Volume 11 Number 18 1983 Nucleic Acids Research

DNA sequence of the tandem ribosomal RNA promoter for *B. subtilis* operon *rrn*B

George C.Stewart* and Kenneth F.Bott+

Department of Microbiology and Immunology, University of North Carolina, Medical School, Chapel Hill, NC 27514, USA

Received 15 July 1983; Accepted 12 August 1983

ABSTRACT

A new ribosomal RNA operon designated rrnB has been identified by screening a Charon 4a library of cloned B. subtilis sequences. Clones containing the promoter region of this operon are unstable in E. coli unless a special vector possessing a transcriptional terminator is used. DNA sequence data suggests that this operon contains two tandem putative promotor regions not unlike those found in E. coli. There are 92 base pairs separating the two "-10 regions" of the promotors. The second is 180 bp upstream from the start site for mature 16S RNA. A potential 29 base pair stem structure necessary for processing of the mature 16S RNA sequence can also be predicted from this analysis.

INTRODUCTION

The ribosomal RNA (rRNA) operons of Escherichia coli are transcribed as a large primary transcript containing precursors for the mature rRNA species plus several transfer RNA (tRNA) determinants. These operons are very active transcriptionally. Although representing only 0.4% of the genome, they account for as much as 50% of the RNA synthesis in rapidly growing cells (1,2). DNA and RNA sequence analyses have indicated that several, perhaps all, of the rRNA operons contain two transcritpion initiation sites (3, 4, 5, 6, 7).

Much less information is available regarding the transcription or initial processing events of the rRNA gene sets of gram positive species. In Bacillus subtilis the production of mature rRNA species will most likely differ from the analogous process of E. coli as evidenced by a lack of tRNA determinants in the majority of the 16S-23S spacer regions (8,9) a different nucleotide sequence in the "stem" surrounding the mature 16S and 23S RNA molecule and a different mechanism involved in the final maturation of 5S rRNA (10).

In this paper we support the hypothesis that there are major differences in processing of B. subtilis rRNA by reporting a stretch of the nucleotide sequence from a cloned rRNA gene set from B. subtilis. Tandem promoter-like sequences are described and, where possible, secondary structure and cleavage

sites which may be involved in the processing of the primary transcript are indicated. In addition, we provide a comparison of sequence generated in our laboratory from two rRNA gene sets with sequence from a gene set located near the putative origin of replication recently reported by Seiki et al. (11).

MATERIALS AND METHODS

Bacterial Strains, Phage, and Plamids. B. subtilis strain 168 was the source of RNA and DNA for recombinant clones. E. coli strain HB101 was the recipient in the subcloning procedure. Phage J25, a Charon 28 clone containing an intact 16S rRNA determinant plus a portion of the 23S rRNA gene from B. subtilis was described in a previous paper (9). The vector plasmid for subcloning was the E. coli plasmid pKM-1, which contains the lambda t_{R1} rho-dependent terminator cloned into the SmaI site of the plasmid pKO-1 (12). Transformation and amplification of plasmid DNA in E. coli were as described by Dagert and Erlich (13) and Norgard (14), respectively.

Chemicals and Enzymes. γ-^{32}P ATP was from New England Nuclear. All restriction enzymes, bacterial alkaline phosphatase, T4 polynucleotide kinase, T4 DNA ligase, and S1 nuclease were from Bethesda Research Laboratories.

Nucleotide Sequence Determination. DNA sequence analysis was by the method of Maxam and Gilbert (15). Restriction fragments were treated with bacterial alkaline phosphatase at 65 C, labelled at their 5'-termini using T4 polynucleotide kinase, denatured by boiling in 50% dimethylsulfoxide and the separated strands resolved by electrophoresis through 5% polyacrylamide (15).

