Nucleotide Sequence of the *Bacillus licheniformis* Homologue of the Sporulation Locus *spoIIA* of *Bacillus subtilis*

By M. D. YUDKIN,* L. APPLEBY AND A. J. SMITH†

Microbiology Unit, Department of Biochemistry, University of Oxford, South Parks Road,
Oxford OX1 3QU, UK

(Received 18 November 1988; revised 8 January 1989; accepted 19 January 1989)

The *spoIIA* locus of *Bacillus licheniformis* was cloned into the phage vector ϕ 105J9; selection was based on the ability of the clone to complement mutations in both the first and the last of the *spoIIA* genes of *B. subtilis*. The *B. licheniformis* DNA was subcloned into M13 and sequenced; it includes three genes of identical lengths to those of *B. subtilis*. The average interspecies difference in nucleotide sequence is 24%; C occurs at the third position of codons 55% more often in the *B. licheniformis* than in the *B. subtilis* sequence. The average difference in predicted amino acid sequence is 11%, but the distribution of these differences is far from random, and there are several long stretches of amino acid sequence that are identical in the two organisms. The distribution of non-conservative amino acid changes between the species is also strikingly non-random; no such changes are found in those regions in which missense mutations are known in *B. subtilis*. Each species has an open reading frame 5' of the *spoIIAA* gene, but the predicted amino acid sequence, and the distribution of differences between the two species in both nucleotides and predicted amino acids, suggest that these open reading frames are not expressed. Both species have regions 3' of the open reading frames which resemble rho-independent terminators of transcription, but that in *B. licheniformis* is longer and could form a more elaborate secondary structure than that in *B. subtilis*.

INTRODUCTION

The *spoIIA* region of *Bacillus subtilis* includes many features of great interest. Mutations in the locus give rise to a variety of phenotypes, but these are not correlated with the genetic positions of the lesions (Errington & Mandelstam, 1983). Cloning and sequencing showed that the locus includes three genes (Fort & Piggot, 1984), but one of these, *spoIIAB*, contains no known mutations. Inspection of the sequence suggests that the first two genes (*spoIIAA* and *spoIIAB*) are translationally coupled, but that a third (*spoIIAC*) is translated independently. An open reading frame occurs to the 5' side of *spoIIAA*, but it is not known whether this is translated. The predicted amino acid sequence of the *spoIIAC* product shows that this protein is homologous to the family of bacterial sigma factors (Helmann & Chamberlin, 1988), with a helix–turn–helix region (Pabo & Sauer, 1984) that possesses an unusual specificity (Yudkin, 1987*a*). Curiously, the various different phenotypes due to *spoIIAC* mutations are exactly similar to those produced by mutations in *spoIIAA*, whose predicted product in no way resembles a sigma factor.

Recent work in this laboratory has included a study of the structure and function of the *spoIIA* proteins, based largely on mutation and reversion (Yudkin *et al.*, 1985; Yudkin, 1987*b*). The results have enabled us to focus attention on those regions of the *spoIIAA* and *spoIIAC* proteins where mutations give rise to changes in sporulation phenotype, and to propose functions for some of them. As a complement to such mutational studies, we wish to establish which regions of the proteins can be allowed to vary without impairing function. One way to achieve this end

† Present address: Perkin Elmer Ltd, Post Office Lane, Beaconsfield, Bucks. HP9 1QA, UK.

would be to direct mutations systematically into each part of the locus, and determine which mutations were silent. An alternative is to study the homologue of *spoIIA* from a related *Bacillus*, in order to identify regions in which differences that do not affect function have occurred in the evolution of the two species. This is the approach that we have adopted in the work reported here, in which we have compared the sequence of the *spoIIA* region of *Bacillus licheniformis* with that of *B. subtilis*.

