Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 6187–6191, October 1983
Biochemistry

Processing of transcription products of the gene encoding the RNA component of RNase P

(*in vitro* transcription/sequence analysis/*in vitro* processing)

HIROSHI SAKAMOTO, NAOHIRO KIMURA, AND YOSHIRO SHIMURA

Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606, Japan

Communicated by Daniel Nathans, July 8, 1983

ABSTRACT **The gene coding for the RNA component of RNase P from *Escherichia coli* was transcribed *in vitro* by using a restriction fragment carrying the gene as template. The terminal sequences of the transcription products were determined and mapped on the DNA sequence of the gene. The signals for transcription initiation and termination of the gene were thus identified. Transcription termination occurs within a region of two bases at positions 413 and 414 from the transcription start site. The transcripts carry extra stretches of 36 and 37 nucleotides at the 3′ end of the RNA molecule isolated from the enzyme. The extra sequences were removed *in vitro* when the transcripts were incubated with a crude cell extract.**

RNase P is a processing endonuclease that cleaves tRNA precursors to generate the 5′ termini of mature tRNA sequences (1). The enzyme is essential for the biosynthesis of all tRNA species in *Escherichia coli* (2, 3). RNase P contains both protein and RNA components that can be dissociated from each other and subsequently reassembled *in vitro* to reconstitute RNase P activity (4, 5). We have obtained evidence indicating that the protein component alone has a nonspecific nuclease activity but, when associated with the RNA component *in vitro*, the specificity of RNase P toward substrate is restored (unpublished results). Thus, the RNA component appears to play a key role in substrate recognition. Recently, the nucleotide sequence of the gene coding for the RNA component from an *E. coli* strain was determined (6). The sequence of the gene for the RNA component from a temperature-sensitive mutant of defective RNase P function (*ts709*) has also been analyzed and compared with that from its parental wild-type strain (unpublished results). The terminal sequences of the RNA component prepared from RNase P were determined and mapped on the DNA sequence. Thus, the complete sequence of the RNA molecule has been determined. We have carried out *in vitro* transcription of the gene by using a restriction fragment containing the gene as template, and we have identified the major transcription products. Structural analysis of the transcripts has revealed the signals for transcription initiation and termination of the gene. In addition, the analysis has also shown that the transcripts of the gene are processed to form the RNA component present in RNase P.

MATERIALS AND METHODS

Chemicals. All chemicals, including ^{32}P-labeled nucleotides, DEAE-cellulose plates, and polyethyleneimine cellulose plates, have been described (7).

Enzymes. RNA polymerase holoenzyme was purified to near homogeneity from *E. coli* A19 as described (7). All other enzymes used were described previously (7, 8). The ρ factor, which was purified as described (9), was a gift from M. Imai of Kyoto University.

Preparation of DNA. DNA from λgrnpR-W, a clone containing the wild-type gene for the RNA component of RNase P, was prepared as described (7, 8). The *Hin*cII fragment of 2.8 kilobase pairs (designated *Hin*c2.8kb) was prepared by digestion of λgrnpR-W DNA with *Hin*cII, followed by electrophoresis on a 1% agarose gel.

***In Vitro* Transcription.** Unless indicated otherwise, RNA synthesis was carried out as described (7). Purified ρ factor (5 μg/ml) was added where indicated.

Polyacrylamide Gel Electrophoresis. The procedures for polyacrylamide gel electrophoresis and for recovery of RNA from the gel were the same as described in ref. 7.

Analysis of Terminal Nucleotide Sequences of Transcription Products. The 5′- and 3′-terminal sequences of transcription products were determined by using the same procedures described in ref. 7.

***In Vitro* Processing of RNA Transcripts.** A dialyzed S30 extract was prepared from *E. coli* Q13 as described (2). The transcripts of the gene for the RNA component of RNase P were labeled with [α-^{32}P]ATP and purified by electrophoresis on a 4% polyacrylamide gel/8 M urea. The RNA (about 2 fmol) was incubated in the reaction mixture containing, in 80 μl, 10 mM Tris·HCl (pH 7.8)/5 mM $MgCl_2$/0.1 mM EDTA/10 mM 2-mercaptoethanol/10 μl of the S30 extract (about 180 μg of protein) at 37°C. After incubation, the RNA was extracted with phenol, precipitated with ethanol, and analyzed by polyacrylamide gel electrophoresis.

