Sequence of the gene for isoleucine tRNA_I and the surrounding region in a ribosomal RNA operon of *Escherichia coli*

Takao Sekiya and Susumu Nishimura

National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo, Japan

Received 1 November 1978

ABSTRACT

A DNA fragment of about 2000 base pairs carrying the gene for tRNA_I^{Ile} has been cloned from a total Eco RI endonuclease digest of *Escherichia coli* DNA. Sequence analyses revealed that about the first 850 base pairs from one end of the fragment contain a nucleotide sequence corresponding to that in the 3'-end of 16S rRNA. The gene for tRNA_I^{Ile} follows the 16S rRNA gene and both genes flank a spacer sequence of 68 base pairs. The spacer region contains a repeating, a hair pin and a symmetrical structure when the sequence is viewed in the single stranded form. A notable hair pin structure is also observed in the region adjacent to the 3'-end of the tRNA_I^{Ile} gene. In addition, about 850 base pairs from the other end of the DNA fragment have been found to contain the nucleotide sequence of the 5'-end of 23S rRNA. The presence of the genes for tRNA_I^{Ile}, 16S and 23S rRNA and the hybridization to tRNA_I^{Ala} suggest that this cloned DNA is part of one of the *E. coli* rRNA operons carrying these two tRNA genes as a spacer.

INTRODUCTION

The primary sequences of most *Escherichia coli* tRNA species have been determined (1) and the genes for many tRNAs have also been mapped on the *E. coli* chromosome (2). However, little is known about the detailed gene structure beyond the tRNA sequence except in tRNA_I^{Tyr}. Sequence analyses of a transducing phage DNA carrying the gene for tRNA_I^{Tyr} have provided information on the fine structure of the gene including the regions for controlling gene expression, that is, the promoter and terminator (3,4). Other information so far obtained on the sequence of tRNA genes was obtained by analyses of precursor RNAs. Analyses of the precursors of tRNA₂^{Gly}-tRNA₃^{Thr} (5), tRNA₆^{Glu} (6), tRNA₆^{Phe} (6) and tRNA₇^{Asn} (7) provided information on the sequences of several nucleotides beyond those for the tRNAs.

For understanding the mechanism of gene expression in *E. coli*, our interest in the structures of tRNA genes prompted us to isolate *E. coli* DNA fragments carrying these genes. Fragments produced by restriction endonuclease Eco RI have been cloned using total *E. coli* tRNA as a probe. One

of the fragments isolated in this way was about 2000 base pairs long and it hybridized with purified tRNA₁^{Ile} and tRNA₁^{Ala}.

This paper reports sequence analysis of 1586 base pairs of this fragment by the method of Maxam and Gilbert (8). The first 850 base pairs from the one end of the fragment contain the 3'-end region of the 16S rRNA gene. The gene for tRNA₁^{Ile} follows this gene and these two genes flank a spacer region of 68 base pairs. The spacer region contains a repeating sequence, a hair pin and a symmetrical structure.

It is known that some *E. coli* rRNA operons (rrnA, rrnD and an unmapped operon) contain clusters of genes for 16S rRNA, tRNA₁^{Ile}, tRNA_{1B}^{Ala}, 23S rRNA and 5S rRNA (9), which are transcribed as unit (10-12). Our results indicate that the DNA fragment cloned and sequenced in this work is part of one of these clusters of genes.

MATERIALS AND METHODS

Restriction endonuclease Eco RI, Alu I, Hind III and Hind II were purchased from Boehringer Mannheim Biochemicals. Hinf I, Hae II, Hpa I and Hpa II were from Bethesda Research Laboratory, Inc. Endonuclease Hae III was a kind gift from Dr. B. Weisblum, University of Wisconsin. [³²P]Orthophosphate was purchased from New England Nuclear and [γ-³²P] ATP was prepared as described by Maxam and Gilbert (8). T₄-Polynucleotide kinase was a kind gift from Dr. H.G. Khorana, Massachusetts Institute of Technology and also from Dr. K. Takeishi, Saitama Cancer Center Research Institute. The enzyme was also purchased from Boehringer Mannheim Biochemicals. Bacterial alkaline phosphatase was purchased from Worthington Biochemicals.

Preparation of a DNA fragment containing the gene for E. coli tRNA₁^{Ile} (tDNA₁^{Ile}). The methods used for cloning *E. coli* C600 DNA fragments produced by restriction endonuclease Eco RI and detection of the plasmid DNAs carrying tRNA genes using [³²P] labeled total *E. coli* tRNA as a probe will be published elsewhere. One of the plasmids thus detected, pTS228, contained an *E. coli* DNA fragment of 2000 base pairs (tDNA₁^{Ile}). It was hybridized with tRNA₁^{Ile} and tRNA₁^{Ala} by Southern's blotting hybridization method (13) using tRNA species separated by two-dimensional polyacrylamide gel electrophoresis (14).

