DNA sequence of the *murE*–*murD* region of *Bacillus subtilis* 168

RICHARD A. DANIEL and JEFFERY ERRINGTON*

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

(Received 13 August 1992; revised 16 October 1992; accepted 28 October 1992)

The sequence of a 4.4 kbp region of DNA from *Bacillus subtilis* 168, lying between sporulation genes *spoVD* and *spoVE*, has been determined as part of the *B. subtilis* genomic sequencing programme. The region contains three genes with high sequence similarity to the *murE*, *mraY* and *murD* genes of *Escherichia coli*. The products of these genes are likely to catalyse various steps in the formation of the precursors for peptidoglycan synthesis in *B. subtilis*. The regions at 133° on the standard genetic map of the *B. subtilis* chromosome, and in the 2 min region of the *E. coli* genetic map, are now shown to contain a large cluster of functionally related genes. Although the linear order of the genes in the cluster is conserved, three genes that are present in the *E. coli* chromosome, and which are likely to be essential for peptidoglycan synthesis in both organisms, are absent from this region of the *B. subtilis* chromosome. In general, the *B. subtilis* cluster differs from that of *E. coli* in having more extensive intergenic regions, with less potential for translational coupling.

Introduction

Peptidoglycan (PG) forms the major structural component of the cell walls of many eubacteria. It is composed of a network of linear polysaccharide chains cross-linked by short peptides (reviewed by Rogers *et al.*, 1980). In both *Escherichia coli* and *Bacillus subtilis* the basic repeating structures are the same: a backbone of repeating disaccharides comprising *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc), linked by β -1,4-glycosidic linkages; and a tetra- or pentapeptide side chain linked to MurNAc, containing L-alanine, D-glutamate, *m*-diaminopimelic acid, and ending either in D-alanine, if the peptide participates in a cross-link, or D-alanyl-D-alanine, if not cross-linked.

The precursors of PG synthesis are likely to be similar in *E. coli* and *B. subtilis*, and their synthesis is thought to occur via a common series of intermediates, as illustrated in Fig. 1. Genes encoding the proteins catalysing most of these steps are clustered in the 2 min region of the *E. coli* chromosome, and have been extensively sequenced and characterized (Ikeda *et al.*, 1989, 1990*a, b*, 1991; Joris *et*

al., 1990; Mengin-Lecreux *et al.*, 1989, 1991; Mengin-Lecreux & van Heijenoort, 1990; Michaud *et al.*, 1990; Murayama *et al.*, 1988; Parquet *et al.*, 1989; Tao & Ishiguro, 1989). Recent results from various laboratories indicate that a functionally equivalent cluster of cell wall synthesis and cell division genes may be present in *B. subtilis*. Thus, likely functional homologues of the *ftsA*, *ftsZ*, *ftsW*, *murD*, *murG* and *pbpB* genes of *E. coli* have all been found in a cluster lying at about 133° on the standard genetic map of *B. subtilis* (Beall *et al.*, 1988; Beall & Lutkenhaus, 1989; Harry & Wake, 1989; Henriques *et al.*, 1992; Ikeda *et al.*, 1989; Joris *et al.*, 1990; R. A. Daniel, J. Errington & C. E. Buchanan, unpublished results). Moreover, the linear order of the various conserved genes so far identified is identical in the two organisms. However, of the genes involved in PG precursor synthesis one, *dll*, was shown some years ago to lie elsewhere on the *B. subtilis* chromosome, near another gene involved in alanine metabolism, *dal*, at about 36° (Buxton & Ward, 1980). Furthermore, Henriques *et al.* (1992) and Beall & Lutkenhaus (1989) have recently sequenced the region following the apparent *B. subtilis* homologue of *murG* but found no sign of the *murC* gene that follows *murG* in *E. coli*.

Here we present the sequence of the remainder of the *mur* operon of *B. subtilis*, including likely homologues of the *murE*, *mraY* and *murD* genes of *E. coli*. Unexpectedly, we did not find a likely homologue of *murF*, which lies between *murE* and *mraY* in *E. coli* (Ikeda *et al.*, 1990*a*; Parquet *et al.*, 1989).

* Author for correspondence. Tel. (0865) 275561; fax (0865) 275556; e-mail ERRING@VAX.PATH.OX.UK.AC.

Abbreviation: PG, peptidoglycan.

The nucleotide sequence data reported in this paper have been submitted to the EMBL Data Library and have been assigned the accession number Z15056.

Fig. 1. Summary of the likely enzymic steps in the synthesis of the precursors of PG in *E. coli* and *B. subtilis*, and the genes encoding the enzymes. Abbreviations: MurNAc, *N*-acetylmuramic acid; P_i, inorganic phosphate; DAP, diaminopimelic acid; GlcNAc, *N*-acetylglucosamine.

Methods

Bacterial strains, plasmids and bacteriophages. All of the plasmid integration and excision experiments were done on *B. subtilis* strain 168, which was obtained from C. Anagnostopoulos, INRA, Jouy en Josas, France. *E. coli* strain DH5 α was used for routine plasmid manipulations. Integration plasmids were based on the vector pSG1301 (Stevens *et al.*, 1992).

The isolation of phage ϕ 105J109 was described previously (East & Errington, 1989). Plasmid pPP73 (Pigot *et al.*, 1986) was used as the source of *spoVE* DNA for the chromosome walking experiments. The recovery of integrated plasmids from the *B. subtilis* chromosome (Niaudet *et al.*, 1982) was done essentially as described by Errington (1990).

General methods. *B. subtilis* strains were transformed according to 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).

