Volume 12 Number 4 1984 Nucleic Acids Research

The complete DNA sequence and regulatory regions of the *Bacillus licheniformis spoOH* gene

N.Ramakrishna, E.Dubnau and I.Smith

Department of Microbiology, Public Health Research Institute of the City of New York, Inc., 455 First Avenue, New York, NY 10016, USA

Received 9 January 1983; Accepted 20 January 1984

ABSTRACT

We have determined the sequence of a 1228 base-pair cloned DNA fragment from Bacillus licheniformis capable of specifically complementing mutations in the spoOH gene, which is required for the early stage of sporulation in B. subtilis. The sequence has only one long open reading frame consisting of 168 codons. In vivo and in vitro transcription mapping studies indicate the size of complementary RNA to be around 1 kb with the 5' initiation site at base 79 and the 3' termination site in the area of base 1138. This indicates the presence of a 5' untranslated RNA and a fairly long 3' extension. The promoter sequence of this gene is 5'TATAAT3' at -10, and 5'TTGACG3' at -35, a typical E. coli-like promoter sequence, and is transcribed in vitro specifically only by RNA polymerase containing σ55 and not σ37-containing holoenzyme.

INTRODUCTION

We have previously isolated a 1.2 kb fragment of the B. licheniformis chromosome which specifically complements mutations in spoOH, an early sporulation gene of B. subtilis (1). RNA complementary to the cloned fragment is transcribed during vegetative growth and sporulation. In minicells a protein with an apparent molecular weight of 27K is coded for by the 1.2 kb DNA segment. Deletions within the insert abolish the spoOH$^+$ complementing activity of the clone and also eliminate the 27K protein. The insert is functional in both orientations which suggests that it possesses its normal promoter. We have recently shown that the 1.2 kb cloned DNA fragment shows homology to the B. subtilis spoOH gene and must therefore contain the B. licheniformis counterpart (ms. submitted).

In this communication we report on in vivo and in vitro transcription of the 1.2 kb DNA fragment and present the complete 1228 bp DNA sequence. These studies show that the B. licheniformis spoOH gene is transcribed by the major vegetative form (σ55-containing) of the B. subtilis RNA polymerase and by E. coli RNA polymerase holoenzyme and that the promoter structure at the -35 and -10 sites is of the canonical E. coli type (TTGACG and TATAAT). An open

reading frame of 168 amino acids corresponding to a protein with a molecular weight of 22,000 is found in the 1.2 kb fragment. It is preceded by a ribosomal binding sequence of AGGAAGG, which is followed by a UUG start codon.

MATERIALS AND METHODS

Bacterial plasmids and strains

Plasmids pIS1 and pIS9A each contain the 1.2 kb B. licheniformis chromosomal fragment in plasmid pBD64 and pBD97, respectively, as described previously (1). pIS52B was created by restricting pIS9A at its unique EcoRI site and ligating it with EcoRI restricted pSK10Δ6 (created and kindly provided by Jan Pero), a derivative of pMC1403 (2). The chimeric plasmid, pIS52B, can replicate in E. coli and in B. subtilis, expressing ampicillin resistance in the former and chloramphenicol resistance in the latter (manuscript in preparation).

B. licheniformis strain FD02 was previously described (1). E. coli strain JM105 was used to propagate pIS52B and was used as a source of RNA for in vivo transcription mapping.

Techniques for the isolation of plasmids, growth of cells, and transformation were as previously described (1,3).

Northern RNA:DNA hybridization

RNA was isolated as previously described (1), and DNA was nick-translated as described by Rigby et al. (4). RNA was separated by electrophoresis on agarose gels. Blotting of the gels, preparation of diazobenzylomethyl (DBM) paper, and hybridization with nick-translated DNA were performed according to the methods of Alwine et al. (5).

In vivo transcription mapping

RNA was isolated, and DNA fragments were 5' end-labeled with [γ-^{32}P]ATP previously described, as were the techniques for RNA:DNA hybridization followed by S1 nuclease digestion and subsequent electrophoresis (1).