E. coli HB101 (15) and *E. coli* HB129 harboring pMB9 (16), the parent plasmid of pTS228, were supplied by Dr. M. Ryan, Massachusetts Institute of Technology. DNA of pTS228 was prepared by chloramphenicol amplification in

host *E. coli* HB101 as described by Clewell and Helinski (17). A sample of 6 mg of pTS228 DNA was digested with 1000 units of restriction endonuclease Eco RI in reaction mixture (total volume, 1 ml) containing 0.1 M Tris-HCl (pH 7.5)-50 mM NaCl-10 mM MgCl₂. The reaction was performed at 37° for 16 hrs and then the mixture was subjected to electrophoresis on 1 % agarose gel to separate tDNA₁^{Ile} from the vector DNA. The bands of DNA in the gel were stained with ethidium bromide (1 µg/ml) and located under ultraviolet light. The fragment was cut out of the gel and eluted electrophoretically as described by Yang *et al.* (18). The eluate was extracted twice with an equal volume of phenol and then NaCl was added to 0.3 M, and DNA was precipitated by addition of 2.5 volumes of ethanol. The final yield of tDNA₁^{Ile} was 155 µg.

Cleavage of fragment tDNA₁^{Ile} with restriction endonuclease Alu I or Hae III. A sample of 22 µg of tDNA₁^{Ile} was digested by incubation with 40 units of restriction endonuclease Alu I or Hae III in 6 mM Tris-HCl (pH 7.5)-6 mM MgCl₂-6 mM β-mercaptoethanol in a total volume of 200 µl for 3 hrs at 37°. The digest was subjected to electrophoresis on a polyacrylamide slab gel (40 cm) consisting of 20 % (10 cm), 10 % (10 cm) and 5 % (20 cm) gel (from the bottom). Electrophoresis was performed at 6 mA for 16 hrs using 45 mM Tris-45 mM boric acid-2 mM EDTA (pH 8.3) buffer. The bands of DNA were stained with ethidium bromide and located under ultraviolet light and the DNAs in individual bands were eluted from the gel as described by Maxam and Gilbert (8).

DNA sequencing. DNA sequencing was performed essentially as described by Maxam and Gilbert (8). The phosphate groups at the 5'-ends of double stranded restriction endonuclease fragments were removed by treatment with 5 µg/ml of bacterial alkaline phosphatase in 20 µl of 10 mM Tris-HCl (pH 7.5)-0.1 mM EDTA at 65° for 30 min. Then the phosphatase was inactivated by heating at 95° for 5 min in the presence of 5 mM EDTA, and the mixture was quickly chilled in ice and subjected to the T₄-polynucleotide kinase reaction. The 5'-ends of the double stranded DNA were labeled using [γ-³²P] ATP and the kinase as described by Maxam and Gilbert (8). Strand separation of the labeled DNA or further digestion of the DNA using a different restriction endonuclease were performed as described by Maxam and Gilbert (8). The resulting DNA carrying labeled phosphate at one end only was subjected to sequence analysis (8).

RESULTS

A plasmid carrying tRNA genes. DNA fragments produced by restriction endonuclease Eco RI were cloned using plasmid pMB9 as a vector. The resulting new plasmids carrying tRNA genes were detected by Southern's method of hybridization (13) using total *E. coli* tRNA labeled with [³²P] as a probe. About 5 % of the plasmids contained a *E. coli* DNA fragment hybridized with tRNA. Of these plasmids, plasmid pTS228 described in this paper hybridized the most extensively with total tRNA and contained the smallest DNA fragment of *E. coli*.

When hybridization was carried out using purified [³²P] labeled *E. coli* tRNA species, obtained by two-dimensional polyacrylamide gel electrophoresis as described by Ikemura and Dahlberg (14), tRNA₁^{Ile} and tRNA₁^{Ala} were the major species hybridized to pTS228 (data not shown). The *E. coli* DNA fragment integrated into pTS228 and carrying these tRNA genes was designated as tDNA₁^{Ile}.