DNA sequencing. Double-stranded plasmid DNA was used as the template for sequencing by the chain-terminator method of Sanger *et al.* (1977). Sequenase (United States Biochemicals) and ³⁵S-labelled nucleotides (Amersham) were used for the sequencing reactions, as recommended by the suppliers. Deleted plasmids for sequencing were made by restriction and religation or exonuclease III/S1 nuclease treatment as described in Sambrook *et al.* (1989). Synthetic oligonucleotides were used to complete the sequence determination of both DNA strands.

Results and Discussion

Cloning the promoter-proximal portion of the mur operon of B. subtilis

Phage ϕ 105J109 was isolated on the basis of its ability to complement the *spoVD156* mutation (East & Errington, 1989). Parts of the 5.4 kbp insert were subcloned readily and sequenced on both strands. The central portion of the cloned DNA segment encoded a single open reading frame that we have shown corresponds to the *spoVD* gene (R. A. Daniel and others, unpublished). The region immediately downstream from *spoVD* was cloned only with considerable difficulty, probably because it contains a strong vegetatively active promoter: similar problems have been reported for other genes recently; for examples, see Cutting *et al.* (1991) and Stevens *et al.* (1992). The region was eventually cloned by the plasmid integration and excision method (Niaudet *et al.*, 1982; see Methods), or 'plasmid walking'. Briefly, a plasmid grown in *E. coli*, but incapable of autonomous replication in *B. subtilis*, containing the distal end of the *spoVD* gene was transformed into *B. subtilis* strain 168. A transformant containing the plasmid integrated into the chromosome via its homologous sequence was purified and chromosomal DNA prepared. The DNA was digested with appropriately chosen enzymes cleaving outside the plasmid, ligated at low DNA concentration to favour recircularization, and transformed back into *E. coli* with selection for the plasmid. Transformants were relatively rare, in accordance with the notion that the desired DNA segment contained sequences that are deleterious to the growth of the *E. coli* host. However, transformants that grew well were obtained at low frequency. Some of these were found to contain the expected DNA fragment together with a second piece of DNA of unknown origin. Presumably, the additional fragment of DNA in some way counteracted the deleterious effects of the sequence downstream from *spoVD*.

Plasmids containing deletions of the DNA immedi-

Fig. 2. 'Gap repair' procedure to allow cloning of the *spoVD*–*spoVE* region of the *B. subtilis* chromosome. The integration plasmid used contained a selectable chloramphenicol resistance determinant *cat*, and two non-contiguous portions of the *B. subtilis* chromosome: internal fragments of the *spoVD* (R. A. Daniel & J. Errington, unpublished results) and *spoVE* (Piggott *et al.*, 1986) genes. These DNA fragments were arranged in the same orientation as their supposed orientations in the chromosome, although this was not initially certain. On transformation of *B. subtilis* strain 168 and selection for chloramphenicol resistance, some transformants arise by a process involving repair of the gap in the original plasmid, and are recognizable because they are both viable and *Spo*⁺. Although we have illustrated this as occurring by a double-crossover mechanism, it is possible that some other mechanism is responsible, e.g. single crossover, followed by gap repair. In the absence of repair, single-crossover insertions involving only the DNA homologous to the *spoVD* gene disrupt *spoVD* and give a *Spo*[–] phenotype. Most insertions involving only the homology to *spoVE* are lethal because they prevent transcription of the essential *murG* gene from the *mur* promoter (*murP*) (see Henriques *et al.*, 1992). Transformants with the structure shown were used to isolate extended portions of the *mur* operon by the excision–religation process shown in the lower part of the Figure. The chromosomal DNA was digested with a suitable restriction endonuclease (RE), ligated at low DNA concentration to favour the formation of circular DNA molecules, and transformed into *E. coli* with selection for a plasmid-borne resistance marker.

Fig. 3. Physical map of the *B. subtilis* *mur* region. The upper part of the Figure shows the restriction map of the *spoVD*–*spoVE* region, together with the inserts in $\phi 105J109$ and the various plasmid derivatives used for DNA sequence analysis. The asterisks indicate the locations of additional DNA fragments of unknown origin, inserted during the construction of plasmids pSG1002 and pSG1004. The lower part of the Figure shows a detailed map of the promoter-proximal part of the *mur* operon, as deduced by DNA sequence analysis. The abbreviations for restriction endonuclease cleavage sites are as follows: A, *Sma*I; B, *Bcl*I; E, *Eco*R I; G, *Bgl*II; H, *Hind*III; M, *Mlu*I; N, *Nco*I; P, *Pst*I; R, *Nar*I; S, *Sst*I; V, *Eco*RV.

---spoVD--->
 D L E S L L V N L N V D A S G K G S K I V K Q S P A A G T K V K E G S K I R V Y
 AGATCTTGAGTCACTATTGGTGAATCTCAATGTGGATGCTTCCGGGAAAGGGAGCAAATCGTTAAACAATCGCCAGCTGCCGGAACGAAAGTAAAGAAGGCTCAAAGATACGCGTGTGA
 BglIII 50 100 MluI

 End of spoVD
 L T E E D E K E A A D *
 TTTGACTGAAGAAGATGAAAGGAGGCAGCCGATTGATTCTCGGGCTGCCTATTCTTTTATTGCAGAAAAAATGCTGATATCTTTCAAGATTAGCTGGATGTATCCTTTTATTACGCTAAA
 150 EcoRV 200

