A 28 kbp segment from the *spoVM* region of the *Bacillus subtilis* 168 genome

David Foulger and Jeffery Errington

Author for correspondence: Jeff Errington. Tel: +44 1865 275561. Fax: +44 1865 275556.
e-mail: erring@molbiol.ox.ac.uk

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

The sequence of a 28 kbp segment of DNA surrounding the *spoVM* gene of *Bacillus subtilis* 168 (lying at approximately 145° on the standard genetic map) has been determined. The region contains 27 ORFs, a number of which have predicted products significantly similar to proteins in sequence databases, particularly to proteins involved in macromolecular synthesis of nucleic acids, proteins and phospholipids. A pair of closely linked genes encode a likely serine protein phosphatase and a serine protein kinase, respectively. Such proteins play important regulatory roles in eukaryotic cells but are rare in prokaryotes.

Keywords: *Bacillus subtilis*, *spoVM*

INTRODUCTION

As part of the *Bacillus subtilis* genome sequencing project (Moszer *et al.*, 1996) we were assigned the region between *adeC* (130°) and *polC* (147°). In previous communications we have described the sequencing of several segments of DNA from this region (Daniel & Errington, 1993; Daniel *et al.*, 1994, 1996; Yanouri *et al.*, 1993). Here we describe the sequence of a 28 kbp segment of DNA, obtained by plasmid integration and excision, using as a starting point the previously characterized *spoVM* gene (Levin *et al.*, 1993). The sequence of the region is analysed and shown to contain genes with a wide range of functions.

METHODS

Bacterial strains. *B. subtilis* strain 168 (from C. Anagnostopoulos, INRA, Jouy-en-Josas, France) was used in all plasmid integration and excision experiments. Subcloning was performed in *Escherichia coli* strain DH5 α [Life Technologies; (F')*endA1 hsdR17(r_k⁻ m_k⁺) supE44 recA1 gyrA96 λ^- thi-1 relA1 Δ (*lacZYA-argF*)U169 ϕ 80 Δ *lacZ* Δ M15]. During chromosome walking experiments plasmids were excised and recovered in *E. coli* strains DH5 α or TP611 (from P. Glaser, Institute Pasteur, France; *thi thr leuB6 lacY1 tonA21 supE44 hsdR hsdM recBC lop11 cya-610 pcnB80 zad::Tn10*).*

DNA manipulations. The technique of chromosome walking was used to clone DNA from *B. subtilis* by the recovery of integrated plasmids from the chromosome (Niaudet *et al.*, 1982) as described by Errington (1990). DNA manipulations

were carried out using standard procedures (e.g. Sambrook *et al.*, 1989). Plasmids pSG1301 (Stevens *et al.*, 1992) or pSG2 (Fort & Errington, 1985) were used as the cloning vectors. To initiate chromosome walking downstream from *spoVM*, plasmid pSG2012 was constructed by cloning the 0.4 kbp *HindIII/XbaI* fragment from pSC603 (Levin *et al.*, 1993) into pSG1301. pSG2012 was then transformed into *B. subtilis* 168 and a transformant was used for chromosome walking. To initiate the chromosome walking upstream from *spoVM*, a second plasmid (pSG2013) was constructed by isolating the 1.3 kb *EcoRI/HindIII* insert from pSC600 (Levin *et al.*, 1993) and subcloning it into pSG1301. The resulting plasmid was then used for chromosome walking as described above.

Transformation. *B. subtilis* 168 was transformed by the method of Anagnostopoulos & Spizizen (1961) as modified by Jenkinson (1983). Transformants of *B. subtilis* containing integration plasmids were selected on Oxoid nutrient agar containing chloramphenicol (5 μ g ml⁻¹). *E. coli* strains DH5 α and TP11 were transformed by the method of Sambrook *et al.* (1989).

PCR. PCR products were generated from template DNA of *B. subtilis* 168 using the conditions described by Innis & Gelfand (1990) or long range PCR (GeneAmp XL PCR kit, Perkin Elmer).