Isolation of the DNA fragment (tDNA₁^{Ile}) from the plasmid and its cleavage by restriction enzymes. When pTS228 (Fig. 1d) was digested with restriction endonuclease Eco RI, the fragment tDNA₁^{Ile} was formed, and it was separated from the vector pMB9 (Fig. 1a and b) by electrophoresis on 1 % agarose gel (Fig. 1c). From its mobility relative to those of marker DNAs of known sizes, tDNA₁^{Ile} was estimated as contain about 2000 base pairs (Fig. 1e).

Fragment tDNA₁^{Ile} was prepared in a large scale and treated with various restriction endonucleases, and the digests were analyzed by polyacrylamide gel electrophoresis. Alu I generated 11 fragments (A to K in Fig. 2). Alu A was a mixture of two fragments A1 and A2 with the same chain length of 400 base pairs. Upon treatment of Alu A with Hind II, only one of these, Alu A1, was further digested. The chain lengths of Alu B, C, D, E, F, G, H and I were estimated as about 340, 220, 185, 175, 120, 100, 50 and 45 base pairs, respectively. The total length calculated from the lengths of Alu A1 to I was 2045 base pairs, which is very close to the size of tDNA₁^{Ile} estimated as described above. The other fragments, Alu J and K, were smaller than 30 base pairs.

Restriction map of tDNA₁^{Ile}. When tDNA₁^{Ile} carrying [³²P] phosphates at the 5' termini was digested with Hind II, two fragments, Hind II A (about 230 bases) and Hind II B (about 150 bases), were generated as labeled DNAs (see Fig. 3). Sequence determination of these terminal fragments together

Fig. 1. Reisolation of the cloned *E. coli* DNA fragment ($tDNA_{I}^{Ile}$) from pTS228. One μ g of pTS228 or pMB9 was digested with 5 units of Eco RI in the reaction mixture (20 μ l) described under Materials and Methods. After incubation at 37° for 1 hr, the mixture was directly subjected to electrophoresis on 1 % agarose gel. Electrophoresis was performed at 15 mA for 16 hrs with 90 mM Tris-90 mM boric acid-4 mM EDTA as buffer. Lambda DNA (5 μ g) was digested with 10 units of Eco RI and Hind III in solution (20 μ l) containing 10 mM Tris-HCl (pH 7.5)-50 mM NaCl-10 mM $MgCl_2$ -14 mM DTT. The reaction was performed at 37° for 1 hr and the digestion products were used as marker DNAs, the chain lengths of which were taken from Allet and Bukhari (19). The DNAs analyzed are, (a) pMB9 treated with Eco RI, (b) pMB9, (c) pTS228 treated with Eco RI, (d) pTS228 and (e) λ DNA treated with Eco RI and Hind III.

with Alu I fragments revealed that the one end of Alu A1 was located at the 41st base pair from the labeled end of Hind II A (Fig. 3, also see Fig. 7). On the other hand, the labeled end of Hind II B was identical to that of one end of Alu D (Fig. 3, also see Fig. 6). These results clearly indicate that Alu A1 was very close to the left end of $tDNA_{I}^{Ile}$ and Alu D was the right end fragment, when the sequence of $tDNA_{I}^{Ile}$ is illustrated as in Fig. 3. In the sequence of 40 base pairs from the labeled end of Hind II A, that is, the left very end of $tDNA_{I}^{Ile}$, there was one Alu I site and so two Alu I fragments of 23 and 17 base pairs were formed (see Fig. 7). From their sizes, these

Fig. 2. Digestion of the cloned *E. coli* DNA fragment ($\text{tDNA}_{\text{I}}^{\text{Ile}}$) with restriction endonuclease Alu I. Fragment $\text{tDNA}_{\text{I}}^{\text{Ile}}$ was digested with Alu I and the digestion products were separated by electrophoresis on polyacrylamide gel as described under Materials and Methods. DNA bands A to K are referred to as Alu A to Alu K in the text and their chain lengths were estimated from their mobilities relative to those of Hpa II digests of bacteriophage fd replicative form DNA (20).

two fragments seemed to correspond to the very small Alu I fragments J and K (Fig. 2), respectively. The order of the other Alu I fragments was determined by digestion of $\text{tDNA}_{\text{I}}^{\text{Ile}}$ with Hae III. One of the fragments produced by the endonuclease, Hae B (Fig. 3), was labeled with $[^{32}\text{P}]$ at its 5'-ends. Upon digestion with Alu I, Hae B produced two labeled 5'-fragments of