Preparation of RNA. B. subtilis 168 was grown in Antibiotic Medium 3 (Difco) until the absorbance at 500 nm was 0.7. The bacterial cells were harvested by centrifugation, resuspended in 50 mM Tris-HCl (pH = 8.0), 50 mM EDTA, 15% (w/v) sucrose, and converted to protoplasts by incubation at 37 C in the presence of 1 mg/ml lysozyme. The protoplasts were lysed with sodium lauryl sulfate (0.5%, w/v) in the presence of 8 ug/ml proteinase K and 0.08% (v/v) diethyl oxidiformate (Eastman Organic Chemicals). Chromosomal DNA was removed by precipitation on ice following the addition of potassium acetate to a final concentration of 0.5M. The supernatant was extracted with phenol, which had been equilibrated with a Tris buffer to a pH of 5.5, and the phenol removed by ether extraction. The nucleic acid was precipitated with ethanol at -20 C and collected by centrifugation. This pellet was resuspended in water and stored at -20 C.

S1 Mapping Assay. S1 nuclease mapping experiments (16,17) were performed under the conditions described by Aiba et al (18). The reaction products were

separated by electrophoresis on an 8% acrlyamide gel containing 8 M urea (15).

RESULTS

Subcloning of the 5'-Terminus Fragments From the Ribosomal Gene Set. Phage J25, a Charon 28 clone containing sequences homologous to 16S rRNA of B. subtilis on two of five EcoRI fragments internal to the Sau3A generated insert has been described (9). The phage has an additional characteristic of generating a pin-point plaque morphology when plated on its E. coli DP50 host. An EcoRI-generated fragment, whose size was determined to be 1.9 kb by agarose gel electrophoresis, contains the sequences corresponding to the 5' -terminus of 16S RNA and approximately 1.2 kb of DNA upstream of the rRNA determinant. Attempts to subclone this fragment into the EcoRI site of the plasmid pBR322 failed. However, we were able to insert this fragment into the EcoRI site of the plasmid pKM-1 which contains the lambda-phage t_{R1} rho-dependent

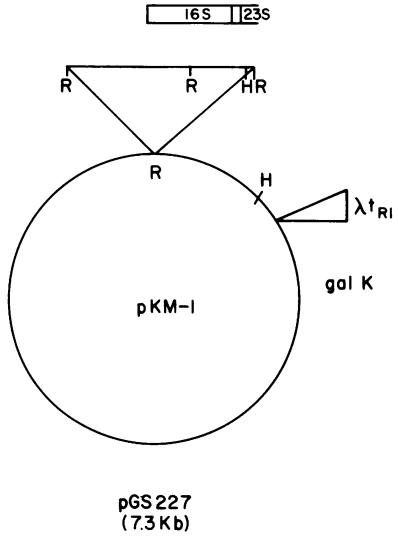

Fig. 1. Construction of plasmid pGS227. A 1.9 kb EcoRI fragment containing the coding sequences for the 5'-terminus of B. Subtilis 16S rRNA and approximately on kb of sequence upstream from the rRNA determinant was ligated into the EcoRI site of the E. coli plasmid pKM-1 (12). Following transformation into E. coli, the plasmid pGS227 was obtained. This plasmid contains, in addition to the 1.9 kb EcoRI fragment, the adjacent 1.1 kb EcoRI fragment from phage J25 (9) which contains the determinants for the 3'-terminus of 16S rRNA and the 5'-terminus of 23S rRNA. The source of DNA in the ligation reaction was a limit EcoRI digest of phage J25 as determined by agarose gel electrophoresis.

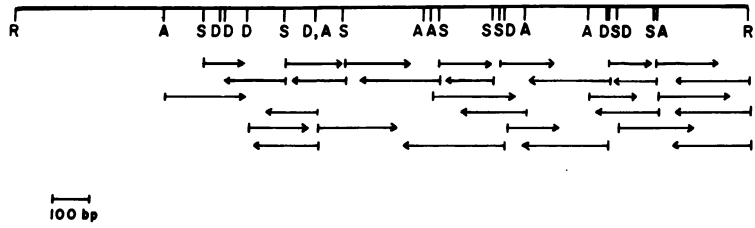

Fig. 2. Strategy for the determination of the nucleotide sequence of the rDNA region of the 1.9 kb EcoRI fragment from pGS227. The arrows indicate the origin and direction of each sequence analysis. Abbreviations used are: A = AluI, D = DdeI, R = EcoRI, and S = Sau3A.

