The RNA Component of the *Bacillus subtilis* RNase P

SEQUENCE, ACTIVITY, AND PARTIAL SECONDARY STRUCTURE*

(Received for publication, December 27, 1985)

Claudia Reich, Kathleene J. Gardiner†, Gary J. Olsen, Bernadette Pace, Terry L. Marsh§, and Norman R. Pace

From the Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, Indiana 47405

The gene defining the catalytic RNA component of RNase P in *Bacillus subtilis* 168 was cloned into bacteriophage λ and plasmid vectors. The nucleotide sequence of the gene and its surroundings was determined from the cloned DNA and by directly sequencing or reverse transcribing the RNase P RNA. The *B. subtilis* RNase P RNA sequence (400-401 nucleotides) is remarkably different from that of *Escherichia coli* (377 nucleotides) (Reed, R. E., Baer, M. F., Guerrier-Takada, C., Donis-Keller, H., and Altman, S. (1982) Cell 30, 627-636; Sakamoto, H., Kimura, N., Nagawa, F., and Shimura, Y. (1983) Nucleic Acids Res. 11, 8237-8251). At best the two are less than 50% similar in sequence. To verify that the RNase P RNA gene was analyzed, a modified, putative gene was cloned adjacent to a bacteriophage T7 promoter and various transcripts were tested for RNase P activity. The intact gene transcript, but not fragments, showed full activity. Full catalytic activity was restored upon mixing the fragments.

The extensive differences between the *B. subtilis* and *E. coli* RNase P RNAs precluded full covariance analysis of secondary structure, but phylogenetically consistent foldings for portions of both molecules could be derived.

Ribonuclease P (RNase P) is the endonuclease responsible for removing 5′ precursor-specific sequences from tRNAs. In vitro and at physiological ionic strength conditions in vivo, RNase P action requires both a protein and an RNA component (1, 2). However, at very high monovalent and divalent cation concentrations, the RNA element is able to catalyze precise tRNA processing in the absence of the protein component (3). The novelty of an RNA enzyme, whose only precedent was the *Tetrahymena* self-splicing intron (4), posed many questions related to its mode of action and the biological rationale for the selection of RNA as a catalytic entity. An important aspect of understanding this functionally novel RNA is gaining insight into its higher order structure; i.e., elucidating its secondary and tertiary foldings. Altman and his colleagues (5, 6) have determined the nucleotide sequence of the *Escherichia coli* RNase P RNA (M1 RNA). They proposed a possible secondary structure based on minimum energy calculations and chemical and enzymatic structure mapping data (7). However, there are credible alternatives to the proposed folding which also are consistent with the structure mapping data.

At this time, the best a priori method for determining the secondary structure of large RNAs is the phylogenetic comparative approach (8). That is, possible helices in an RNA, as indicated by complementary sequences, are tested by seeking the equivalent pairing in the homologous RNA from another organism in which the sequence varies. Helical regions are indicated by covariance in compared sequences; mutations compensate one another to maintain complementarity.

The RNase P RNAs from *E. coli* and *Bacillus subtilis* are homologous in function; the RNA and protein subunits from each organism will complement those from the other in the low salt, holoenzyme reaction (2). Therefore, the structural elements involved in catalysis and RNase P protein interactions likely are similar in the two RNAs. In order to analyze the RNase P RNA structure by phylogenetic comparisons, we have cloned the *B. subtilis* 168 RNase P RNA gene and report here the determination of its nucleotide sequence. In addition, we describe the manipulation of the gene into an efficient in vitro expression vector and show that the gene product is enzymatically active.

**MATERIALS AND METHODS AND RESULTS**

**DISCUSSION**

The isolation of the RNase P RNA gene from *B. subtilis* proved unusually difficult. Not only was the genomic region containing the gene subject to deletion in the vectors and hosts employed, but also the low cellular abundance of the RNA rendered questionable the purity of hybridization probes. We estimate, based on recoveries of the RNA during purification, that only about 20–50 copies of the RNase P RNA are present in each *B. subtilis* cell. Additionally, the termini of the RNase P RNA are not good substrates for polynucleotide kinase and RNA ligase, which were used for

1 Portions of this paper (including “Materials and Methods,” “Results,” and Figs. 1 and 3–5) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 85M-4224, cite the authors, and include a check or money order for $4.00 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

*This work was supported by Grant GM-34527 from the National Institutes of Health. DNA synthesis was partially funded by a grant from the Indiana Corp. for Science and Technology to the Institute for Molecular and Cellular Biology, Indiana University. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

† Present address: Eleanor Roosevelt Institute of Cancer Research, 4200 E. Ninth Ave., Denver, CO 80262.