DNA sequence analysis

DNA fragments were labeled at the 5' ends with T4 polynucleotide kinase (Bethesda Research Laboratories) and [γ-^{32}P]ATP (Amersham). The DNA sequence was determined according to the procedures of Maxam and Gilbert (6). The products of the G, G+A, T+C, and C cleavage reactions were electrophoresed on 0.4 mm 8% or 20% polyacrylamide gels, which were 40 or 60 cm long.

In vitro run-off RNA synthesis

E. coli RNA polymerase (1 μg), kindly provided by J. Krakow, was prebound (10' at 37°C) to 2 μg plasmid DNA template restricted at sites as

indicated in 50 µl reaction mixtures lacking ribonucleotides but containing 10% (vol/vol) glycerol. RNA synthesis was initiated by the addition of 0.15 mM CTP, GTP, and ATP, and 0.4 µM (10 µCi) [α-^{32}P]UTP. After 5 min incubation at 37°C, unlabeled UTP (0.15 mM) was added. The reactions were then stopped after 5 min and RNA was isolated as described by Pero and coworkers (7, 8).

The B. subtilis RNA polymerase assays (performed by Charles Moran and Richard Losick) were carried out as previously described (9).

RESULTS

The cloned 1.2 kb B. licheniformis fragment which complemented spoOH mutation in B. subtilis was functional in both orientations and was therefore believed to possess its normal promoter (1). We have recently shown that the 1.2 kb fragment shows extensive DNA:DNA homology with the B. subtilis spoOH gene and must be the homolog in B. licheniformis (ms. in preparation). Knowledge of early spo gene structure, which of the multiple RNA polymerase holoenzymes transcribes these genes, and the nature of their gene products is of crucial importance in our understanding of the control of sporulation. We therefore proceeded to analyze in vivo and in vitro transcription of the B. licheniformis spoOH gene and to analyze its DNA sequence.

In vivo transcription of the spoOH gene

Total cellular RNA was prepared from wild type B. licheniformis cells which had been induced to sporulate by resuspending in minimal sporulation replacement medium (10). The RNA was electrophoretically fractionated on agarose gels and was probed with ^{32}P nick-translated pIS1 DNA, by means of the "Northern" technique (5). A single RNA-DNA hybrid corresponding to an RNA transcript of approximately 1.0 kb was observed (Fig. 1).

We then used a modified Berk Sharp transcriptional mapping method (11) to study further the size and orientation of the spoOH transcript. We have previously shown, by means of this method, that significant levels of RNA complementary to the spoOH gene were transcribed in vegetative as well as sporulating B. licheniformis (1). For the experiments reported here, RNA prepared from exponentially growing B. licheniformis was hybridized with 5' end labeled restricted pIS1 DNA. As shown in Fig. 2, there are two HindIII sites in the middle of the B. licheniformis insert. Removal of the internal 350 base HindIII fragment destroys the ability of the clone to complement the spoOH mutation and defines some or all of this sequence as part of the spoOH gene. Therefore, pIS1 DNA was restricted with HindIII and the 5' termini were labeled with [^{32}P]ATP and T4 DNA kinase. The HindIII labeled fragments

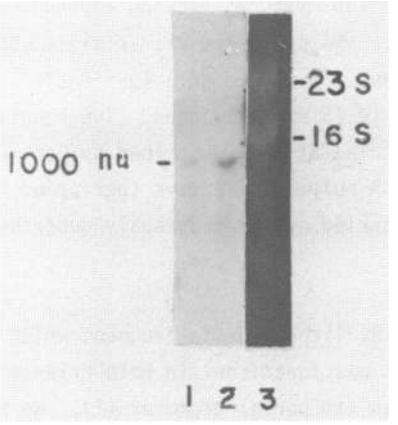

FIG. 1. Northern blotting of in vivo RNA. RNA from B. licheniformis strain FDO2 isolated at different times after shift to sporulation medium was denatured by treatment with glyoxal and electrophoresed on agarose gels (5). The RNA was blotted onto DBM paper and hybridized with ^{32}P-nick translated pIS1 DNA. Lanes 1 and 2 represent RNA isolated 30 and 60 min after transfer to sporulation medium. Lane 3 shows the position of denatured 23S and 16S RNA.