transcription terminator (12). This is in keeping with the expected strong promoter activity of a rRNA gene set and the requirement for terminators of comparable strength to clone strong promoters (19). The resulting plasmid, pGS227, was found to contain the 1.9 kb fragment and also the adjacent 1.1 kb EcoRI fragment from phage J25 (Fig. 1). The plasmid contains an intact 16S rRNA determinant. The 1.9 kb fragment could be stably cloned into the plasmid pKM-1 only in the orientation shown (as evidenced by restriction endonuclease digestions utilizing the single asymetrically located HindIII site with the insert) and in the presence of the adjacent 1.1 kb EcoRI fragment.

Nucleotide Sequence From the 1.9 kb EcoRI Fragment. The strategy used in sequencing this fragment is outlined in Figure 2 and the DNA sequence obtained is presented in Figure 3. The sequence shown originates at a DdeI site approximately 0.82 kb from the left end of the molecule as shown (Fig. 2). Only the sequence from the DNA strand which has the same sense as the rRNA transcript is given. Mature 16S rRNA of B. subtilis would begin at nucleotide 489 of the sequence. The sequence beginning at this site is in agreement with the first five nucleotides of the 5' -terminus of 16S rRNA as determined by Suguira and Takanami (20) and corresponds to the start site of E. coli m16S rRNA when its sequence (21) is aligned with the B. subtilis sequence.

Two promoter-like sequences found at positions 214 (P1) and 306 (P2) are shown in Figure 3. The two putative promoters are separated by 92 bp and the latter sequence is 180 bp upstream from the putative start site for mature 16S rRNA. Both possess Pribnow box (-10) and -35 region sequences typical of E. coli and B. subtilis promoters (22). The -35 and -10 regions are the optimal optimal 17 nucleotides apart and, especially in the case of P1, have the stretches of T's and A's upstream of the -35 region. These are characteristics of promoters which efficiently utilized an RNA polymerase in