‡ Present address: Dept. of Biology, Hamilton College, Clinton, NY 13323.
**B. subtilis RNase P RNA**

The first seven nucleotides of RNase P RNA show only one mismatch when compared to the consensus promoter sequences. This overlapping promoter arrangement conceivably allows subtle control of transcription of the RNase P RNA gene. Although we have not observed an RNA containing extra nucleotides at its 5' end, it could be a less abundant species or be processed over the observed, mature form. Neither of the promoter-like elements pointed out shows perfect agreement to the consensus sequence derived from frequently transcribed B. subtilis genes. This is not surprising in view of the low abundance of RNase P RNA in vivo (above).

An 80-amino acid open reading frame (ORF) begins 97 pairs distal to the 3' terminus of the mature domain of the RNase P RNA gene there occur neighboring, inverted repeats, which in an RNA transcript would be capable of forming a hairpin consisting of a 9-base pair stem and a 3-base pair loop. Such a structure would be similar to numerous, r-independent, transcription termination sites (27). We have no evidence that this structure serves to terminate transcription of the gene. If it does, however, then the mature 3' end of the RNase P RNA must be generated by some processing event, as is the case in E. coli (28).

Twenty base pairs distal to the 3' terminus of the mature domain, extends 53 potential codons to the end of the region sequenced. The lack of a

\[ \text{The abbreviation used is: ORF, open reading frame.} \]
compelling Shine-Dalgarno sequence, the overlap with the first ORF, and the lack of an upstream promoter make the significance of this ORF questionable.

The sequences of the RNA components of RNase P from

B. subtilis and E. coli (5, 6) are strikingly dissimilar. Fig. 6A is a "dot plot" showing those regions of the two sequences in which at least 9 of 12 consecutive nucleotides are identical. For comparison, panel b of the figure is a similar analysis of the first 400 nucleotides of the 16 S rRNAs from these two organisms (30, 31). The divergence of the RNase P RNA sequences makes the identification of corresponding regions in the two molecules impossible over much of their lengths. We have aligned the B. subtilis and E. coli sequences on the basis of similar potential secondary structures in the terminal regions of the two molecules (below) and maximizing primary structural homology in the intervening regions (not shown). The overall sequence similarity (32) in this alignment is 43%, compared with 78% for an alignment of their 16 S rRNA sequences. There occur large regions in one sequence with no homolog in the other, at least in part owing to the significant length difference in the B. subtilis and E. coli RNase P RNAs. If the effects of these regions on the homology are limited by counting only the first five consecutive alignment gaps, the net sequence similarity is still only 49%.

The secondary structure similarities of the B. subtilis and E. coli RNase P RNA terminal regions mentioned above are illustrated in Fig. 7. Except for the pairing of the termini, these foldings are different from those proposed previously for the E. coli RNase P RNA (7). Each of the stems in the models shown in Fig. 7 includes pairs of sequence positions which differ in primary structure between the two species, but which covary so as to preserve their complementarity. These examples of nucleotide covariance suggest a biological importance of the sequence complementarities and hence provide evidence for the secondary structural elements which contain them. Unfortunately, the unexpectedly low homology of the B. subtilis and E. coli sequences has made the alignment of the sequences sufficiently uncertain to preclude extension of the B. subtilis structure model further at this time, without undue speculation. In this vein, we have not unambiguously located the B. subtilis homolog to the presumptive pairing between E. coli sequence positions 152–156 and 161–165, which is supported by a covariance relative to the Salmonella typhimurium sequence (33).

It has been pointed out that the E. coli RNase P RNA (M1 RNA) contains several short sequence repeats (5). The base composition of the B. subtilis RNase P RNA is 28:20:28:23 (per cent A:C:G:U), much more uniform than that of the E. coli sequence (23:27:35:15). An expected consequence of this
pattern of nucleotide usage is a decrease in the number of random recurrences of short sequences (34). Indeed, there are far fewer directly repeated, short (6 to 20 nucleotide) sequences in the B. subtilis RNA than in that of E. coli (data not shown). It is likely that the nucleotide sequence similarities, owing to the base composition, completely explain the sequence repeats in the E. coli RNA P RNA, but their absence in B. subtilis suggests that they are not functionally important.