Sequence determination. The clones obtained by plasmid integration and excision were sequenced by primer walking using custom synthesized oligonucleotide primers. Plasmid DNA was prepared for sequencing using Hybaid recovery plasmid mini preps or the Qiagen Plasmid Midi Kit. The PCR products were purified with a QIAEX II Gel Extraction Kit (Qiagen). The purified DNA was sequenced using the ABI PRISM Dye Terminator Cycle Sequencing Kit (Perkin Elmer) and the Applied Biosystems 373 automatic DNA sequencer (Perkin Elmer). All of the sequence was determined on both strands.

The EMBL accession number for the sequence reported in this paper is Y13937.

Sequence analysis. The DNA sequence was analysed using the Nucleotide Interpretation Program (Staden package) and protein databases were searched using FASTA and BLASTX (GCG). Comparison of the predicted protein sequences to the Prosite library of sequence motifs was searched using Motifs (GCG). TMPred was used to predict membrane-spanning regions and their orientation (Hofmann & Stoffel, 1993).

RESULTS AND DISCUSSION

Cloning and sequencing

The *spoVM*-encoded protein is required for a late step in spore development (Levin *et al.*, 1993) and the gene lies at 145° on a recent genetic map (Anagnostopoulos *et al.*, 1993). The sequence of the previously cloned and sequenced *spoVM* gene was used as a base from which to chromosome-walk by plasmid integration and excision (Fig. 1). To begin walking 0.4 kbp (pSG2012) and 1.3 kbp (pSG2013) segments of DNA from the *spoVM* region were subcloned into the integration plasmid pSG1301. Initial clones obtained extending downstream from *spoVM* showed a different restriction pattern from that obtained previously by Levin *et al.* (1993). In retrospect, it appears that plasmid pSC603 obtained by Levin *et al.* (1993) must have contained two non-contiguous *Hind*III fragments of DNA, rather than having arisen by partial digestion as the authors suggested. We confirmed that our physical map of the *spoVM* region was correct by PCR amplification of DNA from *B. subtilis* strain 168 using appropriately positioned synthetic oligonucleotide primers.

In a series of steps a 28 kbp segment of DNA extending both upstream and downstream of *spoVM* was obtained (Fig. 1). After each round of chromosome walking, we confirmed that the clones obtained were contiguous with previous clones, either by plasmid walking back through the region or by PCR amplification of DNA from *B. subtilis* strain 168 using appropriately positioned synthetic oligonucleotide primers. The plasmids obtained were sequenced by primer walking with synthetic oligonucleotides and an automatic DNA sequencer.

ORFs

The sequence obtained (27779 bp) was analysed for the presence of ORFs that would encode peptides of at least 67 aa (with the exception of *spoVM*, the product of which is only 27 aa long as described by Levin *et al.*, 1993) and 27 were identified. All except *yloA* and *rpmB* would be transcribed from left to right as shown in Fig. 1, which corresponds to the same direction as used by DNA replication forks traversing the region. The translation initiation signals for the putative ORFs are listed in Table 1. The initiation codon for most (17) of the ORFs was ATG, the next most frequent was TTG (5) and 4 ORFs used GTG (not including the two putative start sites of *yloL*; Table 1). Each start codon was preceded by a likely Shine–Dalgarno sequence; most of these were reasonably complementary to the 3' end of the 16S rRNA of *B. subtilis* (3' UCUUCCU-

Fig. 1. Schematic diagram of the *spoVM* region of the *B. subtilis* 168 chromosome. The putative ORFs are represented as thick arrows, the direction of the arrow indicating the predicted direction of translation. Putative transcription units are represented by thin arrows, again denoting the direction of transcription. Putative promoters are represented as vertical lines at the start of the transcription units, and possible transcription termination sites are represented as stem-loop structures. Plasmid inserts (thin horizontal lines) and PCR products (double arrows) used to generate the sequence are also shown. Plasmid clones used to start the chromosome walking are shown as dashed lines. Restriction sites used for plasmid construction are shown.