were secondarily cut with BglII, and the resulting three fragments were isolated on acrylamide gels and were used for the in vivo RNA transcription analysis. As shown in Fig. 3A, lane 1, the left-most BglII-HindIII fragment (see Fig. 2 for orientation) was reduced in size when hybridized to RNA and subsequently treated with S1 nuclease. We consistently observe

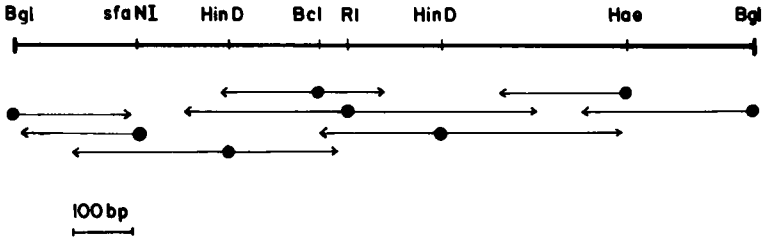

FIG. 2. Restriction sites and sequencing strategy of cloned DNA. The top line indicates the position of restriction sites used in sequencing experiments. The abbreviations denote the following enzymes: Bgl = BglII; HinD = HindIII; Bcl = BclI; Rl = EcoRI; Hae = HaeIII. Filled circles denote the site of 5' end-labeling, and the arrows show the direction and extent of sequencing in each experiment. All labeled fragments were secondarily cut by suitable restriction enzyme prior to sequencing. Regions of DNA which were not sequenced on both strands were sequenced several times.

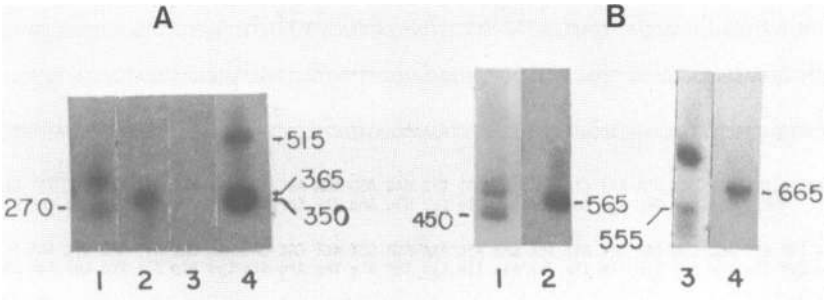

FIG. 3. S1 nuclease transcription mapping of cloned B. licheniformis spoOH gene. pIS9A DNA was 5' end-labeled at either the HindIII or EcoRI sites (Figs. 3A and B, respectively), secondarily cut with BglII, and fragments isolated and subjected to hybridization with RNA isolated from B. licheniformis FDO2. (A) With 5' labeled fragments, lanes 1 and 2 show the size of S1 resistant fragments using the leftward BglII-HindIII fragment (365 bp) and the middle HindIII-HindIII fragment (350 bp), respectively. Lane 3 shows the result of similar hybridization and S1 treatment experiments with the rightward HindIII-BglII fragment (see Fig. 2 for orientation). Lane 4 shows the mobilities of the untreated fragments. (B) Lanes 1 and 3 show the size of S1 resistant DNA hybrids, obtained using the left BglII-EcoRI fragment and the right EcoRI-BglII labeled at 5' (lane 1) and 3' (lane 3) ends, respectively, for the hybridization. Lanes 2 and 4 show the untreated left EcoRI-BglII (565 bp) and the right EcoRI-BglII (660 bp) fragments, respectively.