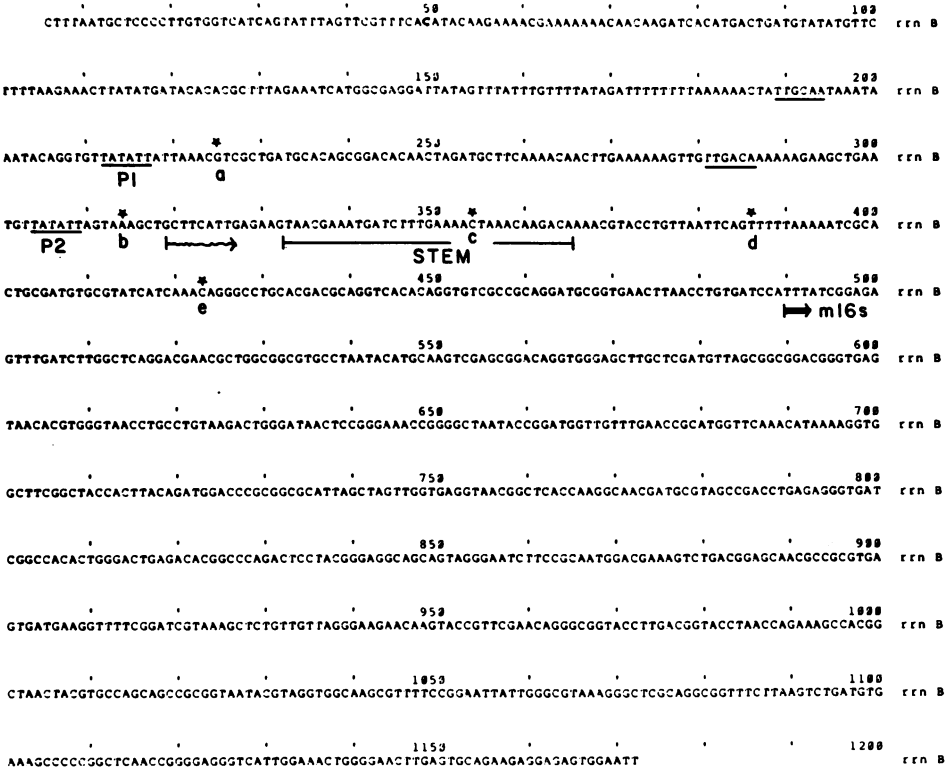

Fig. 3. Nucleotide sequence of the rDNA region of the 1.9 kb EcoRI fragment from pGS227. The tandem promoters are identified as P1 and P2. The "stem" sequence corresponds to the sequence identified in Fig. 5. Positions "a" through "e" are the 5'-terminal positions corresponding to the bands shown in Fig. 4.

association with σ^{55} (22).

S1 Mapping of in Vivo Transcripts. We used the S1 nuclease mapping assay of Aiba et. al. (18) to determine if the putative promoter sequences functioned as such in vivo. The DdeI fragment which corresponds to positions 1-512 and the Sau3A fragment (nucleotides 75-340) were labelled at their 5'-termini with P-32, separated into the individual single strands, and the purified single-stranded DNA hybridized to RNA from B. subtilis. Following treatment with S1 nuclease, transcript-length DNA fragments were resolved on urea-acrylamide gels and visualized by autoradiography. The results of such an experiment are shown in Figure 4. The DNA samples in lanes 1, 4, and 5 were the coding strand of the Sau3A fragment, the non-coding strand of the

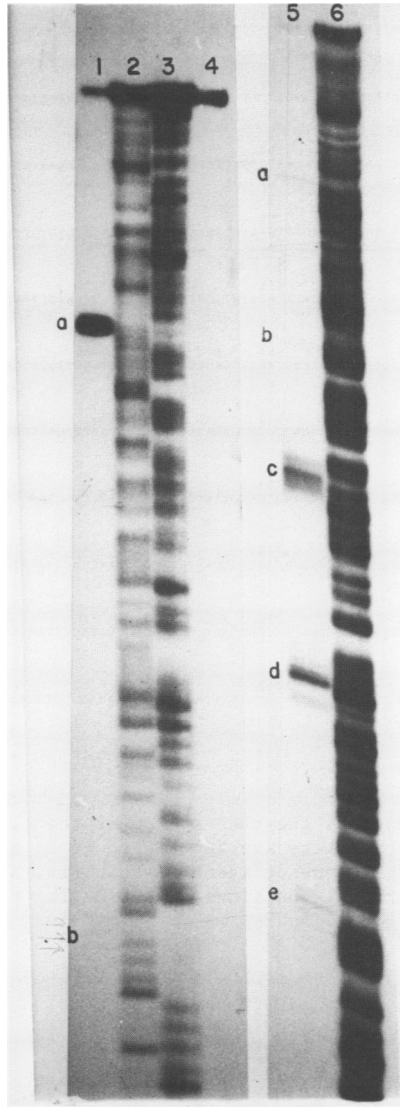

Fig. 4. S1 mapping of in vivo 16S rRNA transcripts. Single-stranded DNA fragments labelled at their 5' -termini with ^{32}P were hybridized to total cellular RNA and treated with S1 nuclease. The following DNA's were used: lane 1, Sau3A (nucleotides 75-340) coding strand, lane 4, the Sau3A noncoding strand, and lane 5, DdeI fragment (nucleotides 1-512) coding strand. The protected DNA fragments were resolved on a 10% acrylamide, 8M urea gels. Lanes 2 and 3 are the Maxam and Gilbert (15) pyrimidine reactions of the Sau3A fragment coding strand and noncoding strand, respectively which were used as size standards. Lane 6 is the purine reaction of the DdeI fragment coding strand. The positions of protected fragments "a" through "e" are identified in Fig. 3.