CCACUAG; Mountain, 1989). The spacing between the 3' end of the Shine–Dalgarno sequence (5' AGGAGG) and the first base of the initiation codon ranged from 11 to 5 bp. The most common spacings were 8 (25%), 6 (21%), 10 (18%) and 7 bp (14%).

ORFs with no known function

The amino acid sequence predicted for each ORF was searched for similarity to previously reported protein sequences using FASTA, BLASTX and Motifs. The results are summarized in Table 2. Three predicted products,

Table 1. Putative translation start signals

ORF	Endpoints (bp)*	Putative ribosome-binding site and translation start†		
<i>yloA</i>	592 < 625	aatgcaggaaga Atga GAGG	gtgTa	TTG cat atg
<i>yloB</i>	690 > 724	agagagcaggg GAg GGAcG	aGcgga	ATG aag ttt
<i>yloC</i>	3445 > 3479	aacagggag Aattg GGtGt	tttaaa	ATG ata cga
<i>yloD</i>	4552 > 4588	attgacgctacat Ac GGacG	aagaaccc	GTG cag ttg
<i>yloH</i>	5287 > 5326	atataagaaa Atgct GGAGG	TGaataatca	ATG tta gat
<i>yloI</i>	5577 > 5610	gtaaagtgcagaggg GGAGa	attca	TTG ctt aac
<i>yloJ</i>	6793 > 6827	ctctcaaaac Aa AccGGAGa	gcgctc	ATG aat ttt
<i>yloK</i>	9235 > 9271	aagatctcagctt GGAGG	gtAaCaaa	TTG gca gta
<i>yloL</i> ‡	9722 > 9758	gaactagcgg At AtgGaAGG	atgatttg	ATG acg aga
	9721 > 9755	tgaactagcggat At GGAAg	gatgat	TTG atg acg
<i>yloM</i>	10665 > 10698	gccggtgatgt Gtt AGGAGt	aaAca	ATG aaa aaa
<i>yloN</i>	12009 > 12045	ttgcagcatg AGAA GaAGG	gataagca	ATG gca gaa
<i>yloO</i>	13106 > 13143	agcgtcaagac GAg AcGAGG	TGATgaggg	TTG tta aca
<i>yloP</i>	13863 > 13901	tgctttacaagttg Aa GAGG	gtgaagatca	GTG cta atc
<i>yloQ</i>	15828 > 15862	aaagatgaat Aac AAGGAGG	gaAaat	ATG cct gag
<i>yloR</i>	16725 > 16763	ttaaagacag Aa AgccGAGG	TatTagcata	ATG ata aag
<i>yloS</i>	17452 > 17488	ttcctgtatgcag Agaa AGG	aatcagag	ATG aaa aca
<i>spoVM</i>	18169 > 18205	ttacgggacgtt At AGGAGG	gGAcaaaa	ATG aaa ttt
<i>rpmB</i>	18565 < 18530	gttttgtgca AGtg AGGAGG	gaAaCaa	ATG gca cgt
<i>yloU</i>	18788 > 18823	ctctatgcattcg AAG GAGG	aacTagt	GTG tcc att
<i>yloV</i>	19166 > 19201	acgaaccctg AGtt AGGAGG	aGtagga	TTG tct atc
<i>yloW</i>	20967 > 21001	agctctggttc at AAGGAGG	aatagc	ATG aaa tac
<i>ylpA</i>	21653 > 21689	tagcgaaaagg Gtc AGGAGG	aGccagtt	ATG ttt cgt
<i>ylpB</i>	22530 > 22569	aaaaaaattt tcg AGGAGc	gctaggttcac	GTG aaa cag
<i>ylpC</i>	24688 > 24726	tagtcttaatt Gtcc GGAtG	gtgTtagat	ATG aga aga
<i>ylpD</i>	25271 > 25306	caaaacattc At AAAGGtGG	aGcatgc	ATG aga ata
<i>ylpE</i>	26287 > 26326	acagatgagt AGtct GGAGG	TttTtcatc	ATG agt aag
<i>ylpF</i>	27235 > 27272	aattcaacgctt AAG GAGG	aGAatgata	ATG ctt aat
16S rRNA		3'-OH-UCUUUCCUCC	ACUAG	

* The direction of the ORF is indicated: > as DNA replication; < counter to DNA replication.