 ---murE--->
 M K L T K L L T Y
 ATATGAACTGTATGCTTAAAAATCATTACCACATCATTGATGTGACTTCAATGATAAAAAGAACGAATTAGAATGTTGAGAGGTTTGATACGATGAAACTTACAAAGCTGCTTACATACT
 250 300 350
 L T T E P S V N D S Q D P E I T S I E M D S R E V K K G S L F V C V K G Y T V D
 TAACAACCGAACCTTCTGTAAACGACTCACAAGACCTGAAATAACTTCAATTGAGATGGACTCCAGGGAAGTAAAGAAAGGGAGTCTTTTTGTATGTGTAAAGGGTATACGGTGGATG
 400 450
 G H D F A Q K A V E N G A A A I V A E R E V D V N V P V I I V R Q S L R A L S V
 GCCACGATTTTGCAGAAAAGCAGTAGAAAACGGTGCAGCGGCTATTGTGGCAGAAAGAGAGGTCGATGTGAATGTTCTGTATTATTGTCCGCGAGTCACTTCGCGCGCTCAGCGTTT
 500 550 600
 L S D A F Y G Q P T K K L Q L I G I T G T N G K T S T T H M V D E I L K K A G K
 TGTCTGATGCGTTTTTACGGACAGCCGACGAAGAAGCTTCAGCTGATTGGGATAACAGGGACAAACGGCAAAACATCAACGACCCACATGGTCGATGAGATTCTGAAAAAGCGGGAAAC
 HindIII 650 700
 R T G L I G T M Y M K I G D E T L P V K N T T P E S V T L Q K T F K K M N D K H
 GCACAGGCTTGATCGGTACAATGTACATGAAAATCGGAGACGAAACACTTCCTGTAAAAACAGACTCCGGAAGCGCTCACGTCGAGAAAACGTTTAAGAAAATGAATGACAAGCAGC
 750 800 PstI
 V D T A I M E V S S H A L S L G R V H G C D Y D I A V F T N L T Q D H L D Y H K
 TTGACACTGCAATCATGGAGGTTTTCTCCACGCTTTGTCTTTAGGACGCGTCCATGGCTGTGATTATGATATTGCCGTGTTTACGAATTTAACACAGGACCATCTGGACTATCATAAA
 850 MluI NcoI 900 950
 T M D E Y R H A K S L L F S Q L G G A F N H E H P K R A V L N A D D E A S A Y F
 CGATGGATGAGTACCGGCATGCAAAAGCTTGCTCTTTTCCCAACTTGGCGGGGCTTTCAACCATGAACATCCAAAGCGTGCCGTCTGAATGCTGACGATGAGGCGAGCGCTTATTTTG
 HindIII 1000 1050
 E K V T A A H I S T Y G I K N D A D V M A K N I S I T A Q G T S F D L V T N K G
 AAAAGGTAACAGCTGCGCATATCTTACATACGGAATAAAAACGATGCTGATGTGATGGCGAAGAACATCAGTATCACTGCACAGGGCACAAAGCTTTGATCTCGTGACAAACAAAGGCA
 1100 1150 HindIII 1200
 T K H I T M S L V G Q F N V Y N V L A A V A T C I A A G I P F E I I T E A V E E
 CGAAGCACATCAGATGTCTCTCGTCGACAATTTAATGTGTATAACGTGCTGGCCGCTGTTGCGACATGTATTGCAGCAGGCATACCGTTTGAATCATCACTGAGGCTGTTGAAGAAT
 1250 1300
 L H G V R G R F E L V N Q Q Q E F P V I V D Y A H T P D S L E N V L E T C R D M
 TGCATGGTGAAGAGGCCGTTTTGAATTGGTCAATCAGCAGCAGGAGTTTCTGTCTATTGTAGACTACGCGCACACACCGGACAGCCTTGAAAATGTGCTGGAGACATGCAGAGACATGA
 1350 1400
 T E G K L F V V V G C G G D R D K T K R P K M A K I A V E L A D E P I F T S D N
 CGGAAGGAAAGCTATTTGTCGTTGTGCGCTGCGGAGGAGACAGAGATAAGACGAAGCGCCCGAAAATGGCTAAAATTGCTGTGGAGCTGGCTGACGAGCCGATTTTCACGTCAGATAACC
 1450 1500 1550