At the same time we have determined whether the particular features of the genetic organization of *spoIIA* mentioned above are conserved between the two bacilli. We find that the apparent translational coupling between *spoIIAA* and *spoIIAB*, and independent translation of *spoIIAC*, occur in both organisms. Furthermore, we find in *B. licheniformis* (like in *B. subtilis*) an open reading frame upstream of *spoIIAA*, but we give reasons for believing that this is not translated in either of the species.

METHODS

Bacterial strains. *B. subtilis* 69.6 was constructed by transforming MB75 (*metC3 lys-1 spo+*) with DNA from strain 69 (strain NG 18.6 of Piggot, 1973), selecting for Met⁺, and obtaining a *spo* recombinant by congression. *B. licheniformis* strain 6346 was obtained from the NCTC.

Preparation of DNA. This was by the method of Marmur (1961).

Restriction endonuclease digestion. Enzymes were bought from Amersham, Boehringer or BRL and used as recommended by the suppliers.

Ligation. This was as described by Yudkin (1987b).

Other cloning techniques. The procedures of Maniatis *et al.* (1982) were followed.

DNA sequencing. This was as described by Yudkin (1987b).

RESULTS AND DISCUSSION

Cloning and sequencing of the B. licheniformis spoIIA region

To clone the *spoIIA* region, we made use of the cloning vehicle ϕ 105J9 (Errington, 1984). *B. licheniformis* DNA was restricted to completion with *Bcl*I, and about 200 ng of digest was ligated to about 500 ng of ϕ 105J9 DNA that had been restricted with *Bam*HI, and transfected into *B. subtilis* protoplasts as described by Errington (1984). The pool of recombinant phage was used to lysogenize strain 69.6. Spo⁺ colonies were selected by the method of Hoch (1971), and those obtained (about 50) were pooled, grown in broth and induced to lyse by treatment with mitomycin C (0.5 μ g ml⁻¹). Cell debris was removed by centrifugation, and the lysate was sterilized by filtration and spotted on to a lawn of strain 69.6. A Spo⁺ lysogen was purified and used to prepare a stock of recombinant phage, which was named ϕ 105AJS1. This phage proved to be capable of complementing both *spoIIA69* and *spoIIA1*; since one of these mutations is in *spoIIAA* and the other in *spoIIAC* it appeared likely that ϕ 105AJS1 carried the whole *spoIIA* region of *B. licheniformis*.

After purification of the phage on a CsCl gradient, DNA was obtained by phenol extraction. Restriction mapping showed that it contained an insert of about 3.0 kb; this was trimmed with *Hind*III and *Eco*RV (which removed fragments of DNA that had been shown not to hybridize with a probe of *B. subtilis spoIIA* DNA). The resulting 2.1 kb of DNA was cloned into pUC13, from which fragments of the insert were subcloned for sequencing in M13.

A diagram of the sequencing procedure is shown in Fig. 1, and the sequence, all of which was determined on both strands, is given in Fig. 2 and Fig. 5.

The coding region

The coding strand of the *B. licheniformis spoIIA* region has 23.5% T, 22.4% C, 29.7% A and 24.4% G. The figures for the same region in *B. subtilis* are 25.3% T, 18.9% C, 30.8% A and 25.0% G. The sequence has three open reading frames that correspond exactly to the three *spoIIA* genes of *B. subtilis*. The lengths of the predicted polypeptide products – 117, 146 and 255 amino acid residues – are identical to those of *B. subtilis*. Nucleotide differences between the two species amount to 25%, 21% and 24% respectively for the three genes. The presumed ribosome-binding

Fig. 1. Partial restriction map of the *spoIIA* region of *B. licheniformis*. The arrows indicate the start-point, direction and extent of sequence determined from each subclone. Restriction sites are abbreviated as follows: A, *AluI*; C, *Clal*; D, *DraI*; H, *HindIII*; M, *MboI*; V, *EcoRV*; X, *XhoI*. The open rectangles at the bottom of the figure show the extent of the three *spoIIA* genes.

sites of both the *spoIIAA* and the *spoIIAC* genes have one additional nucleotide in *B. licheniformis*. The junction between *spoIIAA* and *spoIIAB*, the sequence of which strongly suggests translational coupling between these two genes, is identical in the two species.

