFs whose products have one or more membrane domains, as indicated by their hydropathy profiles.

Keywords: *Bacillus subtilis*, genome sequencing, 100°–102° region, *degA*

As part of the European sequencing project of the *Bacillus subtilis* genome, a 60 kbp DNA fragment was assigned to our group. The work presented here describes the sequencing and cloning of 25.9 kbp immediately upstream of the *nprB*–*argJ* region. As described previously (Levine *et al.*, 1997), we used chromosome walking with the integrative plasmid, pDIA5304 (Glaser *et al.*, 1993). This allowed us to clone the whole region, except a 5 kbp fragment, which was apparently toxic in *Escherichia coli*, and therefore was obtained by long-range PCR. As described in more detail earlier (Levine *et al.*, 1997), the DNA fragments were subject to a nebulization procedure and 1–1.5 kbp fragments were subcloned into pUC18, identified by colony hybridization and sequenced in an automatic sequencer. The sequences were compiled using the Autoassembler program (Perkin Elmer). A part of our sequence is redundant with GenBank Z93940, as indicated in Fig. 1 (Oudega *et al.*, 1997). We searched for possible ORFs, using the Genemark program (Borodowsky & McIninch, 1993). Based on the presence of a putative Shine–Dalgarno sequence preceding an initiation codon ATG, GTG or TTG, 30 putative ORFs were detected. These putative ORFs are listed in Table 1, which also indicates the endpoints, the size of the proteins encoded, putative Shine–Dalgarno sequences and initiation codons. Eighteen of the putative ORFs are transcribed in the direction opposite to that of replication (Fig. 1).

We searched for homology for each of the putative ORFs, using the BLAST program (1.4.9 MP; Altschul

et al., 1990; non-redundant database). The results are shown in Table 1. In particular, note the presence of *degA* and the surrounding genes *yuxC* and *yuxD* (Bussey & Switzer, 1993), previously mislocated (276°) on the genetic map of *B. subtilis* and designated here *yisR* and *yisS* respectively. Three ORFs show strong similarities with sulphate adenylyl metabolism enzymes, *yisZ*, *yitA* and *yitB*, a paralog of the PAPS sulphotransferase (Arz *et al.*, 1994; Kaneko *et al.*, 1996; Mansilla & deMendoza, 1997). A terminator has been found downstream these three genes, suggesting that they could be organized in an operon. Other significant similarities were found with one esterase, *yisY* (Pelletier & Altenbuchner, 1995) and two putative ORFs from *Methanococcus jannaschii* and *Haemophilus influenzae*, *yitB* and *yitK*, respectively (Bult *et al.*, 1996; Fleischmann *et al.*, 1995). Surprisingly, three ORFs, *YisT*, *YisV* and *YisW*, show some homology with two proteins of the 35°–47° region of the *B. subtilis* genome (38% identity between *YisT* and gi\1881369, 50% identity with the N-terminal and C-terminal regions of gi\1881344 for *YisV* and *YisW*, respectively). The function of these two distant proteins is unknown. However, despite the availability of a larger number of genomes, no homology at all was found for 12 of the predicted ORFs in this region.

Using a hydropathy algorithm (Kyte & Doolittle, 1982), we also looked for putative membrane proteins. It is interesting to note that many ORFs in this region appear to encode membrane proteins. This includes a cluster of six ORFs, *yitM*–*yitR*, corresponding to the region which was apparently toxic and impossible to clone in plasmid pDIA5304 in *E. coli*. The hydropathy profiles for all the

Table 1. Putative ORFs in the *yisP*–*yitR* region and similarities of their predicted products to other proteins