Fig. 3. Restriction map of the cloned *E. coli* DNA fragment (tDNA^{Ile}). Cleavage sites on the DNA fragment (tDNA^{Ile}) were mapped for Alu I (—→) and Hae III (—→). The region between Alu F and Alu A2 contains Alu G, H and I, although the order of these fragments has not been analyzed. Detailed map for Alu I in this region has been elucidated independently by Young *et al.* (21). The two Eco RI sites correspond to the termini of the cloned DNA. Digestion of the DNA with Hind II gave two terminal fragments, Hind II A and Hind II B. The locations of the genes for 16S rRNA, tRNA^{Ile}, and 23S rRNA are indicated by bold arrows, the direction of which corresponds to that of transcription of the genes.

about 60 and 50 base pairs. An identical labeled fragment of 60 base pairs was generated when [5'-³²P] Alu F was digested with Hae III, while a fragment of 50 base pairs was obtained by digestion of [5'-³²P] Alu B with Hae III. Alu I digestion of [5'-³²P] Hae B also generated an unlabeled internal fragment that had the same chain length of about 175 base pairs as Alu E. From these results the order was deduced to be Alu B-E-F as shown in Fig. 3. Similar analyses were performed on the other Hae III fragments. From the results on Hae C, D and E, together with those on Alu I fragments, Alu B, C and D, the order was concluded to be Alu D-C-B as shown in Fig. 3. Digestion of Hae A with Alu I produced Alu A1, A2, G and smaller fragments. As described below, sequence determination of Alu A1 and A2 revealed that part of the nucleotide sequence of the 5'-end of 23S rRNA overlapped these two fragments. Therefore, Alu A2 is adjacent to Alu A1. The region between Alu F and Alu A2 has not been analyzed further but the remaining fragments, Alu G, H and I should be located in this region. All these results and assumptions are summarized in Fig. 3.

Nucleotide sequence determination of Alu I fragments. The 5'-termini of each Alu I fragment were labeled with [^{32}P] phosphate and the strands were separated as described under "Materials and Methods". Sequence determination of these DNAs was performed by the method of Maxam and Gilbert (8). As an example, sequence analysis on Alu F, which carries the gene for tRNA^{Ile}₁, is shown in Fig. 4. The single stranded DNAs [5'- ^{32}P] Alu F1 and [5'- ^{32}P] Alu F2 were obtained by strand separation of [5'- ^{32}P] Alu F. The

Fig. 4. DNA sequencing gel. The double stranded DNA Alu F was labeled with [^{32}P] at its 5'-ends. Strand separation gave Alu F1 and F2 and these single stranded DNAs were subjected to the sequencing procedure of Maxam and Gilbert (8). The numbering is described in the legend to Fig. 5.

sequencing gel patterns for these single stranded DNAs are shown in Fig. 4. The nucleotide sequences TTCA (or G) and CCA that are common in tRNAs are observed in Alu F1 (Fig. 4, left) at nucleotide numbers 123 and 143, respectively (the numbering is as described in the legend to Fig. 5). The corresponding nucleotide sequences are found in the complementary strand Alu F2 (Fig. 4, right). The complete nucleotide sequence of Alu F (118 base pairs) thus determined is shown in Fig. 5. Fig. 5 also shows the sequence of Alu E (177 base pairs), Alu B (376 base pairs) and a part of Alu C (45 base pairs), which were determined similarly. The contiguous sequence of Alu B-

Fig. 5. DNA sequence of the region between the 3'-end of the 16S rRNA gene and the gene for tRNA^{Ile}. The sequences of Alu C, B, E and F are shown. Cleavage sites for the restriction endonucleases Alu I, Hae III, Hpa II, Hinf I and Hae II are shown by arrows. The 3'-end region of the 16S rRNA gene (up to nucleotide number 1547) and the gene for tRNA^{Ile} (nucleotides number 69 to 145) are indicated by solid boxes. Under the boxes, sequences of 16S rRNA (22) and tRNA^{Ile} (23) previously determined by the direct RNA sequencing method are shown in smaller print. The base pair at residue 1053 shown by dots did not appear on the sequencing gel, leaving a blank space. Numbering of the gene for 16S rRNA is based on the RNA sequence and the base pair numbered 1050 corresponds to the 1050th residue from the 5'-end of 16S rRNA (22). The gene for tRNA^{Ile} and the surrounding region is numbered from the base pair adjacent to the 3'-end of the 16S rRNA gene.

E-F was deduced from the sequence of Hae B, which contains the regions of junctions between Alu B, E and F. When [5'-³²P] Alu B, E and F were digested with various restriction endonucleases and the products were analyzed by electrophoresis on polyacrylamide gels, the recognition sites for Hpa II, Hae III, Hinf I and Hae II indicated in Fig. 5 were evident, and the results confirm the proposed DNA sequence.