† Bases shown in bold upper case letters represent bases complementary to 16S rRNA. Upper case letters in plain type indicate the putative start codon.

‡ There are two potential start sites for *yloL*. The putative ribosome-binding sites and translation starts are shown for each ORF.

YloA, *YloS* and *YlpC*, shared no significant similarity to any known proteins in the databases. Several other sequences, *yloC*, *yloM*, *yloN*, *yloQ*, *yloU* and *yloV*, appear to encode proteins that are highly conserved in other bacteria but for which no function has yet been assigned. To try and obtain more information from the sequence of these proteins, the program TMpred was used to predict membrane-spanning regions and their orientation. *YloU* and *YloQ* have one strong putative transmembrane helix (N terminus inside the cell). *YloA*, *YloC*, *YloM*, *YloN*, *YloS* and *YloV* have no predicted transmembrane helices and they are therefore unlikely to be secreted or directly anchored to the membrane.

ORFs to which probable functions can be assigned

A substantial proportion of the ORFs showed strong similarity to genes of known function, allowing tentative functions to be assigned to the newly sequenced genes.

The functions uncovered were mainly associated with macromolecular synthesis. Thus, there were genes involved in synthesis of DNA (*priA*, *recG*), protein synthesis (*def*, *fnt* and *rpmB*) and phospholipid synthesis (*plsX*, *fabD* and *fabG*). Others were involved in intermediary metabolism (*dfp*, *araD* and *cfxE*) or nucleotide precursor synthesis (*gmk*). One is probably involved in the movement of cations across the membrane (*pacL*) and its product is predicted to have nine membrane-spanning helices, as has *PacL* from *Synechococcus* sp.

ORFs with unusual features

Two groups of genes found in the *spoVM* region are worthy of more extensive discussion. *yloO* and *yloP* encode proteins with signature motifs and similarity to families of eukaryotic proteins that act as serine protein phosphatases and kinases, respectively. The *yloO* and