RESULTS

Identification of Transcription Products of the Gene for the RNA Component of RNase P. We have found that the gene coding for the RNA component of RNase P from *E. coli* is present within *Hin*c2.8kb, a fragment generated by digestion of the 10-kilobase-pair *Eco*RI fragment of λgrnpR-W with *Hin*cII. Furthermore, the sequence of the RNA component was mapped on the *Hin*c2.8kb fragment, and the complete nucleotide sequence of the gene and its flanking regions has been determined. From the DNA sequence and the 5′- and 3′-terminal sequences of the RNA, the nucleotide sequence of the RNA component has also been determined. These results will be described elsewhere. When the *Hin*c2.8kb fragment was transcribed *in vitro* with *E. coli* RNA polymerase in the absence of ρ factor and RNA products were fractionated by electrophoresis on a 4% polyacrylamide gel/8 M urea, a single major band and at least several minor bands were detected (Fig. 1).

The major RNA band was ≈400 nucleotides long, as judged by the electrophoretic mobility, and represented >85% of the transcription products of the restriction fragment on the basis

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

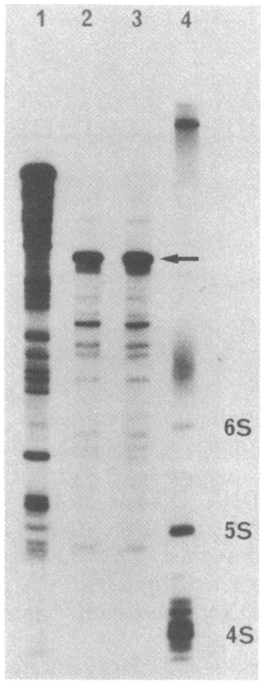

FIG. 1. Identification of *in vitro* transcripts of the gene encoding the RNA component of RNase P. The *Hinc*2.8kb fragment was transcribed as described, with [α-^{32}P]ATP as the radioactive substrate. RNA synthesis was carried out in the absence (lane 2) or presence (lane 3) of ρ factor. The transcripts were fractionated by electrophoresis on 4% polyacrylamide gel/8 M urea and autoradiographed. The major transcript, ≈400 nucleotides long, is marked by an arrow. The *in vitro* transcripts of the *Hae*1.4kb fragment containing the *supB-E* tRNA operon of *E. coli* (7) and RNAs synthesized in *ts241*, a thermosensitive RNase P mutant, at 42°C (2) were electrophoresed on the same gel (lanes 1 and 4, respectively) as size markers.

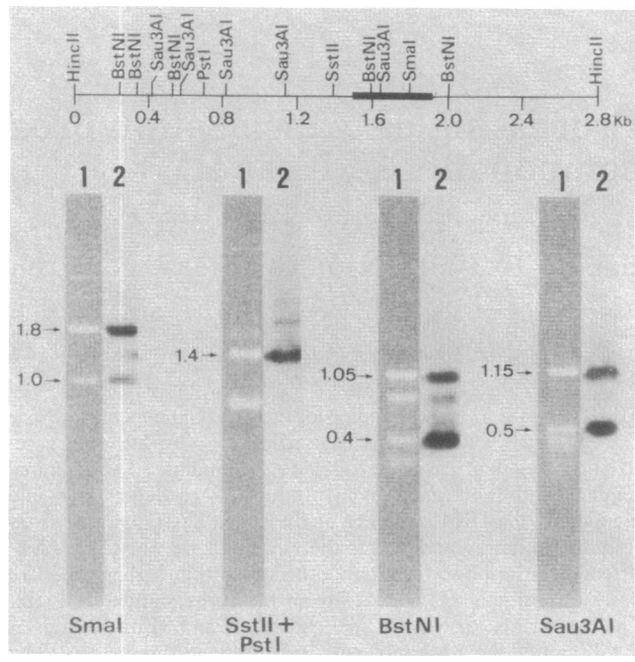

FIG. 2. Map of the major *in vitro* transcript on the *Hinc*2.8kb fragment. (*Lower*) Patterns of Southern blot hybridization of the restriction endonuclease digests of the *Hinc*2.8kb fragment. The *Hinc*2.8kb fragment (0.3 μg) was digested with *Sma* I, *Pst* I/*Sst* II, *Bst*NI, or *Sau*3A1 and then electrophoresed on a 1% agarose gel. The DNA bands were visualized by staining with ethidium bromide (lanes 1), transferred onto nitrocellulose filter, and then hybridized with the major transcript detected in Fig. 1 (lanes 2). (*Upper*) Restriction map of the *Hinc*2.8kb fragment. The region corresponding to the sequence of the RNA component from RNase P is indicated by a thick bar.