a second larger protected fragment when promoter proximal fragments are used for transcriptional mapping (see also Fig. 6), even when the concentration of S1 nuclease is increased or the annealing temperature is raised. This artifact may represent some kind of secondary structure which is resistant to S1 treatment. Full length protection was observed with the middle HindIII fragment (Fig. 3A, lane 2). The right-most HindIII-BglII fragment (Fig. 3A, lane 3) showed no hybridization, i.e., complete sensitivity to S1. These results indicate that the spoOH transcript is initiated within the left BglII-HindIII fragment and traverses the HindIII sites. The size of the protected fragment in Fig. 3A, lane 1, indicates that transcription is initiated at a point approximately 80 bases within the left BglII terminus of the insert.

A similar in vivo transcriptional mapping experiment, using RNA protection of the 3' labeled EcoRI site on the clone (Fig. 3B, lane 3) showed that the transcript terminated about 120 bases from the right BglII terminus. RNA protection of the 5' labeled EcoRI site resulted in a resistant hybrid fragment of approximately 450 nucleotides. This latter result indicated an initiation point about 100 nucleotides in from the left-most BglII site (Fig. 3B, lane 1). We conclude from these results that the spoOH transcript was

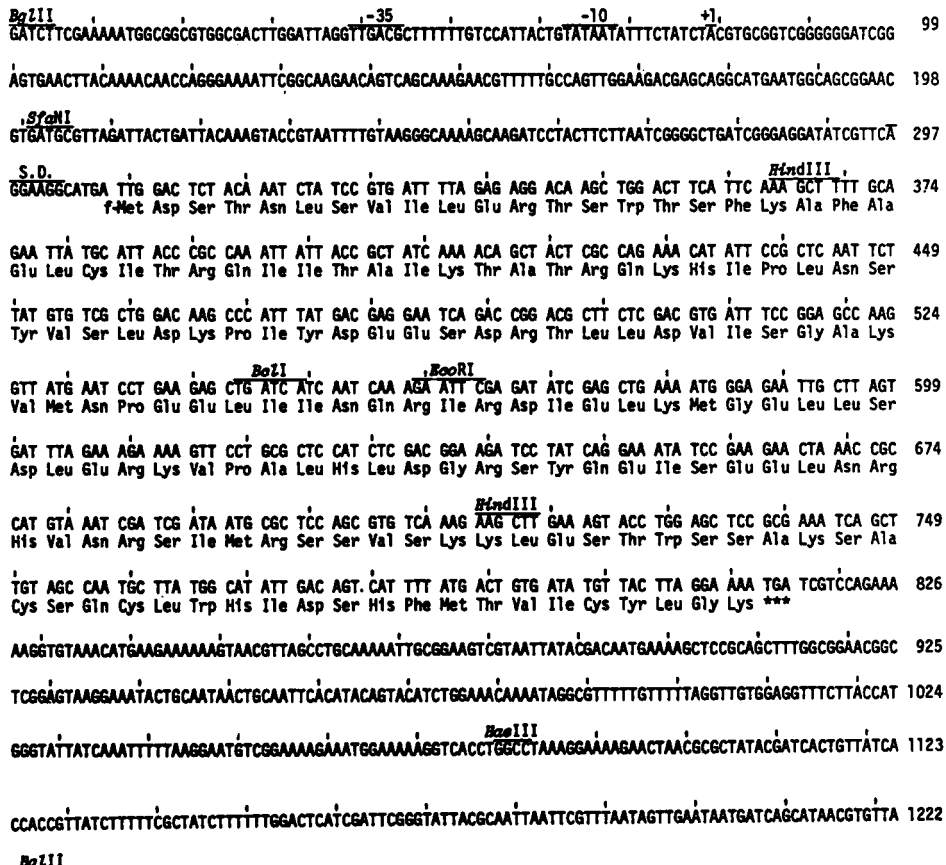

FIG. 4. DNA sequence of B. licheniformis spoOH gene. The coding strand of the nucleotide sequence is shown. The sequence was determined for the most part on both strands, and where it was determined on one strand, it was repeated at least three times. The underlined sequences shown as -35 and -10 refer to the conserved RNA polymerase binding and recognition sites (14). The underlined sequence referred to as "S.D." refers to the conserved ribosome binding site, and "TTG" refers to the initiator codon recently shown to be used in gram-positive organisms (13).

initiated at a point 80-100 nucleotides from the left-most Bgl terminus and terminated approximately 100 nucleotides from the right BglII terminus.