P R S E D P R A I L R D M E A G V E N A Y Y H S I A N R E Q A I F F A I A N A K
 CGAGAAGCGAAGATCCTCGCGGATTTTAAGAGATATGGAAGCCGAGTAGAGAACGCATACATCAGTATTGCAAACCGCGAGCAGCGATCTTTTCGCAATTGCCAATGCCAAAA
 1600 1650
 K G D V V L I A G K G H E T Y Q Q I G N E T F D F D D A E V A A R A I V E L N K
 AAGGCGAGCTTGTCTCATTGCCGAAAAGGACATGAAACCTATCAGCAGATCGGAAATGAACTTTTCGATTTTGATGATGCTGAAGTAGCGGCGAGAGCCATTGTTGAACTCAATAAGA
 1700 1750 1800
 End of *murE*
 N K T N S *
 ATAAGACGAATTCATAGGAAAGAGATGTTACTGAATAATGGAGCTTGAAAAAGATTTTTTCATATGGAAAAAATCATGAATATGATAACTACGTATGTCGTTTGAAAGAAACAGGAGGAG
 EcoRI 1850 1900
mraY--->
 M L E Q V I L F T I L M G F L I S V L L S P I L I P F L R R L K F G Q S I
 AAAAGGACAATGCTTGAGCAAGTCATTCTGTTTACAATTTTAATGGGGTTTTTAATTAGTGTCTGCTCTCTCCGATTCTTATTCGTTTTTAAGAAGATTAAAATTCGGCCAGAGTATT
 1950 2000
 R E E G P K S H Q K K S G T P T M G G V M I I L S I I V T T I V M T Q K F S E I
 AGAGAAGAAGGACCGAAATCACATCAGAAAAAATCAGGGACACCGACAATGGGCGGGGTCAATGATCACTTTCTATCATAGTGACAACAATTGTTATGACACAGAAGTTTTTCAGAAATA
 2050 2100 *BclI* 2150
 S P E M V L L L F V T L G Y G L L G F L D D Y I K V V M K R N L G L T S K Q K L
 AGCCCCGAAATGGTGTCTGCTTCTGTTTGTACGCTAGGCTACGGTTTGCTTGGCTTTTTAGATGATTACATCAAGTTTGTATGAAGCGCAATCTTGGATTGACATCAAAGCAAAGCTG
 2200 2250
 I G Q I I I A V V F Y A V Y H Y Y N F A T D I R I P G T D L S F D L G W A Y F I
 ATCGGACAAATTATTATTGCGGTTGATTTTACGCCGTGTATCATTACTACAATTTTGCGACGATATTTCGATTCCTGGAAGTACTTATCATTTGATCTTGGCTGGGCTTACTTTATC
 2300 2350 2400
 L V L F M L V G G S N A V N L T D G L D G L L S G T A A I A F G A F A I L A W N
 CTTGTGCTCTTTATGCTAGTCGGCGGATCAAACGCAGTTAACCTGACTGACGGCCTTGACGGGTATTATCCGGTACTGCGGCGATTGCCTTTGGCGCCTTTGCCATTCTGGCATGGAAT
 2450 *NarI* 2500
 Q S Q Y D V A I F S V A V V G A V L G F L V F N R D P A K V F M G D T G S L A L
 CAGTCTCAATATGACGTAGCGATTTTCTCAGTTGCCGTTGTCTGGTGCAGTTCTGGGCTTCCTTGTATTTAATCGTGATCCGGCCAAAGTTTTTATGGGAGATACGGGATCGCTTGCATTG
 2550 2600
 G G A I V T I A I L T K L E I L L V I I G G V F V I E T L S V I L Q V I S F K T
 GGAGGAGCCATCGTTACCATTGCCATTTTAACGAAATTAGAGATCCTGCTGGTTATCATCGGCGGTGATTTCGTTATCGAGACATTATCCGTTATTTGTCAGGTCATCAGCTTTAAAACG
 2650 2700 2750
 T G K R I F K M S P L H H H Y E L V G W S E W R V V V T F W A A G L L L A V L G
 ACAGGTAAACGAATCTTTAAATGAGTCCGCTTCATCACCATTATGAGCTTGTGGCTGGTCTGAATGGAGAGTAGTCGTGACGTTTGGGCTGCGGGACTTTTGCTTGCCGTTTTAGGA
 2800 2850
 End of *mraY* *murD*--->
 I Y I E V W L * M E N D Q F L Q K Q H F L I L G L A K S G Y A A A S I L H E K G
 ATTTACATCGAGGTGTTTATAAGTGGAATAATGATCAATTTTACAGAAACAGCATTTTTTAATCCTTGGGCTTGCGAAAAGCGGATATGCGGCGGCTCCATTCTTTCATGAAAAAGGC
 2900 *BclI* 2950 3000
 I Y V A V N D Q K P F E E N E P A Q K L S E K G I E V V C G E H P V S L F D Q H
 ATTTATGTTGCAGTCAATGATCAAAAGCCTTTTGAAGAAAATGAACCGGCACAAAAGCTTCTGAAAAAGGAATCGAAGTTGTTTGCAGAGAGCATCCAGTTTCTGTTTCGATCAGCAT
 BclI 3050 *HindIII* 3100

Fig. 4. For legend see page 366.