of the radioactivity incorporated. To map this RNA on the *Hinc*2.8kb fragment, the major band was recovered from the gel and subjected to Southern blot hybridization with various restriction enzyme digests of the *Hinc*2.8kb fragment. The hybridization profiles of the *Sma* I, *Pst* I/*Sst* II, *Bst*NI, and *Sau*3A1 digests with the major transcript (Fig. 2) were the same as those observed with the end-labeled RNA component prepared from RNase P, and they were consistent with the map of the gene for the RNA component in the *Hinc*2.8kb fragment. These results show that the RNA product of ≈400 nucleotides represents the transcript of the gene.

When the *Hinc*2.8kb fragment was transcribed *in vitro* in the presence of ρ factor, production of the major RNA band as well as the minor bands was essentially unaffected (Fig. 1). When λ phage DNA was transcribed in the presence of ρ factor under the same conditions, the transcripts >800 nucleotides long that were detected in the absence of the termination factor disappeared (data not shown). These results led us to conclude that transcription termination of the gene coding for the RNA component of RNase P is independent of the ρ factor.

Terminal Sequences of the Transcription Product of the Gene. To identify the 5′-terminal nucleotide of the *in vitro* transcript of the gene for the RNA component, the *Hinc*2.8kb fragment was transcribed in the presence of [γ-^{32}P]ATP/[γ-^{32}P]-GTP/[γ-^{32}P]CTP/[γ-^{32}P]UTP. The major transcript ≈400 nucleotides long was purified by polyacrylamide gel electrophoresis and subsequently digested with nuclease P1. The digests were chromatographed on a polyethyleneimine cellulose plate and autoradiographed. As shown in Fig. 3*a*, only pppG was detected. To map the transcription start site more specifically in the DNA sequence previously determined, the 5′-terminal sequence of the transcript was determined. The *Hinc*2.8kb fragment was transcribed in the presence of [γ-^{32}P]GTP, and the major transcript was purified by gel electrophoresis. The RNA was subjected to partial digestion with RNase T1, RNase U2, RNase PhyM, RNase from *Bacillus cereus* (Bc), or alkali, and subsequently the digests were electrophoresed on a 20% polyacrylamide gel. As shown in Fig. 3*b*, the sequence of at least 17 nucleotides from the 5′ end of the transcript was determined. The sequence is the same as the 5′-terminal sequence of the RNA component isolated from RNase P and present uniquely in the DNA sequence previously determined (see Fig. 5). It is worth noting that the 5′-terminal guanine residue of the transcript is seven base pairs downstream from a heptameric T-A-T-A-C-T-G sequence that is consistent with the Pribnow box sequence (13) and 31 base pairs downstream from a hexameric G-T-G-A-C-A sequence that is consistent with the sequence of the −35 region (14). It is likely that these sequences function as the promoter of the gene.

To identify the ρ-independent termination site(s) of the gene, the 3′-terminal sequences of the transcript were determined and mapped on the DNA sequence previously determined. The transcripts of the *Hinc*2.8kb fragment were end-labeled at the 3′ termini with [5′-^{32}P]pCp by T4 RNA ligase and purified by electrophoresis on a 4% polyacrylamide gel. The end-labeled major transcript was digested to completion with RNase T1, and the digests were separated by two-dimensional homochromatography. As shown in Fig. 4*a*, two radioactive spots (A and B) were detected in the digests. These spots were reproducibly observed from the transcript that migrated as a single band on 5% and 6% polyacrylamide gels and also from the transcripts