The predicted polypeptides

Amino acid differences between the species amount to 16%, 11% and 9% respectively for the predicted products of the three genes; the conservation of *spoIIAB*, in which no mutations are known, is particularly interesting. The commonest amino acid difference at corresponding positions in the two species is between glutamine and lysine, which, we suggest, should be reckoned a conservative change in this system. (There are six such changes, as compared with only two changes between valine and isoleucine.) In each predicted gene product, the distribution of changes in amino acid residue is strikingly non-random. There are five stretches of sequence in which more than 30 consecutive residues are completely conserved, including, in *spoIIAC*, three stretches of more than 40 consecutive residues. Equally striking is the distribution of non-conservative changes. In *spoIIAA* and *spoIIAC*, all the non-conservative changes are clustered in those regions in which no missense mutations are known (Fig. 3), suggesting that the corresponding sections of the proteins can vary to some extent without impairment of their function. This feature is particularly striking for the sigma-like predicted product of the *spoIIAC* gene; as Fig. 3(c) shows, the non-conservative differences between the two *Bacillus* species all fall outside those regions of the protein in which the homologies within the family of sigma factors are especially noticeable (Helmann & Chamberlin, 1988). In the presumed helix-turn-helix region the only difference between the two species is a glutamate-aspartate change at position 6 of the motif; the glutamine at position 12 and the very unusual valine at position 13, which are thought to be involved in determining specificity of binding to DNA (Yudkin, 1987*a, b*), are conserved.

Despite the strong conservation of amino acid sequence, the distribution of codons used in *spoIIA* differs markedly for the two species (Table 1). The relatively high content of C in the coding strand of *B. licheniformis* (see above) is almost entirely accounted for by changes at the third position of codons: in *B. subtilis* 100 (19.3%) of the codons have C at the third position, whereas the corresponding figure for *B. licheniformis* is 155 (29.9%). (This large difference between the two organisms is not found for all genes, nor even for all *spo* genes. The recently published sequences for *spo0H* (Dubnau *et al.*, 1988) show that the coding strand of this gene has 17.7% C in *B. subtilis* and 20.3% in *B. licheniformis*; in the former species 21.6% of *spo0H* codons