Q I T I L I K N P G I P Y E N I M V Q E A E K R G I P V W T E I E L A Y Y L T S
 CAAATCAGGATTCGATCAAAAACCGGGAATTCCTTATGAAATATCATGGTTCAGGAAGCAGAGAAACGAGGTATCCGGTGGACGGAATGAGCTCGCTTACTACCTGACCAGC
 . BclI . SmaI 3150 EcoRI . 3200 . SstI .
 A K F I G I T G S N G K T T T T T L I Y E M L K A D S Q K A L I A G N I G T V A
 GCCAAGTTTATCGGGATTACCGGATCGAACGGTAAACGACCACGACAACGCTAATTACGAAATGCTCAAGGCTGACTCGCAAAAAGCACTGATTGCCGGAACATCGGCACAGTCGCA
 3250 . 3300 . 3350 .
 S E V A Y H A D G D E W I V T E L S S F Q L M G T H A F R P E I S L I L N V F D
 AGTGAGGTTGCCTATCATGCTGATGGAGATGAATGGATCGTTACCGAAGTCTTCGTTTCAGCTGATGGGAACACACGCTTTCAGACCGGAAATCAGCCTGATTTAAATGTATTTGAC
 . 3400 . 3450 .
 A H L D Y H H T R E N Y E K A K Q K V Y L H Q T A S D K A I V N Q D D E T V V R
 GCCCATCTGGATTATCACCATACGCGGAAACTATGAAAAAGCGAAACAGAAAGTGTACCTTCAGACTGCTTCAGATAAAGCTATTGTCAACCAGGATGATGAAACGTCGTGCGA
 . 3500 . 3550 . 3600 .
 L A E A G K A E I V P F S V S K T L E Q G A Y V K D S M I M F N G E A I L P L E
 CTTGCTGAAGCCGGAAGCCGAAATCGTCCCATTTTCAGTCAGCAAAACGCTTGAACAGGGCGCATATGAAAAGACAGTATGATTATGTTTAAACGGCGAAGCCATTCTTCTCTGGAA
 . 3650 . 3700 .
 E V V L P G A H N L E N I L A A I A V V K T A G A S N E A V K K V L T S F T G V
 GAGGTGTACTGCCCGGTGCTCATAACCTGGAGAACATTTTAGCAGCTATTGCAGTTGTGAAAACGGCAGGAGCATCCAATGAGGCGGTAAAAAGGTGCTGACAAGCTTTACTGGTGTG
 . 3750 . 3800 . HindIII .
 K H R L Q Y V T T V N G R K F Y N D S K A T N I L A T S K A L S A F D K P V I L
 AAGCACCGCTCCAATATGTCACAACTGTGAATGGCCGAAATTTATAACGACAGTAAAGCAACAAACATTCTGGCGACAAGCAAGGCGCTTTCAGCTTTTGATAAGCCGGTCATTCTG
 3850 . 3900 . 3950 .
 L A G G L D R G N G F D D L K P Y M K H V K A V L T F G Q T A P K L E K L G N E
 CTTGCGGGCGGACTGGATCGCGGAAATGGTTTTGATGACTTGAAACCTTATATGAAGCATGTCAAAGCAGTGCTTACATTGGCCAGACGGCGCCAAACTCGAGAAACTCGGCAATGAA
 . 4000 . 4050 NarI .
 L G I Q H V K R V D N V E Q A V S A A F A L S N E G D V I L L S P A C A S W D Q
 CTGGGAATACAACATGTCAAACGTGTCGATAATGTTGAACAAGCAGTATCTGCGGCGTTTGCCTCTCAAATGAAGGAGATGTCATCCTGCTGTACCCGCGTGTGCGAGCTGGGATCAA
 . 4100 . 4150 . 4200 .
 F K T F E E R G D M F I D A V H M L K *
 TTCAAACATTTGAAGAACGCGGTGACATGTTTATAGATGCCGTGCATATGCTTAAGTAAGGCTTGTCTTGAAGTAAATCGGCAGCCCTAATGACTTTAAGCTTGGGGTGTGATTGCT
 . 4250 . 4300 HindIII .
spoVE---->
 M T T K K T S P D L L L V I I T L L L L T I G L I M V Y S A S A V W A D Y K F D
 TTGACTACTAAAAAACATCGCCTGATTTGTTATTGGTCATCATTACGTTATTATTATTAAACATCGGATTAATTATGGTGTACAGTGCAGTGCAGTATGGGCGGATTATAAATTGAC
 . 4350 . 4400 .

Fig. 4. DNA sequence of the promoter-proximal portion of the *mur* operon of *B. subtilis*. The sequence shown begins near the end of the *spoVD* coding region. The complete sequences of the *murE*, *mraY* and *murD* are given, then the beginning of *spoVE*. Putative ribosome-binding sites are shown in boldface type, and certain restriction endonuclease recognition sites (those shown in Fig. 3) are underlined. Some of the more extensive inverted repeat sequences are shown with arrows below the sequence. Amino acid motifs in *murE* and *murD* that are typical of nucleotide-binding proteins are boxed.

ately adjacent to *spoVD* became stable in *E. coli*, even in the absence of the 'stabilizing' fragment of DNA. In principle, having obtained a stable fragment of DNA from beyond the deleterious DNA segment, it should have been possible to use the plasmid walking method to clone further fragments of DNA extending away in the downstream direction from *spoVD*. However, plasmids containing fragments of DNA from this region could not be transformed into *B. subtilis*, presumably because they disrupt an essential transcription unit (Piggot *et al.*, 1984). On the basis of known linkage data between *spoVD* and *spoVE*, and preliminary DNA sequence analysis (see below), we reasoned that *spoVE* probably lay 5 to 10 kbp downstream of *spoVD*. We therefore devised a 'gap-repair' adaptation of the plasmid walking method to clone the intervening DNA sequences. An integrative plasmid was constructed containing segments of DNA from the *spoVD* and *spoVE* regions in their supposed natural orientation. On transformation of wild-type *B. subtilis* strain 168, transformants could arise by several mechanisms, one of which is represented in Fig. 2. Single-crossover events involving only the homology to *spoVE* should not arise because most such insertions are known to be lethal (Henriques *et al.*, 1992). Similar events causing integration at *spoVD* disrupt the gene, giving rise to a *Spo*⁻ phenotype. Integrational events resulting in repair of the gap (Fig. 2), would, in contrast, be both viable and *Spo*⁺. Stable *Spo*⁺ transformants were readily obtained. DNA was purified from one such transformant, and Southern blotting and hybridization used to confirm that the gross structure of the insertion was as shown in Fig. 2. This strain was then used to recover plasmids containing inserts extending upstream of *spoVE* and overlapping those already obtained by plasmid walking downstream from *spoVD*. The region of DNA between *spoVD* and *spoVE* is shown in Fig. 3, together with the various plasmids from which this information was derived.

DNA sequence

Plasmids derived by subcloning from ϕ 105J109 or by plasmid walking (Fig. 3), were sequenced on both strands, and the sequence of a 4.44 kb segment of DNA extending from the promoter-distal end of the *spoVD* gene to the beginning of the *spoVE* gene is shown in Fig. 4. The sequence between the two sporulation genes contained three long open reading frames (ORFs). All of the predicted amino acid sequences showed significant sequence similarities to protein sequences in the Swiss-prot and EMBL databases. Alignments based on the use of a diagonal matrix comparison (Staden, 1982) are shown in Fig. 5. The first ORF showed significant similarity to the *murE* gene of *E. coli*, and the second, to

Fig. 5. Alignments of the amino acid sequences of the predicted products of *B. subtilis* (Bs) and *E. coli* (Ec) genes: (a) *murE*; (b) *mraY*; (c) *murD*. The sequence alignments are based on diagonal matrix comparisons (Staden, 1982). Hyphens indicate gaps introduced to improve the alignments. Asterisks indicate positions where the two proteins have the same amino acid.