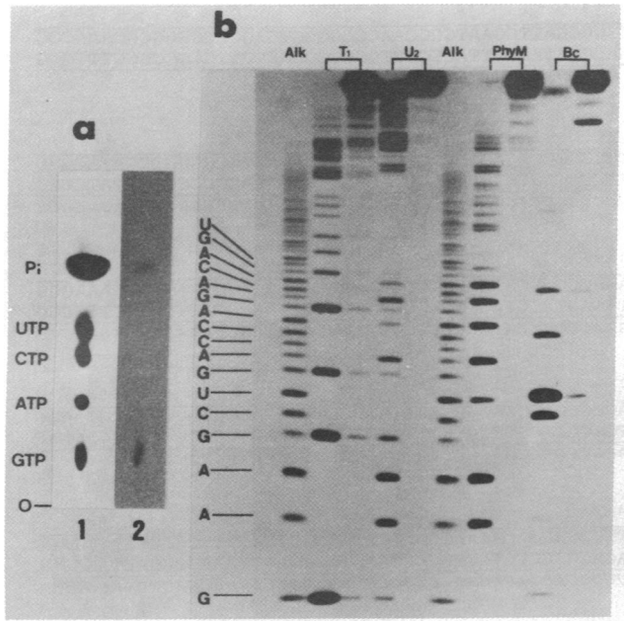

FIG. 3. Analysis of the 5′-terminal sequence of the transcript of the gene for the RNA component of RNase P. (*a*) Identification of the 5′-terminal nucleotide of the transcript. The *Hinc*2.8kb fragment was transcribed in the presence of [γ-^{32}P]ATP/[γ-^{32}P]GTP/[γ-^{32}P]CTP/[γ-^{32}P]UTP. The products were electrophoresed as in Fig. 1, and the 400-nucleotide-long RNA band was recovered from the gel and digested with nuclease P1. The digests were chromatographed on a polyethyleneimine cellulose plate according to the protocol of Cashel *et al.* (10) and autoradiographed. Lane 1, [γ-^{32}P]ATP/[γ-^{32}P]GTP/[γ-^{32}P]CTP/[γ-^{32}P]-UTP used as marker. P_i, orthophosphate; O, origin. Lane 2, the nuclease P1 digests of the transcript. (*b*) Nucleotide sequence of the 5′-terminal region of the transcript. The *Hinc*2.8kb fragment was transcribed in the presence of [γ-^{32}P]GTP and the major transcript was purified as in *a*. The RNA was subjected to partial digestion with RNase T1 (lanes T1), RNase U2 (lanes U2), RNase PhyM (lanes PhyM), RNase from *B. cereus* (lanes Bc), or alkali (lanes Alk), as described by Donis-Keller and colleagues (11, 12). In RNase T1 digestion, RNA was digested with 0.1 unit (left lane) and 0.01 unit (right lane) of the enzyme. In RNase U2 digestion, RNA was digested with 5 units (left lane) and 0.5 unit (right lane) of the enzyme. In RNase PhyM and RNase Bc digestions, RNA was digested with 2 units (left lanes) and 0.2 unit (right lanes) of each enzyme. The digests were electrophoresed on 20% polyacrylamide gel/8 M urea and autoradiographed.