In Alu C, B and E, the sequence up to nucleotide number 1547 (nucleotide number 1 is assumed to be the 5'-end of 16S rRNA) in the top strand corresponds to the 3'-end sequence of 16S rRNA reported by Ehresmann *et al.* (22), shown in smaller print under the DNA sequence in Fig. 5. The sequence observed between nucleotides number 69 and 145 in Fig. 5 contains that determined for *E. coli* tRNA^{Ile} by Yarus and Barrell (23). The spacer sequence flanked by these two genes contains 68 base pairs (nucleotides number 1 to 68 in Fig. 5).

The complete sequence of the right end fragment Alu D (186 base pairs) and part of Alu C (55 base pairs) adjacent to Alu D were also determined and are shown in Fig. 6. The sequence was further confirmed by cleavage of [5'-³²P] Alu D and C with Hpa II, Hae III, and Hind II, and their cleavage sites are indicated in Fig. 6. The contiguous sequence of Alu D-C contains that of 16S rRNA corresponding to a region between about 680 and 920 nucleotides from its 5'-end that has been determined by Ehresmann *et al.* (22) and is shown in Fig. 6 under the DNA sequence.

As described earlier, the fragment Hind II A contains the extreme left

Fig. 6. DNA sequence of the right end region of the cloned DNA (tDNA^{Ile}). The sequences of Alu D and C are shown. Cleavage sites for Eco RI, Alu I, Hae III, Hpa II and Hind II are shown by arrows. Under the DNA sequence, that of the corresponding region of 16S rRNA (22) is shown in smaller print. The numbering is as described in the legend to Fig. 5 and the base pair numbered 700 corresponds to the 700th residue from the 5'-end of the RNA (22). The base pairs at residues 740 and 800 shown by dots did not appear on the sequencing gels, leaving blank spaces.

end of tDNA₁^{Ile} and the one end of Alu A1 located in Hind II A (Fig. 3). The complete sequence of Alu A1 (441 base pairs) and the 40 base pairs at the extreme end of tDNA₁^{Ile} in Hind II A are shown in Fig. 7. Alu A2, which has the same chain length as Alu A1, has also been sequenced. When the sequences of about 100 base pairs from both ends of Alu A2 were elucidated, that of one end was found to contain a sequence corresponding to the 5'-end region of 23S rRNA determined by Branlant *et al.* (24), while the other end had no sequence corresponding to either rRNA or tRNA (data not shown). The sequence of the same region of 23S rRNA was also found in one end of Alu A1, suggesting that Alu A1 and A2 are contiguous, as shown in Fig. 7. In Fig. 7, the matched oligonucleotides in the sequence of the first 480 nucleotides from the 5'-end of 23S rRNA proposed by Branlant *et al.* (24) is shown under the DNA sequences of Alu A2 and A1 (nucleotides number 226 to 510).

Fig. 7. DNA sequence of the 5'-end region of the 23S rRNA gene. The DNA sequences of Alu A2, Alu A1 and Hind II A are shown. Cleavage sites for Alu I, Hae III, Hind II, Hpa I, Hpa II, Hae II and Hinf I are shown by arrows. Under the DNA sequence, the sequences of RNase T₁ oligonucleotides obtained from the 5'-end of 23S rRNA (24,25) are shown. The tentative nucleotide numbers are based on the sequence for the 5'-end region of 23S rRNA proposed by Branlant *et al.* (24) and the base pair numbered 240 corresponds to the 240th residue from the 5'-end of the proposed RNA sequence. The base pair at position 266 shown by dots did not appear on the sequencing gel, leaving a blank space.

Branlant and Ebel have also determined the nucleotide sequences of RNase T₁ digestion products containing more than one uridine residue of the 13S fragment from the 5'-end of 23S rRNA (25). We found 21 of these 43 fragments in the DNA sequence and these are shown under the corresponding regions of DNA in Fig. 7.

The presence of the sequences of 16S and 23S rRNA together with that of tRNA^{Ile}₁ in the DNA (Fig. 5,6 and 7) suggests that fragment tDNA^{Ile}₁ might be part of one of the three rRNA operons that are known to contain the genes for both tRNA^{Ile}₁ and tRNA^{Ala}_{1B} as a spacer between 16S and 23S rRNA genes (9). Hybridization of tDNA^{Ile}₁ with tRNA^{Ala}₁ suggests the presence of the gene for tRNA in the DNA fragment. Sequencing of the region between Alu F and Alu A2 that contains Alu G and other small Alu I fragments is not complete, and the gene for tRNA^{Ala}_{1B} together with a spacer between the tRNA and 23S genes may be in this region.