We then obtained the DNA sequence of the 1228 base insert by the Maxam-Gilbert technique (6) according to the strategy outlined in Fig. 2. The complete DNA sequence is illustrated in Fig. 4. We can summarize our conclusions as follows. An E. coli-like or B. subtilis σ55-like promoter structure

(12, 13) is observed at base 40 to 70 (relative to the left BglII site). The -35 sequence is TTGACG, and the -10 sequence is TATAAT. The presumptive transcription initiation site (+1) is near base 75. Following a ribosome binding sequence of AGGAAGG, which begins at position 300, a long open frame (168 amino acids) starts with a UUG codon at position 310. Rabinowitz and coworkers have shown that this codon can serve to initiate proteins in gram-positive bacteria (14). This indicates an untranslated RNA leader of approximately 230 bases. While we are not sure of the exact transcriptional termination site, there are long rows of T (U in the transcript) at position 1000 and at 1140. Either one of these would be consistent with the overall size of the transcript as determined by Berk-Sharp hybridization (about 1000 bases). Since the open reading frame ends at base 813, this means an untranslated 3' extension of 200 or more bases. There are no other long open frames in the 1228 base fragment, but there are several short peptides following adequate ribosome binding sites. For example, at position 1010 a good gram-positive ribosome binding site (14) of GGAGGT is followed by an ATG start codon at position 1024 and an open reading frame of 19 amino acids.

In vitro and in vivo transcriptional mapping of the spoOH promoter

The in vivo hybridization and DNA sequence data suggested that the spoOH gene should be transcribed by the major vegetative RNA polymerase (containing the σ55 subunit) not the holoenzyme with the σ37 specificity factor (12). Runoff transcription experiments with σ37 and σ55 holoenzymes were performed (Fig. 5A) and our predictions were confirmed, i.e., σ55 polymerase transcription of HindIII cleaved pIS9A was much more efficient than similar experiments with σ37 holoenzyme. As expected from the promoter sequences, E. coli RNA polymerase also transcribed the spoOH gene in vitro (Fig. 5B). However, there was a small discrepancy in the calculated RNA initiation site from these experiments. With the B. subtilis holoenzyme RNA was initiated at approximately 300 bases from the HindIII site, while the E. coli enzyme gave a corresponding value of 275. Since the B. subtilis and E. coli transcription runoff experiments were performed at different times, in different laboratories, it was not clear whether the small differences in the apparent transcriptional start sites were real or due to experimental variation. To answer this question, we performed in vivo transcription mapping experiments, using RNA prepared from wild type B. licheniformis and from an E. coli strain containing pIS52B, a plasmid possessing the B. licheniformis spoOH BglII-EcoRI promoter proximal fragment. RNA was also isolated from B. subtilis strains containing pIS52B and pIS1, which has the intact spoOH gene. These RNA preparations