Finally, we point out that the experiment shown in Fig. 5, which reconstructs fully active RNase P RNA from fragments, suggests that the RNase P RNA tertiary structure, not a simple nucleotide sequence, is required for complete activity. Although the experiment does not prove the point, since the EcoRI site might interrupt the active site, it shows that the regions of the RNase P RNA required for activity can be sought by in vitro recombination of abbreviated fragments.

Acknowledgment—We thank David Waugh for the preparation of labeled RNA precursor.

REFERENCES

1. Stark, B. C., Kole, R., Bowman, E. J., and Altman, S. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 3719-3721
2. Gardiner, K., and Pace, N. R. (1980) J. Bacteriol. 145, 7507-7509
3. Guerrier-Takada, C., Gardner, K., Marsh, T. L., and Altman, S. (1983) Cell 36, 549-549
4. Kruger, K., Grabowski, P. J., Zieg, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Cell 31, 147-157
5. Reed, R. E., Baer, M. P., Guerrier-Takada, C., Donis-Keller, H., and Altman, S. (1982) Cell 30, 627-636
6. Sakamoto, H., Kimura, N., Nagawa, F., and Shimura, Y. (1983) Nucleic Acids Res. 11, 8339-8346
7. Guerrier-Takada, C., and Altman, S. (1984) Biochemistry 23, 6327-6334
8. Gutell, R. R., Weiser, B., Woese, C. R., and Noller, H. F. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 155-156
9. Gardner, K. J., Marsh, T. L., and Pace, N. R. (1985) J. Biol. Chem. 260, 5415-5419
10. Stahl, D. A., Luehrsen, K. R., Woese, C. R., and Pace, N. R. (1981) Nucleic Acids Res. 9, 6129-6137
11. Donis-Keller, H., Maxam, A., and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527-2538
12. Benton, W. D., and Davis, R. W. (1977) Science 196, 180-182
13. Caruthers, M. H., Beaucage, S. L., Becker, C., Efchavit, J. W., Fisher, E. F., Galluppi, G., Goldman, R., de Haseth, P., Matteucci, M., McBride, L., and Stabinsky, Y. (1983) in Gene Amplification and Analysis (Papas, T. S., Rosenberg, M., and Chirikjian, J. G., eds) Vol. 3, pp. 1-26, Elsevier, New York
14. Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., and Pace, N. R. (1986) Proc. Natl. Acad. Sci. U. S. A. 82, 6955-6959
15. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Gene (Amst.) 33, 103-119
16. Bolivar, F., and Backman, K. (1980) Methods Enzymol. 65, 245-265
17. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
18. Saha, B., Strelow, S., and Schlessinger, D. (1983) J. Biochem. Biophys. Methods 7, 277-294
19. Messing, J. (1983) Methods Enzymol. 101, 20-89
20. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
21. Tabar, S., and Richardson, C. C. (1986) Proc. Natl. Acad. Sci. U. S. A. 82, 1074-1078
22. Hassour, S. M., and Whitlock, H. W., Jr. (1974) Anal. Biochem. 59, 162-164
23. Green, M. R., Maniatis, T., and Melton, D. A. (1983) Cell 32, 681-694
24. Kornberg, E. M. (1975) J. Mol. Biol. 88, 503-518
25. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompkson, K., Faber, H. E., Furlong, L. A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Sheldon, E. L., and Smithies, O. (1977) Science 196, 161-169
26. Gotohara, N., Moriya, S., and Yoshikawa, H. (1983) Nucleic Acids Res. 11, 6301-6318
27. Rosenberg, M., and Court, D. (1979) Annu. Rev. Genet. 13, 319-333
28. Reed, R. E., and Altman, S. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 5359-5362
29. Shira, J., and Dalgarino, L. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 1342-1346
30. Green, C. J., Stewart, G. C., Hollis, M. A., Vold, B. S., and Bott, K. F. (1985) Gene (Amst.) 37, 261-268
31. Brosius, J., Dull, T. J., Sleeter, D. D., and Noller, H. F. (1981) J. Mol. Biol. 148, 107-127
32. McCarroll, R., Olsen, G. J., Stahl, Y. H., Woese, C. R., and Sogin, M. L. (1983) Biochemistry 22, 5858-5868
33. Bier, M., and Altman, S. (1985) Science 228, 999-1002
34. Good, W. B., and Kanehsita, M. (1982) Nucleic Acids Res. 10, 247-263
to transform E. coli HB101. Recombinant M13 phage DNA was used to transform E. coli HB101 (19).