DISCUSSION

Using a cloned *E. coli* DNA fragment of about 2000 base pairs produced by restriction endonuclease Eco RI, the DNA sequences of the genes for tRNA^{Ile}₁, the 3'-end of 16S rRNA and the 5'-end of 23S rRNA have been determined. Of seven rRNA operons so far studied, three (rrnA, rrnD and an unmapped operon) are known to code both tRNA^{Ile}₁ and tRNA^{Ala}_{1B} together with 5S, 16S and 23S rRNA, while the others code tRNA^{Glu}₂ as a spacer tRNA (9). Since the DNA fragment described here can also hybridize with tRNA^{Ala}₁, it might be part of one of these three operons carrying the genes for both tRNA^{Ile}₁ and tRNA^{Ala}_{1B}. Sequence determination of the region of 500 base pairs between the genes for tRNA^{Ile}₁ and 23S rRNA has not been completed, and the tRNA^{Ala}_{1B} gene may be in this region.

The DNA sequence determined in this work shows (1) the structure of an rRNA operon, (2) nucleotide sequences for 16S and 23S rRNA and (3) interesting features in non-coding regions surrounding the gene for tRNA^{Ile}₁.

First, with respect to the structure of an rRNA operon-the gene order of 16S rRNA-(spacer tRNAs)-23S rRNA-5S rRNA for an operon carrying the gene for tRNA^{Ile}₁ has been proposed by Lund *et al.* (26). Our sequence provides direct evidence that the gene order in the operon is 16S rRNA-tRNA^{Ile}₁-tRNA^{Ala}_{1B}-23S rRNA. Furthermore, the finding of sequences corresponding to those of 16S rRNA, tRNA^{Ile}₁ and 23S rRNA in this order on the same strand of DNA gives further evidence of the direction of transcription on polycistronic rRNA operons; that is, 16S rRNA-(spacer tRNAs)-23S rRNA-5S rRNA (26).

Recently, Young and Steitz (27) and Young *et al.* (21) have determined the sequence of the 16S-23S spacer region in the *rrnD* operon that contains the genes for both tRNA₁^{Ile} and tRNA_{1B}^{Ala}. Morgan *et al.* (E.A. Morgan, T. Ikemura, L.E. Post and M. Nomura, personal communication) have also analyzed a similar region in another operon carrying the gene for tRNA₂^{Glu}. Our sequence for the tRNA₁^{Ile} gene and the surrounding regions (Fig. 5) is identical to that of Young *et al.* Moreover, their sequence covers the region between the genes for tRNA₁^{Ile} and 23S rRNA, which we have not yet analyzed completely, and suggests that the gene for tRNA_{1B}^{Ala} starts at nucleotide number 189 in our sequence shown in Figs. 5 and 8.

Second, with respect to the nucleotide sequence of rRNA—the sequence determined from those of Alu D, C, B and E cover about 750 nucleotides in the 3'-end region of 16S rRNA. The first 161 nucleotides from 3'-end are exactly the same as those reported by Young and Steitz (27). The sequence of other nucleotides corresponds fairly well to that proposed by Ehresmann *et al.*, who analyzed the RNA directly (22). Several ambiguities about the RNA sequence can be clarified by comparison with the DNA sequence determined (Fig. 5). Recently, Branlant *et al.* reported the sequence of the first 480 nucleotides for the 5'-end of 23S rRNA, determined by direct RNA sequencing (24). The sequence that we determined from those of Alu A2 and A1 (up to nucleotide number 510 in Fig. 7) contains all the RNase T₁ oligonucleotides found in the region between residues 230 and 480 in the RNA, although the alignment of the oligonucleotides in the RNA, somehow, differs from that in the sequence presented here. Discrepancy between our sequence and that by Branlant *et al.* (24) is also caused by lacking many of the nucleotides in the RNA which are present in the DNA sequence. Branlant and Ebel also analyzed a 13S fragment obtained from the 5'-end of 23S rRNA and determined the nucleotide sequences of RNase T₁ digestion products containing more than one uridine residue (25). Twenty one of these 43 oligonucleotides are found in our sequence of about 630 base pairs in Alu A2 and A1. Despite of considerable discrepancies, similarity to the 23S rRNA sequence suggests that the sequence shown in Fig. 7 could be part of the gene for 23S rRNA and correspond tentatively to nucleotide residues 226 to 854 from the 5'-end of 23S rRNA.