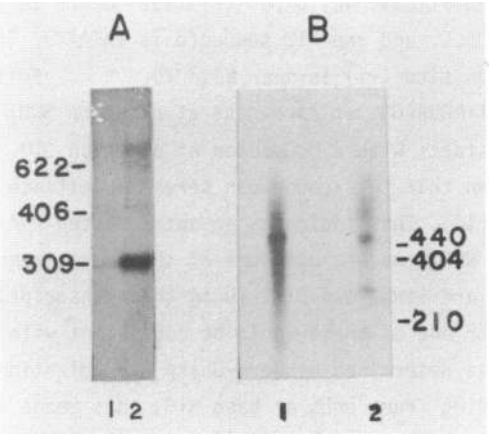

FIG. 5. Runoff transcription with B. licheniformis spoOH gene. In vitro synthesized RNA was labeled with [α-^{32}P]UTP with different templates and RNA polymerase from E. coli and B. subtilis. RNA was prepared after termination of the reaction and was electrophoresed on gels. (A) Pattern obtained by B. subtilis RNA polymerase with pIS9A DNA cleaved at HindIII using different forms of the enzyme. Lane 1 with σ37 holoenzyme and lane 2 with σ55 holoenzyme. (B) Transcription with E. coli RNA polymerase. Lane 1, BglII-EcoRI fragment. Lane 2, pIS9A cleaved at HindIII. The experiments with the B. subtilis RNA polymerases were performed by Charles Moran and Richard Losick.

were used to hybridize with the SfaNI restricted, 5' ^{32}P end-labeled promoter proximal spoOH fragment (secondarily restricted with BglII). The protected samples were electrophoresed on a DNA sequencing gel alongside the SfaNI-BglII fragment which had been subjected to the Maxam-Gilbert DNA sequencing procedure. As the autoradiograms of the S1-treated hybrids show (Fig. 6), DNA fragments protected by the B. licheniformis, the E. coli, and the B. subtilis RNAs are of the same size. Consequently, in vivo transcription must initiate at the same site or very close by in these organisms. By comparison with the DNA sequence ladder, the +1 position for RNA initiation is at base 79, the 'A' in the sequence CTAC.

In these experiments, the other protected fragment discussed above (Fig. 3) was again observed when RNA from B. licheniformis and B. subtilis containing the intact B. licheniformis spoOH gene was used to protect the SfaNI-BglII promoter proximal fragment. However, it was not observed with RNA from B. subtilis or E. coli containing the B. licheniformis spoOH promoter fused to the lacZ gene at the EcoRI site. This site is distal to the SfaNI site (Fig. 1). This observation suggests that this protection is not caused by a