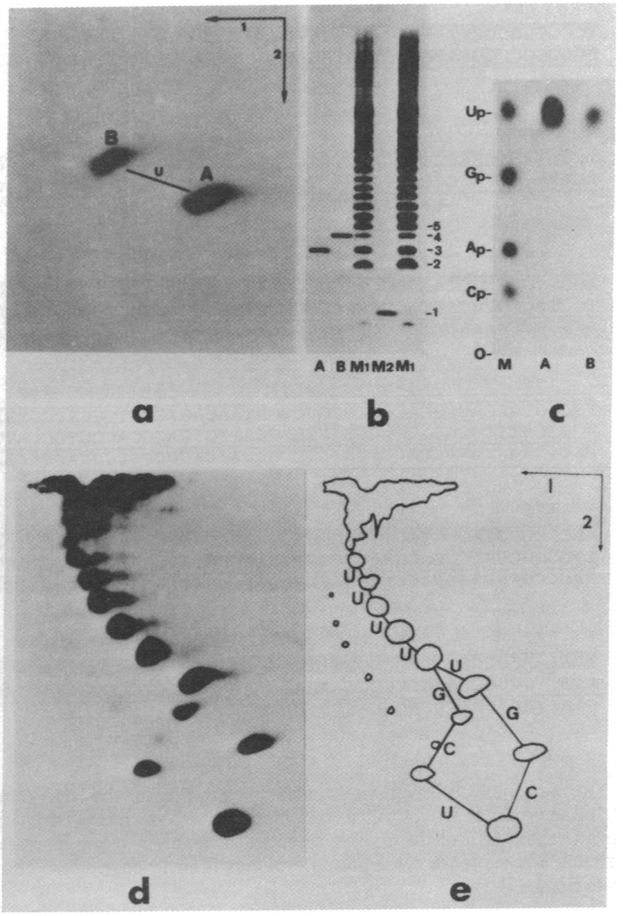

FIG. 4. Analysis of the 3′-terminal sequences of the transcript of the gene for the RNA component of RNase P. (*a*) The transcripts of the *Hinc*2.8kb fragment were end-labeled at the 3′ termini with [5′-^{32}P]pCp by T4 RNA ligase and purified by electrophoresis as in Fig. 1. The major transcript was recovered from the gel and digested to completion with RNase T1. The digests were subjected to two-dimensional separation according to the method of Jay *et al.* (15); first dimension (from right to left) by electrophoresis on cellulose acetate in pyridine acetate/7 M urea, pH 3.5, and second dimension (from top to bottom) by homochromatography on a DEAE-cellulose thin-layer plate in Homo-mix V of Jay *et al.* (15). After chromatography, the plate was dried and autoradiographed. (*b*) Oligonucleotides corresponding to spots A and B in *a* were eluted from the plate with triethylamine carbonate and electrophoresed on 20% polyacrylamide gel/8 M urea. Lanes: A, spot A; B, spot B; M1, partial alkaline digests of the end-labeled transcript; M2, complete RNase T2 digests of the end-labeled transcript showing the position of mononucleotides. Numbers indicate oligonucleotide chain lengths. (*c*) Oligonucleotides corresponding to spots A and B were recovered from the thin-layer plate in *a* and individually digested with RNase T2. The digests were electrophoresed on Whatman 540 paper in pyridine acetate (pH 3.5). The paper was dried and autoradiographed. Lanes: M, mixture of ^{32}P-labeled ribonucleoside 3′-monophosphates as marker; A, digests of spot A; B, digests of spot B. O, origin. (*d*) Major transcript labeled at the 3′ end prepared as in *a* was subjected to partial alkaline hydrolysis as described (11). Digests were subjected to two-dimensional separation as in *a*. (*e*) Schematic interpretation of radioactive oligonucleotides in *d*.

incubated with T4 RNA ligase for prolonged periods. The two spots, which represent 3′-terminal oligonucleotides of the major transcript, differ from each other by the contents of a single UMP residue, as judged by their positions in the chromatogram. The spot B must have an additional UMP as compared to the spot A. From the electrophoretic mobilities on a 20% polyacrylamide gel, the spots A and B were shown to be three and four nucleotides long, respectively, including the 3′-terminal CMP residue ligated to the 3′ end of the transcript (Fig. 4*b*). When the two oligonucleotides were individually analyzed by RNase T2 digestion followed by paper electrophoresis, ^{32}P-labeled UMP was generated in both cases (Fig. 4*c*). These results indicate that transcription terminates at two adjacent positions within a dinucleotide 5′ T-T sequence that is two and three nucleotides downstream from a GMP residue. Quantification of the radioactivity of the spots has revealed that the molar ratio of spot A to B is approximately 13:7. It appears, therefore, that the 5′-proximal thymidine is the preferred termination point, at which about 65% of the chains terminate. To localize the T-T sequence in the DNA sequence of the gene, the major transcript labeled at the 3′ end was subjected to partial alkaline hydrolysis and the resulting digests were fractionated by two-dimensional homochromatography. As shown in Fig. 4 *d* and *e*, two series of oligonucleotides were observed; one series that appeared to be major could be read as 3′ X-C-G-U-U-U-U-U, where X is the 5′-proximal base of the smallest oligonucleotide spot in the figure, while the other series had an additional UMP at the 3′ end of the major series. In the chromatogram shown in the figure, spots for mononucleotides were not seen. Because the majority of transcripts have the 3′-ter-