Fig. 6. Comparison of the chromosomal organization of the '2 min region' of *E. coli* and the '133° region' of *B. subtilis*. The two regions are drawn approximately to scale, with boxes representing coding regions and connecting lines, non-coding regions, where these exist. The spacing between each pair of coding sequences is given above the *B. subtilis* map and below that of *E. coli*. The spacing is defined as the number of nucleotides between the last nucleotide of the termination codon and the first nucleotide of the initiation codon. Homologous sequences are indicated by broken lines. Between the aligned chromosomal regions, the predicted protein sequences of the genes dealt with in this paper are compared. The number given by each gene is the M_r of the predicted product, with the number of amino acids in parentheses. The percentage of identical residues for each pair of proteins, taken from Fig. 5, is also shown. The *pbbB* and *spoVD* gene products are also closely related (R. A. Daniel and others, unpublished), as are those of *spoVE* and *ftsW* (Ikeda *et al.*, 1989; Joris *et al.*, 1990), and the *murG* homologues (Henriques *et al.*, 1992). In *E. coli*, *murG* is followed by *murC* and *ddl* (Ikeda *et al.*, 1990b), but these genes are not present in this part of the *B. subtilis* chromosome (Beall & Lutkenhaus, 1989; Henriques *et al.*, 1992). The function of the gene in this position of the *B. subtilis* chromosome (*orf2*) is unknown. The organization and sequence of this region of the *E. coli* chromosome is described in the following references: Ikeda *et al.* (1989, 1990a, b, 1991), Joris *et al.* (1990); Mengin-Lecreulx *et al.* (1989, 1991), Mengin-Lecreulx & van Heijenoort (1990), Michaud *et al.* (1990), Murayama *et al.* (1988), Parquet *et al.* (1989) and Tao & Ishiguro (1989).

the *mraY* gene of *E. coli*. The distal portion of the third ORF corresponded to the *murD* gene of *E. coli*, as described previously by Henriques *et al.* (1992). The sequence of the entire *murD*-like gene is given here, together with the beginning of the *spoVE* gene, to facilitate linking this sequence to the previously reported sequences. Our sequence was in complete agreement with that previously published by Henriques *et al.* (1992).

Given the likelihood that these sequences correspond to the functional homologues of the *E. coli* *mur* genes we propose to designate them according to their *E. coli* counterparts, in accordance with recent practice (e.g. Beall *et al.*, 1988; Henriques *et al.*, 1992).

Predicted gene products

The properties of the predicted products of the sequenced genes are summarized in Fig. 6, and compared with those of their *E. coli* counterparts. The *murE* gene of *B. subtilis* was similar in size to that of *E. coli*. The two proteins shared 40% identity over 456 aligned residues. Several short stretches of sequence showed particularly high conservation. Among such sequences were motifs found in many ATP-binding proteins (boxed in Fig. 4).

The *mraY* gene product was the most closely related to its *E. coli* counterpart, of the genes in this cluster, showing almost 50% sequence identity. The *B. subtilis* protein was somewhat smaller, mostly due to residues 'missing' from the N-terminal region, and from a single

central gap. Ikeda *et al.* (1991) noted that the *E. coli* *MraY* protein shows limited sequence similarity to the *Saccharomyces cerevisiae* ALG7 protein, which encodes the supposed dolichyl-phosphate-dependent GlcNAc-1-phosphate transferase (Hartog & Bishop, 1987). The *B. subtilis* *MraY* protein also shows limited sequence similarity to this protein, suggesting that the sequence similarity represents a functional relationship (data not shown).

The *B. subtilis* *murD* gene was slightly larger than that in *E. coli*. This pair of proteins showed rather less conservation than the other pairs in the cluster. However, as in the case of *murE*, sequence motifs characteristic of ATP-binding proteins represented some of the most highly conserved regions.

Characterization of these *B. subtilis* genes may have important consequences for further biochemical analyses of the proteins because the regions of particularly extensive sequence similarity are presumably particularly important functionally.

Non-coding DNA

The intergenic regions in and around the *mur* operons of *B. subtilis* and *E. coli* are also summarized in Fig. 6. All of the *E. coli* genes overlap to some degree. In several cases the stop codon of one gene overlaps the start codon of the next in the sequence ATGA, giving an overlap defined as '-4' in the Figure. In marked contrast, there

are gaps of 60 to 175 bp between all of the *B. subtilis* genes except between *mraY* and *murD*, in which the stop and start codons are adjacent. The overlap between the *E. coli* genes strongly suggests that translational coupling will occur (Gold & Stormo, 1987), and that the proteins could be made in roughly stoichiometrically equivalent amounts. In *B. subtilis*, translational coupling is absent, allowing for different amounts of protein to be made, and even for the presence of transcriptional terminators or internal promoters. However, we have not attempted to measure the transcriptional regulation of this gene complex.

Immediately after the end of the *spoVD* coding region is a large inverted repeat, which ends in a T-rich sequence. This seems likely to function as a rho-independent transcription terminator. Curiously, the TGA stop codon of the *spoVD* gene overlaps a potential initiation codon TTG, which is preceded by a good ribosome-binding site sequence. Although translation might be expected to reinitiate here to produce a short polypeptide (30 amino acids), this could only occur if transcription continued through the apparent rho-independent terminator. We have no explanation for this interesting sequence feature.