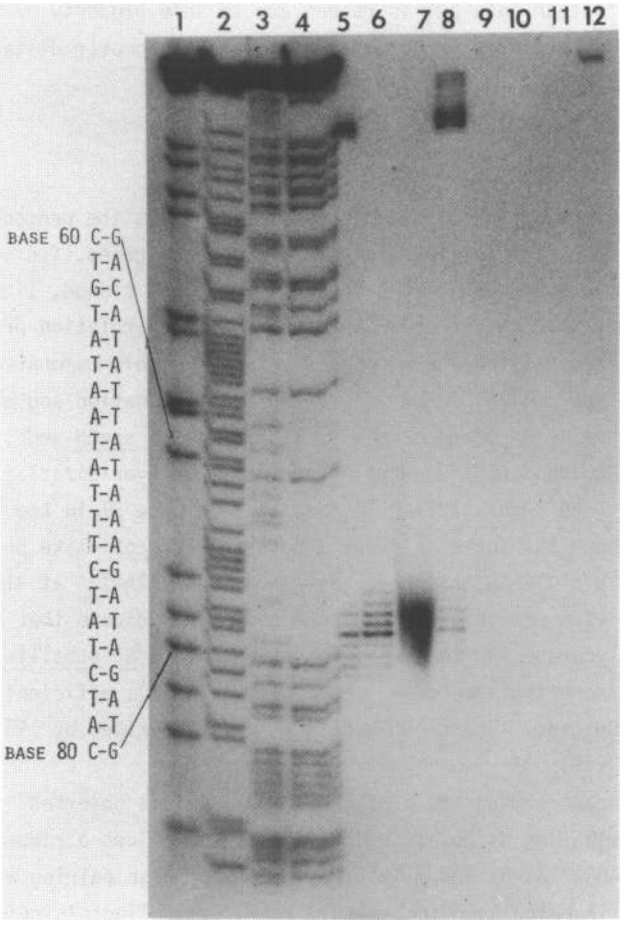

FIG. 6. High-resolution transcription mapping in E. coli, B. licheniformis, and B. subtilis. The SfaNI-BglII fragment of pIS9A (promoter proximal) 5' labeled at the SfaNI site was used for RNA:DNA hybridization followed by S1 treatment to map the transcription site. Lanes 1 through 4 represent the DNA sequencing reactions G, G+A, C+T, and C, respectively. Lane 5, protected hybrid with RNA from wild type B. licheniformis. Lane 6, protected hybrid with RNA from B. subtilis containing the plasmid pIS52B. Lane 7, protected hybrid with RNA from E. coli containing plasmid pIS52B. Lane 8, protected hybrid with RNA from B. subtilis containing plasmid pIS1. Lanes 9, 10, and 11 are controls for the specificity of the Berk-Sharp hybridization. Lane 9, RNA was used from B. subtilis containing the vector pBD64 which does not contain the B. licheniformis spoOH gene. Lane 10 shows RNA from E. coli lacking pIS52B, and lane 11, no RNA was used. Lane 12 represents the untreated SfaNI-BglII fragment used in the hybridization experiments.

The DNA base sequence adjacent to lane 1 shows the sequence of the probe 5' labeled at the SfaNI site and its complement, which is the sense strand. The numbering (base 60-base 80) refers to distance from the left (promoter proximal) BglII site (Fig. 4).

second RNA initiation site but is rather due to some property of the intact spoOH mRNA. This artifact is not present when the promoter distal part of the RNA is removed.

DISCUSSION

The initiation of sporulation in Bacilli requires the presence of several genes, the spoO loci. While several of these genes, i.e., spoOA (15), spoOB (16,17,18), spoOF (18), and spoOH (1) have been cloned, little is known about their gene products or their function in the sporulation process. The spoOF gene has recently been sequenced (19), and in this communication we have described the region of RNA initiation and termination and have determined the nucleotide sequence of the B. licheniformis spoOH gene.

In vivo RNA transcription mapping and Northern hybridization experiments have shown that RNA transcription is initiated at base 79 in the cloned insert. The spoOH DNA sequence shows a consensus E. coli-like promoter sequence, with a 5'TTGACG3' sequence at -35, and 5'TATAAT3' at -10. This agrees with in vivo and in vitro experiments which indicate that both the E. coli RNA polymerase and the major (σ55) vegetative B. subtilis RNA polymerase, which share the same promoter specificity (12), efficiently transcribe the spoOH gene. The spoOF gene is also transcribed by σ55 containing RNA polymerase (19).

One long open reading frame of 168 amino acids is observed, starting at bp 310 and terminating at bp 813. The open frame follows a ribosome binding site of 5'AGGAAGG3' at bp 300. The free energy of base pairing with the 3' end of 16S RNA is calculated to be -13.7 Kcal, using Tinoco's rule (20). The initiating codon is UUG, which is known to function very efficiently in gram-positive bacteria (14).

The calculated m.w. of the spoOH protein, 22,000 daltons, differs from the apparent m.w. of the gene product observed in minicells, 27,000 (1). We cannot presently explain this difference, although other proteins have shown discrepancies of this type. The most notable example is the E. coli σ factor. Its estimated m.w. from SDS gel analysis was calculated to be in excess of 90,000, whereas DNA sequence data give a m.w. of 70,000 (21).

Our sequence data indicate an untranslated RNA leader of approximately 230 bases and a 3' RNA extension of approximately 300. The function of these long tracts of untranslated RNA preceding and following the spoOH structural gene is presently unknown. However, there is a potential for secondary structure formation (hairpin loops) at bp 150 and 226 in the 5' leader region.

Likewise in the 3' RNA extension region there is a potential stem-loop structure at bp 985. The possibility that the untranslated RNA sequences and these potential secondary structures are involved in the regulation of spoOH expression is currently being studied in our laboratory.