Finally, our sequence reveals several interesting features in the non-coding regions surrounding the gene for tRNA₁^{Ile}. The regions are very rich in A-T, compared with the tRNA₁^{Ile} gene, which contains 41 % of A-T pairs: the spacer of 68 base pairs between the genes for 16S rRNA and tRNA₁^{Ile}

contains 40 A-T pairs (59 %) and the region beyond the CCA-end of the tRNA^{Ile}₁ gene so far sequenced contains 64 % of A-T pairs. In the duplicated gene for tRNA^{Tyr}₁, the 3'-end of the first gene has been found to be followed by an A-T rich sequence of the intergenic spacer, and an A-T rich sequence in the terminator region is also adjacent to the 3'-end of the second gene (28, 29). These findings suggest that an A-T rich sequence may be important during transcription of genes or processing of transcripts.

Since the primary transcript 30S rRNA is known to contain spacer tRNAs together with 16S, 23S and 5S rRNA (11, 12), the sequence of the single stranded DNA carrying of 16S rRNA and tRNA^{Ile}₁ corresponds to part of the precursor RNA. Fig. 8 shows the possible secondary structure of the single stranded DNA carrying the sequence for tRNA^{Ile}₁ and surrounding regions. Three remarkable structures are observed in the spacer region of 68 nucleotides between the genes for 16S rRNA and tRNA^{Ile}₁. First, the pentanucleotide sequence of CCTTA at the 3'-end of 16S rRNA is repeated in the spacer region as indicated by parallel arrows. Second, including the two terminal nucleotides of the repeating unit described above, a structure composed of 5 base pairs and a 7 membered loop is possible in the spacer. During processing of 30S rRNA precursor, one of the sites for RNase III cleavage proposed by Young and Steitz (27) is between nucleotides number 33 and 34 in Fig. 8. Therefore, the biological function, if any, of these structures might be related to processing of the 3'-end of 16S rRNA. Third, the heptanucleotides AAAATGA and AGTAAAA, indicated by antiparallel arrows in Fig. 8, are symmetric in the region of 7 nucleotides upstream from the 5'-end of tRNA^{Ile}₁. The center of symmetry is between nucleotides number 53 and 54. A similar symmetric structure has been found in the tRNA^{Tyr}₁ precursor in the region of 10 nucleotides beyond the 3'-CCA end (Fig. 8 inset), although its biological function, if any, is unknown (28). More striking, is the fact that exactly the same heptanucleotide AGUAAAA is present in a similar region of 11 nucleotides upstream from the 5'-end of tRNA^{Tyr}₁ in its precursor shown in Fig. 8 inset (30). In consideration of the recent discovery of an RNA component in RNase P that is involved in processing of the 5'-end of tRNA (31), the fact that the heptanucleotide sequence is present 7 to 11 nucleotides from the 5'-ends of both tRNA^{Ile}₁ and tRNA^{Tyr}₁ in their precursors is very suggestive of the importance of this sequence in processing. The bold arrows in Fig. 8 indicate the sites for enzymatic cleavage of precursor RNA to expose the 5'- and 3'-ends of tRNA^{Ile}₁, probably by RNase P and a 3'→5' exonuclease, respectively (33). Adjacent to the 3'-end of tRNA^{Ile}₁

Fig. 8. Possible secondary structure of the single stranded DNA carrying the sequences for tRNA^{Ile} and the 3'-end of 16S rRNA. The sequences corresponding to 16S rRNA and tRNA^{Ile} are enclosed by continuous lines. The repeating units and the components of a symmetric structure are shown by thin arrows. The bold arrows indicate the cleavage sites for exposing the 3'- and 5'-ends of tRNA^{Ile}. The numbering is described in the legend to Fig. 5. *Inset*: The sequence of the precursor for tRNA^{Tyr}_{su3} (28-30). Since the gene for tRNA^{Tyr}_{su3} differs from that for tRNA^{Tyr}₃ by single base pair at the 5'-end of the anticodon, both genes and therefore the precursors for tRNA^{Tyr}_{su3} and tRNA^{Tyr}₃ are essentially identical. In the text these tRNAs are simply referred to as tRNA^{Tyr}₁. The heptanucleotide sequence identical to one of the components of the symmetric structure found in the region between 16S rRNA and tRNA^{Ile} gene is boxed. The bold arrow indicates the cleavage site for an endonuclease. The components of a symmetric structure are shown by thin arrows.