Table 2. Properties of ORFs and their predicted products

ORF	Motifs	Size (aa)	Name of related gene	Description (putative product/function)	Database accession no.*	Percentage identity (over aa)
<i>yloA</i> †	–	203†	–	No significant identities found	–	–
<i>yloB</i>	E1–E2 ATPase phosphorylation site (cation transport ATPases)	890	<i>pacL</i>	Cation-transporting ATPase – <i>Synechococcus</i> sp. Calcium-transporting ATPase – <i>Rattus norvegicus</i>	gp:D16436 gp:M99223	45 (926) 33 (994)
<i>yloC</i>	–	291	–	Hypothetical protein – <i>E. coli</i> †	sp:P23839	30 (291)
<i>yloD</i>	(i) Guanylate kinase signature (ii) ATP/GTP-binding motif A (P-loop)	244	<i>guk1</i> <i>gmk</i>	5' Guanylate kinase – <i>Saccharomyces cerevisiae</i> 5' Guanylate kinase – <i>E. coli</i>	sp:P15454 sp:P24234	45 (183) 42 (199)
<i>yloH</i>	–	67	<i>rpoZ</i>	DNA-directed RNA polymerase Ω subunit – <i>E. coli</i> †	sp:P08374	31 (60)
<i>yloI</i>	–	406	<i>dfp</i>	Pantothenate metabolism flavoprotein – <i>E. coli</i> †	sp:P24285	41 (408)
<i>yloJ</i>	ATP/GTP-binding motif A (P-loop)	805	<i>priA</i>	Primosomal protein N' (replication factor Y) – <i>E. coli</i> †	sp:P17888	38 (579)
<i>yloK</i>	–	160	<i>def</i>	N-Formylmethionylaminoacyl-tRNA deformylase – <i>E. coli</i> †	sp:P27251	47 (143)
<i>yloL</i>	Phosphoribosyl-glycinamide formyltransferase active site	317	<i>fnt</i>	Methionylaminoacyl-tRNA formyltransferase – <i>E. coli</i> †	sp:P23882	46 (299)
<i>yloM</i>	NOL1/NOL2 Fmu family signature	447	<i>fmu</i>	Fmu protein – <i>E. coli</i> †	sp:P36929	33 (404)
<i>yloN</i>	–	363	–	Hypothetical 43.1 kDa protein – <i>E. coli</i> †	sp:P36979	36 (357)
<i>yloO</i>	–	254	–	Serine/threonine protein phosphatase – <i>Mycoplasma genitalium</i> Serine/threonine protein phosphatase – <i>S. cerevisiae</i>	sp:P47354 sp:P35182	28 (249) 25 (260)
<i>yloP</i>	(i) Protein kinase signature (serine/threonine catalytic activity) (ii) ATP/GTP-binding site motif A (P-loop)	648	<i>ptc1</i> <i>pkn2</i>	Serine/threonine protein kinase – <i>Mycobacterium tuberculosis</i> Serine/threonine protein kinase – <i>Myxococcus xanthus</i>	ge:Z80233.11 gp:M94857	34 (564) 35 (272)
<i>yloQ</i>	ATP/GTP-binding site motif A (P-loop)	298	–	Hypothetical 37.7 kDa protein – <i>E. coli</i> Hypothetical protein – <i>M. genitalium</i>	sp:P39286 sp:P47356	37 (229) 39 (229)
<i>yloR</i>	Ribulose-phosphate 3 epimerase family signature	217	–	Pentose-5-phosphate 3-epimerase precursor – <i>Solanum tuberosum</i> D-Ribulose-5-phosphate 3-epimerase – <i>Alcaligenes eutrophus</i> DOD protein – <i>E. coli</i>	sp:Q43843 sp:Q40539 sp:P32661	61 (218) 53 (215) 52 (213)
<i>yloS</i>	–	214	–	No significant identities found	–	–
<i>spoVM</i>	–	26	<i>spoVM</i>	Stage V sporulation protein M – <i>B. subtilis</i>	sp:P37817	100 (26)
<i>rpmB</i>	–	62	<i>rpmB</i>	50S Ribosomal subunit protein L28 – <i>B. subtilis</i> 50S Ribosomal subunit protein L28 – <i>E. coli</i>	sp:P37807 sp:P02428	100 (62) 44 (57)
<i>yloU</i>	–	120	<i>asp23</i>	Alkaline shock protein 23 – <i>Staphylococcus aureus</i>	gp:S76213	31 (106)
<i>yloV</i>	–	553	–	Hypothetical protein MG369 – <i>M. genitalium</i>	sp:P47609	31 (573)
<i>yloW</i>	–	220	<i>yhaQ</i>	L-Serine dehydratase part 1 – <i>E. coli</i>	sp:P42630	47 (73)
			<i>sdaA</i>	L-Serine dehydratase 1 – <i>E. coli</i> (see text)	sp:P16095	23 (212)
			<i>sdaB</i>	L-Serine dehydratase 2 – <i>E. coli</i> (see text)	sp:P30744	27 (223)
<i>yloA</i>	ATP/GTP-binding site motif A (P-loop)	244	<i>yhaP</i> <i>sdaA</i> <i>sdaB</i>	L-Serine dehydratase part 2 – <i>E. coli</i> Hypothetical protein ORF6† – <i>Lactobacillus sake</i> L-Serine dehydratase 1 – <i>E. coli</i> . (see text) L-Serine dehydratase 2 – <i>E. coli</i> . (see text)	sp:P42629 ge:X98238.1 sp:P16095 sp:P03744	37 (263) 64 (209) 43 (201) 42 (201)
<i>yloB</i>	AT/GTP-binding site motif A (P-loop)	684	<i>recG</i>	ATP-dependent DNA helicase – <i>E. coli</i> †	sp:P24230	39 (655)
<i>yloC</i>	–	188	–	No significant identities found	–	–
<i>yloD</i>	–	333	<i>plsX</i>	Fatty acid/phospholipid synthesis protein – <i>E. coli</i> Fatty acid/phospholipid synthesis protein – <i>M. genitalium</i>	gp:M96793 sp:Q49427	37 (326) 32 (328)
<i>yloE</i>	–	317	<i>fabD</i>	Malonyl CoA-acyl carrier protein transacylase – <i>E. coli</i> †	gp:M87040	44 (309)
<i>yloF</i> †	–	172†	<i>fabG</i>	3-Ketoacyl-acyl carrier protein reductase – <i>E. coli</i> †	sp:P25716	50 (171)