Sau3A fragment, and the coding strand from the DdeI fragment, respectively. Two DNA fragments were obtained with the coding strand of the Sau 3A fragment. The larger species, labelled "a", corresponds to fragments of approximately 117 nucleotides and therefore to position 224 in the sequence (Fig. 3). A fragment of this size would be expected if the first promoter (P1) were functional in vivo. Three very faint bands centered at position "b" and corresponding to a DNA fragment of 27-29 nucleotides or position 312-314 in the sequence were reproducibly obtained but are not apparent in the photograph. This would correspond to sequences protected from the nuclease by an RNA transcript originating from the second promoter. The intensity of these bands could be low artifically due to strand displacement from "a" but it may also reflect poorer hybridization of the transcript due to the short region of homology. Such bands are not obtained when the noncoding strand (the strand with sequence homology to E. coli 16S rRNA) is used as template. When the coding strand of the DdeI fragment (nucleotides 1-512) was hybridized to the RNA, five species were protected from nuclease digestion. Two of these, "a" and "b" result from the putative transcripts generated from the P1 and P2 promoter sequences. In addition, three smaller species are obtained. Fragment "c", 160 nucleotides in length which corresponds to position 353, fragment "d" of 128 nucleotides (nucleotide position 385 of the sequence), and fragment "e" of 91 nucleotides (position 422). The intensity difference between bands "d" and "e" are reproducible and may indicate a lower concentration of the latter RNA species in the cell. The RNA species with 5'-termini at positions 353, 385, and 422 of the sequence are potential processing intermediates in the rRNA maturation process of B. subtilis.

Possible Secondary Structure of the RNA Transcript. The RNA gene set whose 5'-terminal sequences are carried by phage J25 has its 3'-sequence present on the Charon 4A phage G4 (9, unpublished data). Preliminary nucleotide sequencing from the 3'-end of 16S rRNA (the 1.1 kb EcoRI fragment of phage J25) confirms the sequence generated with the same 1.1. kb fragment from phage G4 (23, L.M. Anderson, personal communication). When the transcript from position "a" through the 16S-23S junction is examined for possible secondary structure, only one significant stretch of homology is present. The sequence from positions 336-364 is capable of forming a hydrogen bonded duplex stem structure with sequence distal to the 3'-terminus of mature 16S rRNA (Fig. 5). The stem-loop structure has a stem of 29 nucleotides including 19 A-U pairs, 7 G-C pairs, 1 G-U pair, and 2 unpaired positions (ΔG (25 C) = -23.2 kcal [24]). The loop portion begins 124