Recent experiments from our laboratory using B. licheniformis spoOH-lacZ fusions have shown that the spoOH gene is growth regulated, i.e., shows greater gene expression at the end of vegetative growth and is repressed by glucose (ms. in preparation). This indicates that other factors must act in conjunction with the vegetative RNA polymerase, in vivo, to transcribe spoOH. The nature of these postulated regulatory factors and the environmental signals which control spoOH expression are currently under investigation.

The function of the spoOH gene product, the 22K protein, is also unknown. Comparison of its derived amino acid sequence with that of the B. subtilis spoOF gene product (19) shows no significant homology, even though codon usage in both these genes is grossly similar. We have been hampered in our studies of the spoOH gene product by the low in vivo levels of the 22K protein. Current studies in our laboratory, using immunological procedures, are enabling us to isolate large quantities of the protein and to study its function in sporulation.

ACKNOWLEDGMENTS

The authors would like to thank Joseph Krakow for his gift of E. coli RNA polymerase, and Jan Pero for her plasmid pSK10Δ6 and helpful advice on its use. We are very grateful to Charles Moran and Richard Losick for performing the in vitro transcription assays with the B. subtilis RNA polymerases and for valuable discussions. Special thanks are given to Carol Lusty and Hiroshi Nyunoya for their help with some aspects of the Maxam-Gilbert procedure. We would also like to thank David Dubnau and Leonard Mindich for helpful discussions, and we are especially grateful to Annabel Howard for secretarial assistance and Kettly Cabane for technical help. This work was supported by the National Institutes of Health, research grant GM-19693.

REFERENCES

1. Dubnau, E., Ramakrishna, N., Cabane, K. and Smith, I. (1981) J. Bacteriol. 147, 622-632
2. Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) J. Bacteriol. 143, 971-980
3. Gryczan, T.J., Contente, S. and Dubnau, D. (1978) J. Bacteriol. 134, 318-329
4. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251

5. Alwine, J.C., Kemp, D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R. and Wahl, G.M. (1979) Methods Enzymol. 68, 220-242
6. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
7. Talkington, C. and Pero, J. (1978) Biochemistry 75, 1185-1189
8. Lee, G., Hannett, N.M., Korman, A. and Pero, J. (1980) J. Mol. Biol. 139, 407-422
9. Moran, C.P.,Jr., Lang, N., Banner, C.D.B., Haldenwang, W.G. and Losick, R. (1981) Cell 25, 783-791
10. Sterlini, J.M. and Mandelstam, J. (1969) Biochem. J. 113, 29-37
11. Favaloro, J., Treisman, R. and Kamen, R. (1980) Methods Enzymol. 65, 718-749
12. Moran, C.P.,Jr., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982) Mol. Gen. Genet. 186, 339-346
13. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353
14. McLaughlin, J.R., Murray, C.L. and Rabinowitz, J.C. (1981) J. Biol. Chem. 256, 11283-11291
15. Ikeuchi, T., Kudoh, J. and Kurahashi, K. (1983) J. Bacteriol. 154, 988-991.
16. Jayaraman, K., Keryer, E. and Szulmajster, J. (1981) FEMS Microbiol. Letters 10, 273-276
17. Kobayashi, Y., Hirochika, H., Kawamura, F., and Saito, H.(1981) in Sporulation and Germination, Levinson, H.S., Sonenshein, A.L. and Tipper, D.J., Eds., pp. 114-118, American Soc. for Microbiol.,Wash. D.C.
18. Ferrari, F.A., Lang, D., Ferrari, E. and Hoch, J.A. (1982) J. Bacteriol. 152, 809-814
19. Shimotsu, H., Kawamura, F., Kobayashi, Y. and Saito, H. (1983) Proc. Natl. Acad. Sci. USA 80, 658-662
20. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41
21. Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S. and Gross,C.A. (1981) Nucleic Acids Res. 9, 2889-2903