* gp, Genpept; sp, SWISS-PROT; ge, translation of a GenBank/EMBL DNA sequence. The number following an underscore denotes the position of the ORF in the sequence (i.e. _1 denotes the first ORF etc.).

† Partial coding sequence.

‡ A similar protein is found in *Haemophilus influenzae*.

yloP genes are adjacent and slightly overlapping, so it seems likely that their products act on a common target or targets. Proteins of this class are relatively rare in prokaryotes but possible homologues of *yloO* and *yloP* were found in other bacteria (Table 2). In eukaryotes these proteins generally act as regulators controlling a range of different functions (Hanks *et al.*, 1988; Hunter, 1995). In bacteria, there are two well characterized examples of serine phosphorylation. One system serves to regulate the activity of σ factors, such as σ^F (Errington,

1996; Stragier & Losick, 1996). The SpoIIAB kinase and SpoIIIE phosphatase have opposing activities that regulate the state of phosphorylation of a specific serine residue on the SpoIIAA protein. Accumulation of the non-phosphorylated form of SpoIIAA leads to release of σ^F activity. The other system involves regulation of the sugar transport system (PTS), which is effected in part by phosphorylation of the Hpr protein on a specific serine residue (Voskuil & Chambliss, 1996). Since *yloO* and *yloP* appear to be embedded in an operon that

contains a gene involved in sugar metabolism (*yloR/cfxE*), it is attractive to suppose that *yloP* encodes the missing serine protein kinase for Hpr.

The organization of the *yloO–yloQ* region is also interesting because there is a similar arrangement of ORFs in *Mycoplasma genitalium*. Thus, MG108 encodes a putative serine/threonine protein phosphatase, MG109 encodes a putative serine/threonine protein kinase and MG110 encodes a hypothetical protein similar to YloQ of *B. subtilis*. This region (135337–140022 bp) of the *M. genitalium* chromosome also contains an ORF similar to CfxE of *Alcaligenes eutrophus* and YloR of *B. subtilis*, which also lies in close proximity to the *yloO–yloQ* region (Fig. 1).

Another interesting question is posed by the two contiguous ORFs, *yloW* and *ylpA*. YloW is similar to YhaQ and to the N-terminal regions of two full-length L-serine dehydratases of *E. coli* (Table 2), while YlpA is similar to YhaP and the C-terminal regions of L-serine dehydratases of *E. coli*. Moreover the organization of *yloW* and *ylpA* in *B. subtilis* is similar to that of *yhaQ* and *yhaP* in *E. coli*. Although the alignment of YloW and YlpA overlaps in the central region of L-serine dehydratase, this is a region of only weak similarity (YloW shows 9% identity and YlpA shows 16% identity to SdaA of *E. coli* in this region). Database searches using a six-frame translation of the DNA sequence from the *yloW/ylpA* region, using BLASTX, revealed no evidence of frame shifts in the DNA sequence. Thus, they almost certainly represent two separate genes. It seems possible that they encode two subunits of an enzyme with a function related to L-serine dehydratase.

ACKNOWLEDGEMENTS

This work was funded by the BIOTECH programme of the European Community. We thank Annette Presscott and Alice Taylor for help with the DNA sequencing and Simon Cutting for providing the plasmids pSC600 and pSC603.