TCCAGGCCCTGTCAAAAAAGGCACAGAGCTTGGGACGCTCGTTTTGAAAAAGGATGGA
 ATT TG TCC ATTC CA C T T C
 AAGGTGCTCGCTGAAAGTCTCTTGTGCTGAAAAAGATATGGACAAAGCGGGGATGTGG
 G A A G CT AA G A G C T TATC
 ACGATGTTCAAGCGGACGATGACCCACTGGACGAAGTGGAGTGAATAATGCCGAACGGTC
 AT C A GGAG A A TT AGT T T AC
 ACTAGTTTTGTACGGTGAAGGAATTTATAAAGTCTGAAGCGAAACACTCATTATCCGAT
 C TCC GA T
 M N
 TTAACCAAGGAGGAATGAGGATGAGCCTCGGAATCGATATTACGTCAAAGAATCCGTA 300
 CAT T - C T T C GA T T G
 L C I R L T G E L D H H T A E T L K Q K
 TTATGCATTTCGGTTGACAGGTGAACTCGATCACCATACAGCTGAAACCTTGAGAAACAA
 C T T A A C G C AGC A
 T Q S K D L E
 V S G H L E Q T D I R H I V M N L A D L
 GTCAGTGGCCATCTGGAACAAACCGACATTCGCCATATTGTGTCATGAATCTTGAGAACCTT
 G C CAATCAT GA GGAT T C GC CT G AG
 I Q
 S F M D S S G L G V V L G R Y K E I K Q
 TCGTTTATGGACAGTTTCAGGGCTTGGCGTCTGCTCGGCAGATATAAGGAGATTAAAGCA
 C C G TA TT A A C G A A
 I V
 L G G E M I V C A I S P A V K R L F D M
 CTCGGCGGAGAAATGATCGTCTGCGCCATTTCCCTGCTGTGTCAAACGTTTATTGATATG
 A T G T T T C T G G G A G
 F Q L
 S G L F K I I R L E Q S E Q R A L E T L
 TCCGGGTGTTTAAATCATTCGCCTTGAGCAATCAGAGCAGCGTGCCTTGAAACGTTG 600
 G T T C AT A T A AG CCTG AC
 H L E S
 G V A S M* K N E M N I Q F T A L S Q N E S
 GGGGTGGCGTCATGAAAAATGAATGAACATTCAGTTTACAGCGCTCAGCCAAAATGAAT
 A C C G T T C T G
 S
 F A R V T V A A F I A Q L D P T M D E L
 CGTTTGCACGGGTGACAGTCGCTGCTTTTATCGCTCAGCTTGACCCGACGATGGATGAAC
 C C T T T A A G A
 T E I K T V V S E A V T N A I I H G Y E
 TGACCGAAATTAACCGTTCGATCCGAAGCGGTACAAACGCGATCATCGGTTATG
 T C A G A G T G T T C T A
 E N C E K V M
 N S G Q G N V Y I S V T L E D H I V Y L
 AAAACTCAGGGCAGGGAAACGTATATATTCCGTCACCTCTCGAGGACCATATTGTCTATT
 G GAACT TG A G A T C A G G G A T G C A A
 L T
 T I R D E G V G I P D L E E A R Q P L F 900
 TAACGATCCGCGACGAAGGAGTCGGCATCCCTGATCTTGAAGAAGCGCGCCAGCCCTGT
 G T T T T CT A TA A C T T A
 T T K P E L E R S G M G F T I M E N F M
 TCACGACAAAGCCTGAACCTCGAGCGGTGCGGAATGGGCTTTACGATCATGGAATAATTTCA
 T T G T C T C T
 V R
 D D I S I D S S P E M G T T I H L T K H
 TGGATGATATTTCGATCGACTCCTCAGTGGGAACCAATACACTTAACAAAGC
 G CAGT T A G A A G T G
 L S K S K A L C N * M D V E V K
 ACTTATCAAAAAGCAAAGCGCTTTCGAATTAAGGGAGATTGTGTTATGGATGTGGAGGTTA
 T -

```

      N G K   A
      K E N Q N T Q L K D H E V K E L I K N S
AAAAAGAAAACCAGAACACTCAGCTTAAAGACCATGAAGTGAAAGAACTGATTAAAAACA
      G   A CGG A A   G   G G T   A G   T A C   C A

      Q
      Q N G D Q K A R D L L I E K N M R L V W
GCCAGAACGGCGATCAAAAAGCAAGGGACCTCCTCATAGAAAAAATCATGCGTCTTGTGT 1200
      A T   C GC G   A

      S V V Q R F L N R G Y E P D D L F Q I G
GGTCTGTCGTTTCAGCGTTTTTTGAACAGAGGCTATGAGCCTGACGACCTCTTTCAAATCG
      A   G   A   A   T   C G

      T
      C I G L L K S V D K F D L S Y D V R F S
CTGTCATCGGCTCTTGAAGTCGGTGGACAAATTCGATCTTTCCTATGACGTTTCGGTTTT
      G G A A T T   T T A A   T G T

      T Y A V P M I I G E I Q R F I R D D G T
CCACCTACGCGGTTCCGATGATTATCGGCGAGATTCAGCGGTTTATCAGAGATGACGGAA
      A G T A G   A A C A A   C T

      D
      V K V S R S L K E L G N K I R R A K E E E
CCGTCAAAGTGAGCCGCTCGCTGAAAGAACTCGGCAACAAAATCCGGCGGGCGAAAGAGG
      A G ATCA G AT A   G T A   C   G T

      T L   V   H
      L S K S N G R I P T V Q E I A D Y L E I
AGCTTTCCAAGTCAAAACGGCCGGATTCCGACCGTTTCAGGAAATCGCCGATATCTCGAAA 1500
      G AA CTG   A AG G   G G   G   T CC T G G

      A
      S S E E V V M A Q E A V R S P S S I H E
TCAGTTCAGAAGAGGTCTGTGAGGCCAGGAAGCGGTCCGCTCTCCCTCGTCCATTATC
      TGAAG T G T T AC   A G   AA GG   A T G   C

      N
      T V Y E N D G D P I T L L D Q I A D Q S
AAACGGTTTACGAAAACGACGGGGACCCGATTACACTTCTGGACCAAATCGCCGATCAGT
      C   T   T   A T   C G T T   T C A C

      S   E
      E E K W F D K I A L K E A I K D L D E R
CGGAAGAGAAGTGTTTCGATAAAATCGCTTGAAGGAAGCCATCAAGGATTGATGAAC
      A   A A   T C   T GC   A   G   GC   G A

      Q
      E K L I V Y L R Y Y K D K T Q S E V A D
GGGAAAAGCTGATTGTTTATTTAAGGTATTACAAAGACAAGACGCAATCTGAAGTTGCCG
      A A C C   C C A   T   C   A G C G G T

      R L G I S Q V Q V S R L E K K I L K Q I
ACCGCTCGGCATATCCCGAGGTCAAGTCTCGAGGCTTGAGAAAAAGATTTAAAGCAA 1800
      G G   G C T   G T C   A G A A   A G

      V   T D G
      K N Q M D H F E S *
TCAAAAACCATGATGGACCATTTTGAAAGCTGA
      GGTT A   T   ACG TG AG