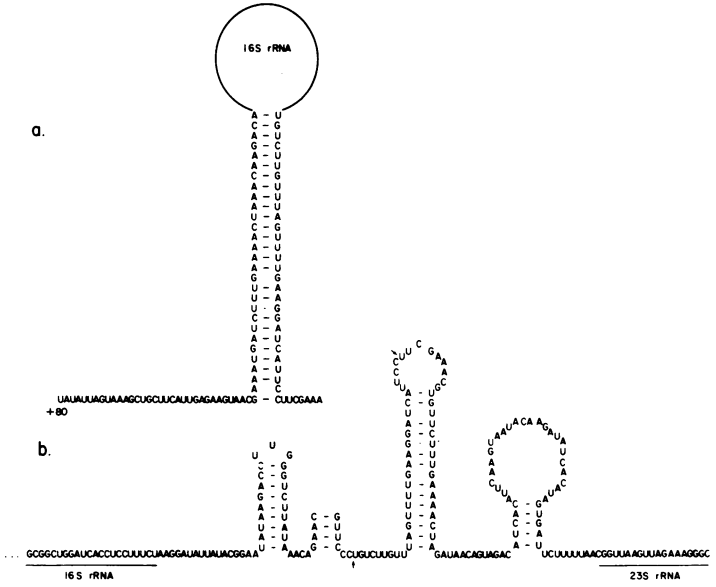

Fig. 5. Possible secondary structure of the RNA transcript. (a): Partial sequence of the primary transcript from position +80 (initiation of transcription from the first promoter) through the 5' -terminus of 23S rRNA is shown. A stem structure of 29 nucleotides (nucleotides 336-364) can be formed with only two mismatches with a sequence distal to the 16S rRNA determinant. The loop portion of this structure contains the coding sequence for mature 16S rRNA as well as 124 nucleotides prior to and 52 nucleotides distal to this sequence. (b): RNA sequence (determined from the DNA sequence) from the 16S-23S junction of the rRNA gene set (23). The arrows indicate the sequence which participates in the formation of the stem structure.

nucleotides prior to the start of mature 16S rRNA and ends 52 nucleotides distal to the mature RNA sequence. Figure 5B shows the sequence of the spacer sequence between mature 16S and 23S rRNA. The region comprising the right-half of the stem structure is delineated by the arrows. That this stem structure might be involved in processing of the transcript is suggested by the observation that in E. coli there is a clip site in the stem at an analogous position on the RNA precursor. Here we see that that an RNA species is present in B. subtilis whose 5'-terminus appears to be the cytosine residue of the C-G pair, 18 base pairs up the stem (position 353 of the sequence) as indicated by band "C" of the nuclease mapping (Fig. 4). This site has a C-C pair flanked on each side by four A-U pairs. The significance of the sequence and/or secondary structure to the mechanism of rRNA maturation in B. subtilis will require identification and characterization of the

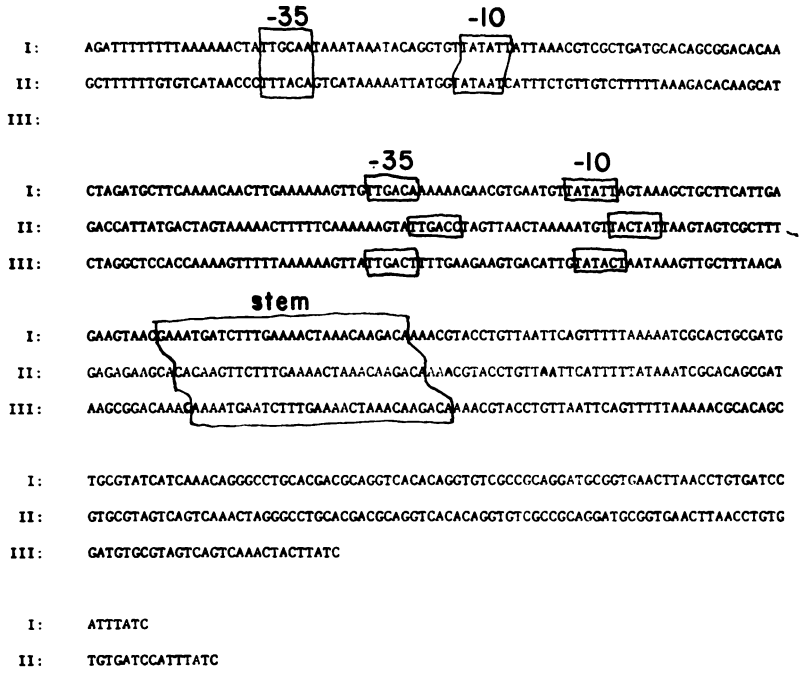

Fig. 6. Nucleotide sequence of the promoter regions from three rRNA gene sets of B. subtilis. I: the sequence from rrnB (this paper), II: the sequence of rrn0 (11,25), III: sequence from p14B1 (9). The start sites of the sequences were arbitrarily chosen to align promoter sequences. The sequences through the first six nucleotides of mature 16S rRNA are given. The -35 and -10 regions of the promoter sequences are indicated and the nucleotides which comprise the left-half of the stem structure (as shown in Fig. 5) are indicated.

enzyme(s) involved.