ORF	Gene name	Endpoints (nt)	Product (aa)	SD consensus sequence (upper case) and initiation codon (bold)	Similar protein in database	BLAST score
1	<i>yisP</i>	2 > 655	218	Incomplete ORF	Phytoene synthase (<i>Synechococcus</i> sp.)	1.9e–17
2	<i>yisQ</i>	691 < 2055	455	gtagcacagattgGGGAA	P45272 (<i>Escherichia coli</i>)	1.3e–47
3	<i>yisR</i>	2179 > 3039	287	GGAGGatgattcaaatg	Previously <i>yuxC</i> (<i>Bacillus subtilis</i>)	
4	<i>degA</i>	3060 > 4070	337	gacGAGGGatgaacatg	Known gene	
5	<i>yisS</i>	4282 > 5307	342	GGAGAGgatgggacgtg	Previously <i>yuxD</i> (<i>Bacillus subtilis</i>)	
6	<i>yisT</i>	5364 < 5870	169	gtagtcgtacggGGGAA	gi\1881369 (<i>Bacillus subtilis</i>)	5.6e–42
7	<i>yisU</i>	5923 < 6582	220	gtgatttcAGAGG	YggA (<i>Escherichia coli</i>)	7.5e–17
8	<i>yisV</i>	6649 > 7359	237	AAGGGggcatgaccttg	gi\1881344 (<i>Bacillus subtilis</i>)	9.0e–76
9	<i>yisW</i>	7463 > 8101	213	GAcAGGcacagtcatg	gi\1881344 (<i>Bacillus subtilis</i>)	1.1e–77
10	<i>yisX</i>	8115 < 8750	212	gtacactactgtAGGAA	None	
11	<i>yisY</i>	8956 > 9759	268	aGGAGGaatgcggaatg	Arylesterase (<i>Pseudomonas fluorescens</i>)	9.5e–61
12	<i>yisZ</i>	9793 < 10389	199	gtaacgttaagGGGcGA	Adenylylsulphate kinase (<i>Arabidopsis thaliana</i>)	2.4e–72
13	<i>yitA</i>	10389 < 11363	325	gtagagaAAGtG	Sulphate adenylyltransferase (<i>Synechococcus</i> sp.)	2.6e–134
14	<i>yitB</i>	11671 < 12258	196	gtgttagaaaAGAGG	PAPS sulphotransferase (<i>Bacillus subtilis</i>)	8.4e–80
15	<i>yitC</i>	12563 > 13246	228	GGAGAGgtagtatg	None	
16	<i>yitD</i>	13246 > 14001	252	AGAGAGgggtcataatatg	U67480 (<i>Methanococcus jannaschii</i>)	9.7e–58
17	<i>yitE</i>	14052 < 14678	209	gtaattcagtaaacGAGGA	D83026 (<i>Bacillus subtilis</i>)	1.1e–12
18	<i>yitF</i>	14777 < 15889	371	gtgtgacccacGAtGGAA	Muconate cycloisomerase (<i>Acinetobacter calcoaceticus</i>)	7.6e–17
19	<i>yitG</i>	15901 < 17178	426	gtgaggaaaaGAAGA	NorA (<i>Staphylococcus aureus</i> ; quinolone resistance)	2.2e–12
20	<i>yitH</i>	17281 < 18126	282	gtattctagtGGAcGA	None	
21	<i>yitI</i>	18134 < 18580	149	gtattacaaGGAGG	None	
22	<i>yitJ</i>	18673 < 20508	612	gtaacagacGGAGG	D90908 (<i>Synechococcus</i> sp.)	4.8e–34
23	<i>yitK</i>	20825 < 21313	163	gtatacttttGGAGG	YA34 (<i>Haemophilus influenzae</i>)	1.5e–40
24	<i>yitL</i>	21460 > 22305	282	GGAAGgacgatagcatg	None	
25	<i>yitM</i>	22364 < 22921	186	gtttttatgactttgcGGAAG	None	
26	<i>yitN</i>	22945 < 23310	122	gtgggtaaGGAcGA	None	
27	<i>yitO</i>	23445 < 23654	70	gtatatcaataaAAGcGG	None	
28	<i>yitP</i>	23858 < 24391	178	gtaagctgacGAGGA	None	
29	<i>yitQ</i>	24763 > 25497	245	GAGAAgggatg	None	
30	<i>yitR</i>	25520 > 25810	97	AGGAGtctgtaagatg	None

Fig. 1. Physical map and gene organization of the 25.9 kbp region on the *B. subtilis* chromosome, 100°–102° area. The upper part of this map shows the location of the fragments cloned by chromosome walking (filled bar) and by long-range PCR (striped bar). The primers used are indicated (a, GGCCCATTAATGTGACTG) (b, GTATTCGATGCTTCTTC, near *nprB*). The numbers 1–30 indicate the ORFs, with the arrows indicating the orientation of the putative genes with respect to transcription. *degA* and its region already present in the database is also indicated (dotted bar).

Fig. 2. Hydropathy analysis of potential membrane proteins.

predicted membrane proteins are shown in Fig. 2. YisQ, YisU, YitE, YitG, YitM, YitQ and YitR appear to be multitopic membrane proteins with, in some cases, many transmembrane domains. YisV, YitN and YitO are predicted to have at least one transmembrane domain, whilst YitP has a predicted N-terminal signal sequence which, if cleaved, would result in secretion.

ACKNOWLEDGEMENTS

We are pleased to acknowledge support from CNRS, University Paris XI, the EU Biotechnology Programme (BIO2-CT94-2011) and GREG (contract no. 118/96).

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.
- Arz, H. E., Gisselmann, G., Schiffmann, S. & Schwenn, J. D. (1994). A cDNA for adenylyl sulphate (APS)-kinase from *Arabidopsis thaliana*. *Biochim Biophys Acta* 1218, 447–452.
- Borodowsky, M. & McIninch, J. D. (1993). Genemark: parallel gene recognition for both DNA strands. *Comput Chem* 17, 123–133.
- Bult, C. J., White, O., Olsen, G. J. & 37 other authors (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058–1073.

- Bussey, L. B. & Switzer, R. L. (1993).** The *degA* gene product accelerates degradation of *Bacillus subtilis* phosphoribosylpyrophosphate amidotransferase in *Escherichia coli*. *J Bacteriol* **175**, 6348–6353.
- Fleischmann, R. D., Adams, M. D., White, O. & 37 other authors (1995).** Whole-genome random sequencing of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Glaser, P., Kunst, F., Arnaud, M. & 14 other authors (1993).** *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol Microbiol* **10**, 371–384.
- Kaneko, T., Sato, S., Kotani, H. & 21 other authors (1996).** Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* **3**, 109–136.
- Kyte, J. & Doolittle, R. F. (1982).** A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.
- Levine, A., Vannier, F., Roche, B., Autret, S., Mavel, D. & Séror, S. J. (1997).** A 10·3 kbp segment from *nprB* to *argJ* at the 102° region of the *Bacillus subtilis* chromosome. *Microbiology* **143**, 175–177.
- Mansilla, M. C. & deMendoza, D. (1997).** α -Cysteine biosynthesis in *Bacillus subtilis*: identification, sequencing and functional characterization of the gene coding for phosphoadenylylsulfate sulphotransferase. *J Bacteriol* **179**, 976–981.
- Oudega, B., Vandenbol, M. & Koningstein, G. (1997).** A 12 kb nucleotide sequence containing the alanine dehydrogenase gene at 279 degrees on the *Bacillus subtilis* chromosome. *Microbiology* **143**, 1489–1491.
- Pelletier, I. & Altenbuchner, J. (1995).** A bacterial esterase is homologous with non-haem haloperoxidases and displays brominating activity. *Microbiology* **141**, 459–468.

Received 21 April 1997; revised 30 June 1997; accepted 8 July 1997.