DNA sequencing — DNA sequence determinations were carried out by the dideoxynucleotide-terminated chain elongation method using either single-strand phage M13 or double-strand plasmid templates. DNA primers were vector-specific primers (19) or oligonucleotides complementary to sequences on the ends of either the 5' or 3' terminus of the plasmid DNA molecule, as indicated in all cases, they were generated during subsequent growth, presumably through rearrangement and deletion of cloned material. The template sequences used were those of R. P. RNA genes from *Helicobacter pylori* (unpublished).

In vivo synthesis of run-off transcripts — Recombinant plasmid DNA carrying either all or part of the RNase P RNA gene adjacent to a single restriction site was linearized and used to transform competent *E. coli* DH5α cells. The linearized plasmid products were resolved by electrophoresis through 6% polyacrylamide/8 M urea gels. The transcription reaction mixture in the presence of 1 nM of each of the four ribonucleoside triphosphates and 60 nM RNA polymerase (a gift from Dr. O. Ohnishi) was incubated with T7 RNA polymerase (supplied by Dr. O. Ohnishi) in T7 transcription buffer (21). The transcription reaction products were resolved by electrophoresis through 6% polyacrylamide/8 M urea gels.

RNA cleavage by RNase H — The RNase P RNA gene of *E. coli* was cloned in *E. coli* HB101. Recombinant M13 phage DNA was used to transform E. coli HB101. Recombinant M13 phage DNA was used to transform E. coli HB101. Recombinant M13 phage DNA was used to transform E. coli HB101 (19).

3' termini of RNase P RNA were assayed for RNase P activity in the absence of 

The described plasmid DNAs were isolated, linearized with an appropriate restriction endonuclease, and used to transform competent *E. coli* DH5α cells as HindIII-EcoRI and EcoRI-HindIII fragments, respectively. Each fragment was subcloned as a HindIII-PstI unit into the in vitro expression vector for RNase P RNA

In vitro synthesis of in vitro expression vectors for RNase P RNA — The sequence of the cloned, synthesizable, RNase P RNA gene proved to be identical to that of strain 108-8, as determined by the method of cocktail labeling (see Discussion). The extent of sequence divergence between the two strains was less than 1% of the entire sequence, with the exception of the 5' termini. The complete sequence of the RNA and its flanking regions, together with the sequences determined from the RNA, are shown in Fig. 2. Other features, which will be referred to in Discussion, are also highlighted.

Construction of an in vitro expression vector for RNase P RNA — The sequence of the cloned, synthesizable RNase P RNA gene proved to be identical to that of strain 108-8, as determined by the method of cocktail labeling (see Discussion). The extent of sequence divergence between the two strains was less than 1% of the entire sequence, with the exception of the 5' termini. The complete sequence of the RNA and its flanking regions, together with the sequences determined from the RNA, are shown in Fig. 2. Other features, which will be referred to in Discussion, are also highlighted.

![FIG. 1. Partial restriction map of the *H. pylori* RNase P RNA gene. The solid box indicates the mature RNase P RNA. Arrow indicates a bulge in the predicted secondary structure.](image-url)
FIG. 4. Gel-electrophoretic analysis of purified in vitro transcripts. The various gel purified run-off transcripts (see Materials and Methods) were resolved by electrophoresis on an 8% polyacrylamide gel containing 8M urea at 15 V/cm for 2 h. The RNAs were visualized by ethidium bromide staining. Lanes 1-6 are labeled as in Fig. 3. Lane 7 is 5S RNA isolated from B. subtilis cells.

FIG. 5. Activity of in vitro transcripts of RNase P RNA. Run-off transcripts were assayed for their ability to process uniformly [32P]-labeled precursor tRNA in a protein-free reaction (see Materials and Methods). Reaction products were resolved on an 8% polyacrylamide gel containing 8M urea and visualized by autoradiography. RNAs are designated as in Fig. 3. For reactions containing both, the 5' and 3' fragments were either "mixed" without incubation, or were "annealed" by heating to 70°C and slowly cooling to 25°C.