Comparison of rRNA Gene Set 5'-Terminal Sequences of B. subtilis. In Figure 6 are presented sequences from the three (of ten) rRNA gene sets where sequence data is available. The rRNA gene set (rrn0), which is located near the origin of replication, has tandem promoter sequences (11,25). It barely differs from the sequence of the gene set reported here (rrnB, positioned at approximately 280° in the B. subtilis genetic map [manuscript in preparation]) in the spacing and location of the promoter sequences. The rrn0 promoter sequences are 92 base pairs apart and the second promoter lies 183 base pairs upstream from the 5'-terminus of mature 16S rRNA. The values are 92 and 180 (for rrnB). The upstream promoter sequence has a 16 base sequence between the -35 and -10 regions while the second promoter has 17 nucleotides as seen in

the rrnB promotors. The rRNA gene set represented by the plasmid p14b1 (9) is the second gene set of a tandem. The two are separate by approximately 500 base pairs. This gene set contains only a single promoter sequence which has an 18 nucleotide space between the -35 and -10 regions. All three gene sets contain the stem sequences. Plasmid p14B1 has two changes from the rrnB sequence while rrn0 differs at four sites. There is extensive sequence homology with all three gene sets throughout this region. The major exception to this is that the 14B1 operon has 58 fewer nucleotides in the 5'-region of the loop prior to the start of mature 16S rRNA.

DISCUSSION

DNA sequence analysis of rRNA gene sets of B. subtilis has revealed both similarities and differences between the rRNA operons of E. coli and those of this gram-positive organism. While tandem promoters have so far been the rule with the E. coli operons, at least one of Bacillus rRNA gene sets has only a single promoter sequence. This particular gene set (from p14B1) is the transcriptional unit of the second gene set of tandem arrangement (8). It will be interesting to determine the relative activity of this particular gene set in vivo. There are three such tandems on the B. subtilis chromosomes (9, 26), the pairs being separated by approximately 100 (see phage P11, ref. 9), 200, and 500 (p14B1) base pairs. Perhaps single promoters are a feature of the downstream gene sets. This clustering of gene sets is not found on the E. coli chromosome, which interestingly, has three fewer rRNA operons.

The S1 mapping results are consistent with these promotor regions acting as start sequences for transcription (although they could also be processing points). If these do act as start sequences the promoter strength of the three Bacillus gene sets must vary considerably. The gene set represented by p14B1 has a single promoter sequence which has a less than the concensus spacing between the -35 and -10 regions (22). These sequences are quite stable when cloned onto pBR313 or pBR322 vectors. The sequence of rrn0 contains tandem promoters but only one of these has the concensus spacing typical of strong σ^{55}-RNA polymerase-associated promoters. This sequence is cis-inhibitory to the replication of plasmid vectors (11). The rrnB gene set described in this paper has tandem promoters, each with the 17 nucleotide spacing. We have been able to maintain these sequences only on a plasmid vector which contains a properly oriented transcription terminator. It is convenient to equate "optimum" promoter activity with concensus spacing of 17 nucleotides although "optimal" certainly could be influenced by accessory

proteins which take 35 bp of DNA binding.

A second feature of the sequence is the presence of a stem-loop structure with a 29 base pair stem and flanking a single-stranded loop which contains all the sequence of mature 16S rRNA. Similar stem-loop structures are a prominant feature of RNase III processing of the primary RNA transcript of the rRNA operons of E. coli to yield precursors of 16S and 23S rRNA (27). The stem structures cleaved by RNase III are 30-40 base pairs in length. The Bacillus stem structure may play a similar role in the RNA processing as evidenced by an RNA species in the cell whose 5'-terminus corresponds to a cleavage site within the stem structure. Two other putative processing sites are identified within the loop. One ("d") is associated with a sequence of 5 T's followed by five A's while the second site ("e", 37 nucleotides downstream from "d") is located immediately upstream from a G-C rich region. The significance, if any, of these observations is unknown and must await the isolation of the processing ribonucleases. Care must be exerted in interpreting the regions of this sequence, especially with regard to A and T stretches in that the overall sequence is very A-T rich. The sequence corresponding to mature 16S rRNA (up to the internal EcoRI site) is approximately 74% homologous to the corresponding E. coli sequence.