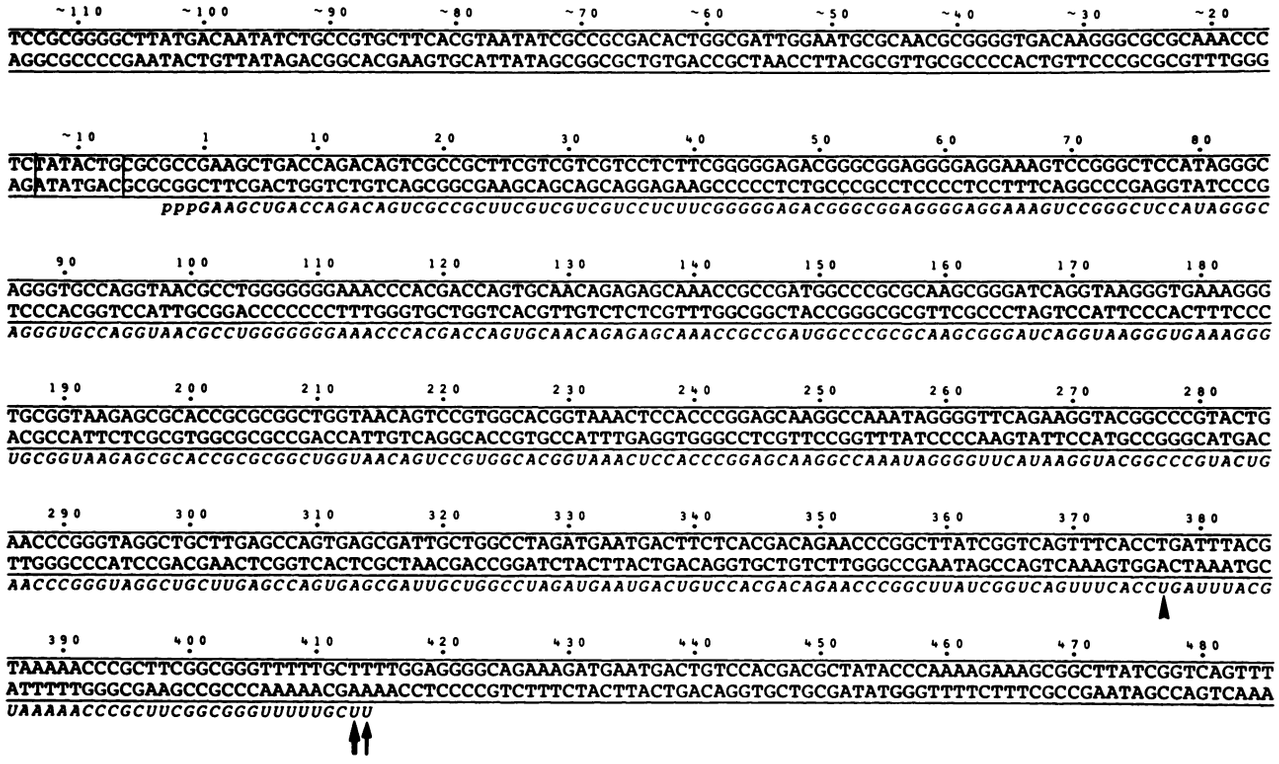

FIG. 5. Nucleotide sequences of the gene coding for the RNA component of RNase P and its transcripts. The nucleotide sequences of the transcription products are in italics starting with the 5′-terminal triphosphate. Arrows indicate the 3′ termini of the transcripts; the larger arrow represents about 65% of the chains, whereas the smaller arrow represents about 35% of the chains. The transcription start site is at position 1. The Pribnow box sequence in the promoter region is enclosed in a box. The arrowhead at position 377 indicates the 3′ end of the RNA component isolated from RNase P.

minal sequence of G-U-3′ as described above (Fig. 4 *b* and *c*), the unidentified base X in the oligonucleotide must be uracil. These results show that the 3′-terminal sequences of the transcripts are 5′ U-U-U-U-U-G-C-U (major) and 5′ U-U-U-U-U-G-C-U-U (minor). A DNA sequence (5′ T-T-T-T-T-G-C-T-) corresponding to the major oligonucleotide sequence is present in the gene 406–413 base pairs downstream from the transcription start site. This is the only DNA sequence that is consistent with the 3′-terminal sequence of the transcript. On the basis of these results, we conclude that transcription terminates within a region of two bases at positions 413 and 414 from the transcription start site as shown in Fig. 5.

The termination sites are immediately preceded by a sequence of 28 nucleotides (between positions 385 and 412 in the figure) that could form a stem-and-loop structure of 13 base pairs in the transcripts as shown in Fig. 6. The presence of this potential secondary structure in the transcripts was supported by the following experimental results. When the major transcripts labeled at the 3′ end were subjected to partial digestion with RNase T1 or RNase U2, and the digests were electrophoresed on a 20% polyacrylamide gel, the region corresponding to the potential secondary structure was relatively resistant to the nucleases and very few partial digestion products were detected in the gel, although there were eight guanines and five adenines in the region (data not shown). These results strongly suggest that the region of the transcripts may form a secondary structure in solution as predicted from the sequence analysis. This is consistent with the consensus structural feature of ρ-independent termination sites (14). The complete nucleotide sequence of the transcript of the gene for the RNA component of RNase P is shown in Fig. 5.