```

Fig. 2. The sequence of part of the *spoIIA* region (see also Fig. 5). The continuous sequence of nucleotides (60 to a line) is that of *B. licheniformis*. Below is given the sequence of *B. subtilis* where it differs from that of *B. licheniformis*, the first nucleotide shown being that numbered 155 in the sequence of Fort & Piggot (1984). (Hyphens mark the positions of two gaps that have been introduced into the *B. subtilis* sequence in order to maximize the homology.) Immediately above the nucleotides are shown (in one-letter code) the sequences of the predicted products of the *spoIIA*, *spoIIAB* and *spoIIAC* genes of *B. licheniformis*; and on the top line are shown the corresponding amino acid sequences for *B. subtilis* at positions where they differ from those of *B. licheniformis*. Asterisks mark the positions of the ends of reading frames. The *Cla*I site is at nucleotide 274; the *Xho*I sites are at nucleotides 819, 918 and 1769.

Fig. 3. Positions of non-conservative amino acid differences between the two *Bacillus* species and of missense mutations in *B. subtilis*. (a) *spoIIA*; (b) *spoIIB*; (c) *spoIIAC*. The genes are shown (to scale) as open boxes, and the positions of non-conservative amino acid differences between the two species are marked as vertical lines at the appropriate places within each box. At each such position the *B. licheniformis* residue is shown (in one-letter code) above the box, the *B. subtilis* residue below the box. Non-conservative amino acid differences are defined as those that cross the boundaries between the following groups: 1, alanine, serine, threonine; 2, aspartic acid, glutamic acid, asparagine, glutamine; 3, isoleucine, leucine, methionine, valine; 4, arginine, glutamine, lysine. The vertical lines above the boxes mark the positions of missense mutations, taken from the work of Yudkin *et al.* (1985), Yudkin (1987*b*) and M. D. Yudkin (unpublished). The shaded areas in (c) represent the homologous regions common to bacterial sigma factors (Helmann & Chamberlin, 1988).

have C at the third position and in the latter species 25.1%.) There is one striking difference in the distribution of codons between the compilation of 6121 *B. subtilis* codons made by Piggot & Hoch (1985) and those used by the same organism in *spoIIA*, namely that the glycine codon GGU is remarkably underrepresented here.