ACKNOWLEDGEMENTS

We wish to thank F. E. Wilson and M. A. Hollis for valuable discussion, technical assistance and enthusiastic criticism, as well as C. P. Moran and E. P. Geiduschek for critique of the manuscript. This project was supported by grant PCM 8022015 from NSF and National Research Science Award GM07900 to GCS.

+To whom correspondence should be addressed

*Present address: Department of Microbiology, University of Kansas, Lawrence, KS 66045, USA

REFERENCES

1. Kennell, D. (1968) J. Mol. Biol. 34, 85-103.
2. Lazzarini, R. and Dahlberg, AE. (1971) J. Biol. Chem. 246, 420-429.
3. Glaser, G. and Cashel, M. (1979) Cell 16, 111-121.
4. deBoer, H.A., Gilbert, S.F. and Nomura, M. (1979) Cell 17, 201-209.
5. Young, R.A. and Steitz, J.A. (1979) Cell 17, 225-234.
6. Lund, E. and Dahlberg, J.E. (1979) Proc. Natl. Acad. Sci. USA 76, 5480-5484.
7. Shen, W.-F., Squires, C. and Squires, C.L. (1982) Nuc. Acids Res. 10, 3303-3313.

8. Loughney, K., Lund, E. and Dahlberg, J.E. (1982) Nuc. Acids Res. 10, 1607-1624.
9. Stewart, G.C., Wilson, F.E. and Bott, K.F. (1982) Gene 19, 153-162.
10. Sogin, M.L., Pace, B. and Pace, N.R. (1977) J. Biol. Chem. 252, 1350-1357.
11. Seiki, M., Ogasawara, N. and Yoshikawa, H. (1982) Proc. Natl. Acad. USA 79, 4285-4289.
12. McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M. (1981) in Gene Amplification and Analysis, Vol. II, eds. Chirikjian, J.G. and Takis, S. (Elsevier-North Holland, New York) pp. 384-415.
13. Dagert, M. and Ehrlich, S.D. (1979) Gene 6, 23-28.
14. Norgard, M.V. (1981) Anal. Biochem. 113, 34-42.
15. Maxam, A.W. and Gilbert, W. (1980) Meth. Enzymol. 165, 499-580.
16. Berk, A.J. and Sharp, P.A. (1978) Proc. Natl. Acad. Sci. USA 75, 1274-1278.
17. Weaver, R.F. and Weissmann, C. (1979) Nucleic Acids Res. 7, 1175-1193.
18. Aiba, H., Adhya, S. and de Crombrugghe, B. (1981) J. Biol. Chem. 256, 11905-11910.
19. Gentz, R., Langner, A., Chang, A.C.Y., Cohen, S.N. and Bujard, H. (1981) Proc. Natl. Acad. Sci. USA 78, 4936-4940.
20. Sugiura, M. and Takanami, M. (1967) Proc. Natl. Acad. Sci. USA 58, 1595-1602.
21. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) Proc. Natl. Acad. Sci. USA 75, 4801-4805.
22. Moran Jr., C.P., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonnenshein, A.L., Pero, J. and Losick, R. (1982) Mol. Gen. Genet 186, 339-346.
23. Bott, K.F. and Hollis, M.A. (1982) in Molecular Cloning and Gene Regulation in Bacilli, eds. Ganesan, A.T., Chang, S. and Hoch, J.A. (Academic Press, New York) pp. 3-10.
24. Tinoco, Jr., 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.
25. Henckes, G., Vannier, F., Seiki, M., Ogasawara, N., Yoshikawa, H. and Seror-Laurent, S.J. (1982) Nature 299, 268-271.
26. Chow, L.T. and Davidson, N. (1973) J. Mol. Biol. 75, 265-279.
27. Bram, R.J., Young, R.A. and Steitz, J.A. (1980) Cell 19, 393-401.