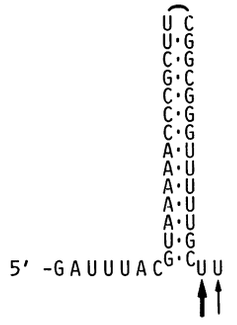

FIG. 6. Stem-and-loop structure near the transcription termination site. The arrows indicate the 3′ termini of the transcripts shown in Fig. 5.

***In Vitro* Processing of the Transcript of the Gene for the RNA Component of RNase P.** The transcript of the gene is 413 or 414 nucleotides long. On the other hand, the RNA component isolated from RNase P is 377 nucleotides long as will be described elsewhere. Comparison of the two RNA sequences has revealed that the two RNAs have the same 5′-terminal sequence, but the transcript carries an extra stretch of 36 or 37 nucleotides at the 3′ end of the RNA component from the enzyme as shown in Fig. 5. It appears, therefore, that the transcript is processed at the specific site to form the mature RNA component.

This idea is supported by the following experiment. The *Hinc*2.8kb fragment was transcribed in the presence of [α-^{32}P]-ATP, and the major transcripts were purified by polyacrylamide gel electrophoresis. When the transcripts were incubated with an S30 extract and then electrophoresed on 8% polyacrylamide gel/8 M urea, they were converted to a RNA species that had

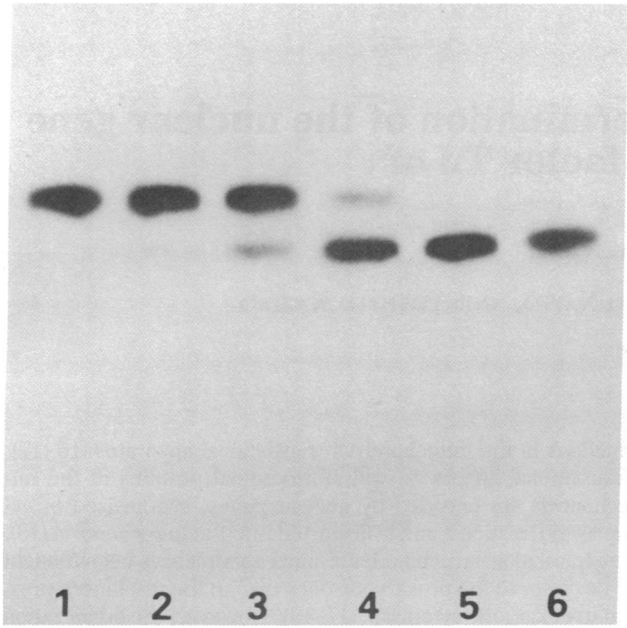

FIG. 7. *In vitro* processing of the transcripts of the gene for the RNA component of RNase P. The *Hinc*2.8kb fragment was transcribed in the presence of [α-^{32}P]ATP as the radioactive substrate, and the major transcript was purified by gel electrophoresis as in Fig. 1. The RNA (approximately 2 fmol, 1×10^4 cpm) was incubated at 37°C with the S30 extract. Incubation times (min) were as follows: 0 (lane 1), 10 (lane 2), 30 (lane 3), 60 (lane 4), and 120 (lane 5). After incubation, RNA was extracted with phenol and electrophoresed on 8% polyacrylamide gel/8 M urea. Migration is from top to bottom. The RNA component isolated from purified RNase P and end-labeled at its 5′ terminus with [γ-^{32}P]ATP by T4 polynucleotide kinase was electrophoresed on the same gel (lane 6).

the same electrophoretic mobility as the RNA component of RNase P (Fig. 7). When the same experiment was performed with transcripts labeled with [γ-^{32}P]GTP, the product still contained radioactivity, indicating that the 5′ end of the transcript was intact (data not shown). A preliminary experiment showed that the product has the 3′-terminal sequence indistinguishable from that of the RNA component from RNase P. The details of the *in vitro* processing reaction will be described elsewhere. The results show that there is a nuclease activity in *E. coli* that cleaves the transcript of the gene to form the RNA component of the enzyme.