As discussed above, the following region caused extreme instability when cloned in *E. coli*, and so this region probably corresponds to the promoter for the *murE* gene and perhaps one or more of the following genes. In the intergenic region there were several sequences similar to the '−10' consensus recognized by RNA polymerase containing the major vegetative sigma factor, σ^A , but no sequence similar to the expected '−35' consensus was apparent.

As noted above, integrational plasmids containing internal fragments of the *mur* complex could not be transformed into *B. subtilis*, in support of the notion that the genes are coordinately regulated. Henriques *et al.* (1992) have previously shown that the *murD*, *spoVE* and *murG* genes are coordinately transcribed from a promoter somewhere upstream of *murD*. We suggest that the major promoter for this operon lies upstream of *murE*, and that the *murE*, *mraY* and *murD* genes comprise the first three genes in this operon.

Genome organization

As shown in Fig. 6, the chromosomal regions lying at about 2 min on the standard *E. coli* map (Bachman, 1990) and at about 133° on that of *B. subtilis* (Piggot *et al.*, 1990) are closely related. In addition to the genes shown, homologues of the *ftsA* and *ftsZ* genes of *E. coli* are found nearby, just to the right of the region shown in Fig. 6 (Beall *et al.*, 1988; Harry & Wake, 1989). The linear order of all of the conserved genes is the same in

both organisms. In *E. coli*, this cluster includes all of the genes needed for the formation of the pentapeptide side chain of the peptidoglycan precursor, together with the *mraY* and *murG* genes encoding the proteins catalysing the first two steps of the lipid cycle reactions (see Fig. 1). The peptide side chains of the PG precursors of *E. coli* and *B. subtilis* are similar, being synthesized by the sequential addition of the following amino acids to UDP-MurNAc: L-alanine, D-glutamate, *m*-diaminopimelate and D-alanyl-D-alanine. The first three bonds are formed by the sequential action of the *murC*, *murD* and *murE* products. The final dipeptide is first synthesized by the action of the *ddl* (D-alanine:D-alanine ligase) product, and then added by the *murF* product (see Fig. 1). All of these gene products lie in the cluster of genes in the 2 min region of the *E. coli* chromosome and are probably coordinately regulated. Remarkably, however, the genes encoding three of the ligases needed for the formation of the peptide side chain are absent from the *B. subtilis* cluster: *murF*, as reported here, and *murC* and *ddl*, as is apparent when the nucleotide sequence of this region of the *B. subtilis* chromosome is assembled from published sequence data (Beall & Lutkenhaus, 1989; Henriques *et al.*, 1992). It seems likely that *B. subtilis* must contain homologues of the *murC* and *murF* genes. Since both are concerned with ligations involving D- or L-alanine, it is tempting to speculate that they may lie close to the *dal* and *ddl* genes.

This work was funded by the EC SCIENCE programme. We thank David Foulger and Adriano Henriques for helpful comments on the manuscript.

References

- ANAGNOSTOPOULOS, C. & SPIZIZEN, J. (1961). Requirements for transformation in *Bacillus subtilis*. *Journal of Bacteriology* **81**, 741–746.
- BACHMAN, B. J. (1990). Linkage map of *Escherichia coli* K-12, edition 8. *Microbiological Reviews* **54**, 130–197.
- BEALL, B. & LUTKENHAUS, J. (1989). Nucleotide sequence and insertional inactivation of a *Bacillus subtilis* gene that affects cell division, sporulation, and temperature sensitivity. *Journal of Bacteriology* **171**, 6821–6834.
- BEALL, B., LOWE, M. & LUTKENHAUS, J. (1988). Cloning and characterization of *Bacillus subtilis* homologs of *Escherichia coli* cell division genes *ftsZ* and *ftsA*. *Journal of Bacteriology* **170**, 4855–4864.
- BUXTON, R. S. & WARD, J. B. (1980). Heat-sensitive mutants of *Bacillus subtilis* 168 blocked at three different stages of peptidoglycan synthesis. *Journal of General Microbiology* **120**, 283–293.
- CUTTING, S. M., ROELS, S. & LOSICK, R. (1991). Sporulation operon *spoIVF* and the characterization of mutations that uncouple mother-cell from forespore gene expression in *Bacillus subtilis*. *Journal of Molecular Biology* **221**, 1237–1256.
- EAST, A. K. & ERRINGTON, J. (1989). A new bacteriophage vector for cloning in *Bacillus subtilis* and the use of $\phi 105$ for protein synthesis in maxicells. *Gene* **81**, 35–43.
- ERRINGTON, J. (1990). Gene cloning techniques. In *Molecular Biological Methods for Bacillus*, pp. 175–220. Edited by C. R. Harwood & S. M. Cutting. Chichester, UK: Wiley.
- GOLD, L. & STORMO, G. (1987). Translation initiation. In *Escherichia*

- coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, pp. 1302–1307. Edited by F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter & H. E. Umbarger, Washington, DC: American Society for Microbiology.
- HARRY, E. J. & WAKE, R. G. (1989). Cloning and expression of a *Bacillus subtilis* division initiation gene for which a homolog has not been identified in another organism. *Journal of Bacteriology* **171**, 6835–6839.
- HARTOG, K. O. & BISHOP, B. (1987). Genomic sequence coding for tunicamycin resistance in yeast. *Nucleic Acids Research* **15**, 3627.
- HENRIQUES, A. O., DE LENCASTRE, H. & PIGGOT, P. J. (1992). A *Bacillus subtilis* morphogene cluster that includes *spoVE* is homologous to the *mra* region of *Escherichia coli*. *Biochimie* **74**, 735–748.
- IKEDA, M., SATO, T., WACHI, M., JUNG, H. K., ISHINO, F., KOBAYASHI, Y. & MATSUHASHI, M. (1989). Structural similarity among *Escherichia coli* FtsW and RodA proteins and *Bacillus subtilis* SpoVE protein, which function in cell division, cell elongation, and spore formation, respectively. *Journal of Bacteriology* **171**, 6375–6378.
- IKEDA, M., WACHI, M., ISHINO, F. & MATSUHASHI, M. (1990a). Nucleotide sequence involving *murD* and an open reading frame ORF-Y spacing *murF* and *ftsW* in *Escherichia coli*. *Nucleic Acids Research* **18**, 1058.
- IKEDA, M., WACHI, M., JUNG, H. K., ISHINO, F. & MATSUHASHI, M. (1990b). Nucleotide sequence involving *murG* and *murC* in the *mra* gene cluster region of *Escherichia coli*. *Nucleic Acids Research* **18**, 4014.
- IKEDA, M., WACHI, M., JUNG, H.-K., ISHINO, F. & MATSUHASHI, M. (1991). The *Escherichia coli mraY* gene encoding UDP-N-acetylmuramoyl-pentapeptide:undecaprenylphosphate phospho-N-acetylmuramoyl-pentapeptide transferase. *Journal of Bacteriology* **173**, 1021–1026.
- JENKINSON, H. F. (1983). Altered arrangement of proteins in the spore coat of a germination mutant of *Bacillus subtilis*. *Journal of General Microbiology* **129**, 1945–1958.
- JORIS, B., DIVE, G., HENRIQUES, A., PIGGOT, P. J. & GHUYSEN, J. M. (1990). The life-cycle proteins RodA of *Escherichia coli* and SpoVE of *Bacillus subtilis* have very similar primary structures. *Molecular Microbiology* **4**, 513–517.
- MENGIN-LECREULX, D. & VAN HEIJENOORT, J. (1989). Nucleotide sequence of the *murD* gene encoding the UDP-MurNAc-L-Ala-D-Glu synthetase of *Escherichia coli*. *Nucleic Acids Research* **18**, 183.
- MENGIN-LECREULX, D., PARQUET, C., DESVIAT, L. R., PLA, J., FLOURET, B., AYALA, J. A. & VAN HEIJENOORT, J. (1989). Organization of the *murE-murG* region of *Escherichia coli*: identification of the *murD* gene encoding the D-glutamic-acid-adding enzyme. *Journal of Bacteriology* **171**, 6126–6134.
- MENGIN-LECREULX, D., TEXIER, L., ROUSSEAU, M. & VAN HEIJENOORT, J. (1991). The *murG* gene of *Escherichia coli* codes for the UDP-N-acetylglucosamine:N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. *Journal of Bacteriology* **173**, 4625–4636.
- MICHAUD, C., PARQUET, C., FLOURET, B., BLANOT, D. & VAN HEIJENOORT, J. (1990). Revised interpretation of the sequence containing the *murE* gene encoding the UDP-N-acetylmuramyl-tripeptide synthetase of *Escherichia coli*. *Biochemical Journal* **269**, 277–278.
- MURAYAMA, I. N., YAMAMOTO, A. H. & HIROTA, Y. (1988). Determination of gene products and coding regions from the *murE-murF* region of *Escherichia coli*. *Journal of Bacteriology* **170**, 3786–3788.
- NIAUDET, B., GOZE, A. & EHRLICH, S. D. (1982). Insertional mutagenesis in *Bacillus subtilis*: mechanism and use in gene cloning. *Gene* **19**, 277–284.
- PARQUET, C., FLOURET, B., MENGIN-LECREULX, D. & VAN HEIJENOORT, J. (1989). Nucleotide sequence of the *murF* gene encoding the UDP-MurNAc-pentapeptide synthetase of *Escherichia coli*. *Nucleic Acids Research* **17**, 5379.
- PIGGOT, P. J., CURTIS, C. A. M. & DE LENCASTRE, H. (1984). Use of integrational plasmid vectors to demonstrate the polycistronic nature of a transcriptional unit (*spoIIA*) required for sporulation of *Bacillus subtilis*. *Journal of General Microbiology* **130**, 2123–2136.
- PIGGOT, P. J., CHAK, K. F. & BUGAICHUK, U. D. (1986). Isolation and characterization of a clone of the *spoVE* locus of *Bacillus subtilis*. *Journal of General Microbiology* **132**, 1875–1881.
- PIGGOT, P. J., AMJAD, M., WU, J.-J., SANDOVAL, H. & CASTRO, J. (1990). Genetic and physical maps of *Bacillus subtilis* 168. In *Molecular Biological Methods for Bacillus*, pp. 493–543. Edited by C. R. Harwood & S. M. Cutting. Chichester, UK: Wiley.
- ROGERS, H. J., PERKINS, H. R. & WARD, J. B. (1980). *Microbial Cell Walls and Membranes*. London: Chapman & Hall.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SANGER, F., NICKLEN, S. & COULSEN, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- STADEN, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Research* **10**, 2952–2961.
- STEVENS, C. M., DANIEL, R., ILLING, N. & ERRINGTON, J. (1992). Characterization of a sporulation gene, *spoIVA*, involved in spore coat morphogenesis in *Bacillus subtilis*. *Journal of Bacteriology* **174**, 586–594.
- TAO, J. S. & ISHIGURO, E. E. (1989). DNA sequence of the *murE* gene of *Escherichia coli*. *Canadian Journal of Microbiology* **35**, 1051–1054.