There are 11 examples of the use of synonymous codons that differ in two nucleotides, eight for leucine and three for arginine. In addition, there are two positions at which codons occur that specify serine, but in which all three nucleotides differ between the two species. It is not obvious how this situation has arisen, since it is impossible for one of these serine-specifying codons to change by single-base-pair substitution to the other without passing through codons that specify other amino acid(s). We speculate that the common ancestor of the two *Bacillus* species had a different residue at the position now occupied by serine. For example, serine-110 of the *spoIIAC* protein is specified by AGC in *B. licheniformis* and by TCA in *B. subtilis*. Possibly an ancestor had threonine at this position, specified by (say) ACC; this codon could then mutate to ACA (still threonine) and thence to TCA (serine), or in another line of descent directly to the serine codon AGC. The presence of these alternative serine codons strongly suggests that there has been selective pressure in favour of a serine residue at this position (and at position 121 of *spoIIB*); it is perhaps pertinent that alteration of the adjacent residue (arginine-111 of the *spoIIAC* protein) results in a mutant in which sporulation is almost abolished.

The 5' region

An open reading frame runs from the beginning of the sequence that we have determined (see Fig. 2), capable of encoding a 55-residue polypeptide. The corresponding region of the *B. subtilis* sequence could specify a polypeptide that would terminate one residue earlier. We do not know whether these reading frames are expressed, but we think it unlikely. In the first place, their amino acid sequences (Fig. 4) are extremely unusual. Secondly, the distribution of nucleotide differences between the two species is quite unlike that within the other three reading frames

Table 1. Codon usage in *spoIIA*

The number of occurrences of each codon is given first, in normal type, for *B. subtilis*, and then, in italics, for *B. licheniformis*.

Phe UUU 13 12	Ser UCU 9 3	Tyr UAU 9 9	Cys UGU 3 0
Phe UUC 3 4	Ser UCC 3 11	Tyr UAC 2 3	Cys UGC 2 4
Leu UUA 11 7	Ser UCA 11 8	Ter UAA 1 1	Ter UGA 1 2
Leu UUG 5 7	Ser UCG 6 9	Ter UAG 1 0	Trp UGG 2 2
Leu CUU 13 14	Pro CCU 5 5	His CAU 7 8	Arg CGU 6 4
Leu CUC 8 14	Pro CCC 0 2	His CAC 4 4	Arg CGC 4 7
Leu CUA 2 0	Pro CCA 1 0	Gln CAA 11 10	Arg CGA 4 0
Leu CUG 14 8	Pro CCG 4 4	Gln CAG 16 16	Arg CGG 4 9
Ile AUU 18 21	Thr ACU 4 2	Asn AAU 8 5	Ser AGU 2 3
Ile AUC 19 20	Thr ACC 6 7	Asn AAC 8 14	Ser AGC 5 6
Ile AUA 3 3	Thr ACA 10 9	Lys AAA 24 24	Arg AGA 5 4
Met AUG 17 17	Thr ACG 8 8	Lys AAG 14 13	Arg AGG 3 3
Val GUU 12 9	Ala GCU 11 6	Asp GAU 24 16	Gly GGU 1 2
Val GUC 7 17	Ala GCC 4 7	Asp GAC 14 20	Gly GGC 11 14
Val GUA 7 3	Ala GCA 4 4	Glu GAA 31 34	Gly GGA 11 7
Val GUG 15 9	Ala GCG 7 8	Glu GAG 18 14	Gly GGG 5 5

SRPLSKKGTELGLTVLKKDGVLAESPLLAEKDMDKAGMWTMLKRTMTHTWTKWKE*
ISAPIQ Q E VA KE K FI F GD F *

Fig. 4. The predicted amino acid sequences of the open reading frames 5' of *spoIIAA*. The continuous sequence is that of *B. licheniformis*. Residues are shown for *B. subtilis* only where they differ from those for *B. licheniformis*. Asterisks mark the positions of the ends of the reading frames.