DISCUSSION

We have identified the major transcription product of the gene coding for the RNA component of *E. coli* RNase P. From the terminal sequences of the RNA product, the sites for transcription initiation and termination were mapped on the DNA sequence of the gene previously determined and the complete nucleotide sequence of the transcript was determined. Structural analysis of the transcription products has revealed the signals for transcription initiation and termination of the gene. The transcription start site indicates that the region including the Pribnow box sequence (T-A-T-A-C-T-G) located 7–13 base pairs upstream from the start site and the sequence (G-T-G-A-C-A) in the −35 region represent the promoter of the gene. The GC-rich sequence (C-G-C-G-C-C) between the Pribnow box sequence and the transcription start site appears to be related to the stringent control of the gene reported by Reed *et al.* (6) as are the cases with tRNA genes and rRNA genes in *E. coli* (7, 16, 17). Transcription of the gene coding for the RNA component of RNase P terminates at two different sites within the 5′ T-T sequence, which is located 413 or 414 base pairs downstream from the start site. The T-T sequence is followed by two additional thymidines and preceded by a sequence of 28 nucleotides that could form a stem-and-loop structure of 13 base pairs. This is consistent with the observation that transcription termination of the gene is ρ independent (14). Among the two termination sites in the 5′ T-T sequence, the 5′ proximal thymidine is the preferred position, where more than 65% of the chains terminate.

The sequence analysis of the transcription products has also revealed that the transcripts carry extra sequences of 36 and 37 nucleotides at the 3′ end of the RNA component isolated from RNase P. It is highly likely that the RNA of the enzyme is formed from the transcript by some processing event. In fact, the *E. coli* extract appears to contain a nuclease activity that cleaves the transcripts to generate the 3′ end of the RNA component. At present, the nature of the nuclease activity is not understood. It also remains to be elucidated whether the processing of the transcripts is a prerequisite for assembly of the protein and RNA components of the enzyme. In any case, it seems that we have encountered an interesting situation; that is, the RNA component of a RNA processing enzyme itself is a product of RNA processing. Apparently, more work is needed to clarify the nature of this processing reaction.

We thank Dr. H. Ozeki for his valuable comments and discussions. We also thank Dr. M. Imai of Kyoto University for the gift of ρ factor. This work was supported by a Grant-in-Aid for Special Research Project from the Ministry of Education, Science, and Culture of Japan.

1. Altman, S. & Smith, J. D. (1971) *Nature (London) New Biol.* **233**, 35–39.
2. Sakano, H. & Shimura, Y. (1978) *J. Mol. Biol.* **123**, 287–326.
3. Shimura, Y., Sakano, H., Kubokawa, S., Nagawa, F. & Ozeki, H. (1980) in *Transfer RNA: Biological Aspects*, eds. Söll, D., Schimmel, P. & Abelson, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 43–58.
4. Stark, B. C., Kole, R., Bowman, E. J. & Altman, S. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3717–3721.
5. Kole, R. & Altman, S. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3795–3799.
6. Reed, R. E., Baer, M. F., Guerrier-Takada, C., Donis-Keller, H. & Altman, S. (1982) *Cell* **30**, 627–636.
7. Nakajima, N., Ozeki, H. & Shimura, Y. (1982) *J. Biol. Chem.* **257**, 11113–11120.
8. Nakajima, N., Ozeki, H. & Shimura, Y. (1981) *Cell* **23**, 239–249.
9. Shigesada, K. & Imai, M. (1978) *J. Mol. Biol.* **123**, 467–486.
10. Cashel, M., Lazzarini, R. A. & Kalgacher, B. (1969) *J. Chromatogr.* **40**, 103–109.
11. Donis-Keller, H., Maxam, A. M. & Gilbert, W. (1977) *Nucleic Acids Res.* **4**, 2527–2538.
12. Donis-Keller, H. (1980) *Nucleic Acids Res.* **8**, 3133–3142.
13. Pribnow, D. (1975) *J. Mol. Biol.* **99**, 419–443.
14. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
15. Jay, E., Bambara, R., Padmanabhan, R. & Wu, R. (1974) *Nucleic Acids Res.* **1**, 331–353.
16. Travers, A. A. (1980) *J. Mol. Biol.* **141**, 91–97.
17. deBoer, H. A., Gilbert, S. F. & Nomura, M. (1979) *Cell* **17**, 201–209.