REFERENCES

- Anagnostopoulos, C. & Spizizen, J. (1961). Requirements for transformation in *Bacillus subtilis*. *J Bacteriol* **81**, 741–746.
- Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993). The genetic map of *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 425–461. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Daniel, R. & Errington, J. (1993). DNA sequence of the *murE–murD* region of *B. subtilis* 168. *J Gen Microbiol* **139**, 361–370.
- Daniel, R. A., Drake, S., Buchanan, C. E., Scholle, R. & Errington, J. (1994). The *Bacillus subtilis spoVD* gene encodes a mother-cell-specific penicillin-binding protein required for spore morphogenesis. *J Mol Biol* **235**, 209–220.
- Daniel, R. A., Williams, A. M. & Errington, J. (1996). A complex four-gene operon containing essential cell division gene *pbpB* in *Bacillus subtilis*. *J Bacteriol* **178**, 2343–2350.
- Errington, J. (1990). Gene cloning techniques. In *Molecular Biological Methods for Bacillus*, pp. 175–220. Edited by C. R. Harwood & S. M. Cutting. Chichester: Wiley.
- Errington, J. (1996). Determination of cell fate in *Bacillus subtilis*. *Trends Genet* **12**, 31–34.
- Fort, P. & Errington, J. (1985). Nucleotide sequence and complementation analysis of a polycistronic sporulation operon, *spoVA*, in *Bacillus subtilis*. *J Gen Microbiol* **131**, 1091–1105.
- Hanks, K. H., Quinn, A. M. & Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52.
- Hofmann, K. & Stoffel, W. (1993). TMbase – A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler* **374**, 166.
- Hunter, T. (1995). Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signalling. *Cell* **80**, 225–236.
- Innis, M. A. & Gelfand, D. H. (1990). Optimization of PCRs. In *PCR Protocols: A Guide to Methods and Applications*, pp. 3–12. Edited by M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White. London: Academic Press.
- Jenkinson, H. F. (1983). Altered arrangement of proteins in the spore coat of a germination mutant of *B. subtilis*. *J Gen Microbiol* **129**, 1945–1958.
- Levin, P. A., Fan, N., Ricca, E., Driks, A., Losick, R. & Cutting, S. (1993). An unusual small gene required for sporulation by *Bacillus subtilis*. *Mol Microbiol* **9**, 761–771.
- Moszer, I., Kunst, F. & Danchin, A. (1996). The European *Bacillus subtilis* genome sequencing project: current status and accessibility of the data from a new World Wide Web site. *Microbiology* **142**, 2987–2991.
- Mountain, A. (1989). Gene expression systems for *Bacillus subtilis*. In *Bacillus*, pp. 73–114. Edited by C. R. Harwood. New York: Plenum Press.
- Niaudet, B., Goze, A. & Ehrlich, S. D. (1982). Insertional mutagenesis in *Bacillus subtilis*: mechanism and use in gene cloning. *Gene* **19**, 277–284.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Stevens, C. M., Daniel, R., Illing, N. & Errington, J. (1992). Characterisation of a sporulation gene, *spoIVA*, involved in spore coat morphogenesis in *Bacillus subtilis*. *J Bacteriol* **174**, 586–594.
- Stragier, P. & Losick, R. (1996). Molecular genetics of sporulation in *Bacillus subtilis*. *Annu Rev Genet* **30**, 297–341.
- Voskuil, M. I. & Chambliss, G. H. (1996). Significance of HPr in catabolite repression of α -amylase. *J Bacteriol* **178**, 7014–7015.
- Yanouri, A., Daniel, R. A., Errington, J. & Buchanan, C. E. (1993). Cloning and sequencing of the cell division gene *pbpB*, which encodes penicillin-binding protein 2B in *Bacillus subtilis*. *J Bacteriol* **175**, 7604–7616.

Received 6 August 1997; revised 31 October 1997; accepted 7 November 1997.