(shown in Fig. 2). In *spoIIAA*, *spoIIAB* and *spoIIAC*, 75% of the nucleotide differences that occur between the two species are found at the third position of codons; but in the region 5' of *spoIIAA* only 48% are at the third position. As a result, 35% of the amino acid residues that would be encoded by the reading frames differ between the species, as against only 11% in the rest of the region. Finally, many of these amino acid differences are strikingly non-conservative (see Fig. 4). We think that the sequence 5' of *spoIIAA* may form an open reading frame only by chance, although we note that in *B. subtilis* the entire upstream region sequenced by Fort & Piggot (1984) represents an open reading frame. We presume that in both species the sequence

A few dozen nucleotides further on, we have sequenced what is clearly the beginning of the *B. licheniformis spoVA* region; we shall report on this later.

REFERENCES

- DUBNAU, E., WEIR, J., NAIR, G., CARTER, L., MORAN, C. & SMITH, I. (1988). *Bacillus* sporulation gene *spo0H* codes for sigma³⁰ (sigma^H). *Journal of Bacteriology* **170**, 1054–1062.
- ERRINGTON, J. (1984). Efficient *Bacillus subtilis* cloning system using bacteriophage vector ϕ 105J9. *Journal of General Microbiology* **130**, 2615–2628.
- ERRINGTON, J. & MANDELSTAM, J. (1983). Variety of sporulation phenotypes resulting from mutations in a single regulatory locus, *spoIIA*, in *Bacillus subtilis*. *Journal of General Microbiology* **129**, 2091–2101.
- FORT, P. & PIGGOT, P. J. (1984). Nucleotide sequence of sporulation locus *spoIIA* in *Bacillus subtilis*. *Journal of General Microbiology* **130**, 2147–2153.
- HELMANN, J. D. & CHAMBERLIN, M. J. (1988). Structure and function of bacterial sigma factors. *Annual Review of Biochemistry* **57**, 839–872.
- HOCH, J. A. (1971). Selection of cells transformed to prototrophy for sporulation markers. *Journal of Bacteriology* **105**, 1200–1201.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MARMUR, J. (1961). A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *Journal of Molecular Biology* **3**, 208–218.
- MORAN, C. P., LANG, N., LEGRICE, S. F. J., LEE, G., STEPHENS, M., SONENSHEIN, A. L., PERO, J. & LOSICK, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Molecular and General Genetics* **186**, 339–346.
- PABO, C. O. & SAUER, R. T. (1984). Protein–DNA recognition. *Annual Review of Biochemistry* **53**, 293–321.
- PIGGOT, P. J. (1973). Mapping of asporogenous mutations of *Bacillus subtilis*: a minimum estimate of the number of sporulation operons. *Journal of Bacteriology* **114**, 1241–1253.
- PIGGOT, P. J. & HOCH, J. A. (1985). Revised genetic linkage map of *Bacillus subtilis*. *Microbiological Reviews* **49**, 158–179.
- TINOCO, I., BORER, P. N., DENGLER, B., LEVINE, M., UHLENBECK, O. C., CROTHERS, D. M. & GRALLA, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology* **246**, 40–41.
- YUDKIN, M. D. (1987*a*). The prediction of helix–turn–helix DNA-binding regions in proteins. *Protein Engineering* **1**, 371–372.
- YUDKIN, M. D. (1987*b*). Structure and function in a *Bacillus subtilis* sporulation-specific sigma factor: molecular nature of mutations in *spoIIAC*. *Journal of General Microbiology* **133**, 475–481.
- YUDKIN, M. D., JARVIS, K. A., RAVEN, S. E. & FORT, P. (1985). Effects of transition mutations in the regulatory locus *spoIIA* on the incidence of sporulation in *Bacillus subtilis*. *Journal of General Microbiology* **131**, 959–962.