there is a hair pin structure composed of 21 nucleotides including the two terminal nucleotides of the tRNA. As shown in Fig. 8 inset, the precursor for tRNA^{Tyr}₁ has been found to contain a similar structure composed of 5 base pairs and 2 unpaired nucleotides on top of the hair pin in the region of two nucleotides beyond the 3'-CCA end (28,29). Two enzyme reactions have been suggested to be involved in processing of the 3'-end of tRNA^{Tyr}₁ (32); first an endonuclease cleaves the precursor RNA at the position between the 7th and 8th nucleotide beyond the 3'-CCA end (indicated by a bold arrow in Fig. 8 inset), where the 7th nucleotides is located at the 5th base pair

in the double stranded region and the 8th nucleotide is the first nucleotide in the unpaired region. The 7 extra nucleotides downstream from the CCA end in the resulting precursor RNA are then removed by the stepwise action of an exonuclease. If the 3'-end of tRNA₁^{Ile} is processed in a similar way, the endonuclease cleavage might occur in the hair pin structure at the position between nucleotides number 152 and 153 in Fig. 8. As with tRNA₁^{Tyr}, the resulting RNA carries 7 extra nucleotides at the 3'-end, which could be removed one by one by the exonuclease.

To understand the details of the mechanism of tRNA gene expression and processing of the transcripts, the findings discussed above must be compared with findings on other tRNA genes. We have cloned several other *E. coli* DNA fragments carrying different tRNA genes. Sequence analyses of these genes will provide more information on these problems.

After completion of this work, the complete nucleotide sequence of the 16S rRNA gene (1541 bases) from the *rrnB* cistron of *E. coli* and the revised sequence of the *E. coli* 16S rRNA (1542 bases) have been reported independently by Brosius *et al.* (34) and by Carbon *et al.* (35), respectively. Within their complete sequences, only nine residues are different. Our sequence of 3'-end region of 16S rRNA gene includes seven of these nine residues and our data for all of the seven residues completely agree with those of Brosius *et al.* rather than those of Carbon *et al.*

Including the seven residues mentioned above, our result of about 750 residues for the 3'-end of the 16S rRNA gene is in good agreement with that of Brosius *et al.* (34) except eight residues. When the sequence of the top strand of our DNA is compared with those of Brosius *et al.* (34) and Carbon *et al.* (35), our sequence lacks a G residue at position 859-860 and C residues at positions 1069-1070 and 1074-1075. Since these residues correspond to the third base pair from the cleavage sites for Alu I, lacking of these residues might be due to our technical difficulty of reading the sequence of first several nucleotides from the labeled 5'-end on a sequencing gel pattern. Two A-T pairs at residues 1156 and 1211 in our sequence are not present in others. Other discrepancies are C and G residues in our sequence at positions 1043 and 1228, respectively. Both of these are replaced by a T residue in the sequences reported by Brosius *et al.* (34) and by Carbon *et al.* (35). Our sequence belongs to *rrnA*, *rrnD* or an unmapped operon, while that of Brosius *et al.* (34) is from the *rrnB* operon. Besides technical errors, therefore, possible explanation for the single-base discrepancies would be cistronic heterogeneities, although the posi-

tional heterogeneities suggested by Ehresmann *et al.* (22) differ from the residues in question.

The base pairs at positions 740 and 800 shown by dot in our sequence are assigned as C-G and G-C, respectively, in the *rrnB* operon (34) and also in 16S rRNA (35). Because the corresponding nucleotides do not appear on our sequencing gel patterns and leave blank spaces, C residues of these pair must have methylated bases (36).

ACKNOWLEDGEMENTS

We are grateful to Dr. A. Maxam for a detailed protocol of the DNA sequencing method and to Drs. J. Steitz and M. Nomura for information on unpublished observations. We also thank Dr. Y. Tanaka and Miss Y. Kanemaru for their technical assistance. This work was partly supported by Grants-in-Aid for Cancer Research and Scientific Research from the Ministry of Education, Science and Culture, Japan.

REFERENCES

1. Sprinzl, M., Grütter, F. and Gauss, D.H. (1978) *Nucleic Acids Res.* 5, r15-r26.
2. Ikemura, T. and Ozeki, H. (1977) *J. Mol. Biol.* 117, 419-446.
3. Sekiya, T., Gait, M.J., Noris, K., Rammamoorthy, B. and Khorana, H.G. (1976) *J. Biol. Chem.* 251, 4481-4489.
4. Küpper, H., Sekiya, T., Rosenberg, M., Egan, J. and Landy, A. (1978) *Nature* 272, 423-428.
5. Chang, S. and Carbon, J. (1975) *J. Biol. Chem.* 250, 5542-5555.
6. Vögeli, G., Grosjean, H. and Söll, D. (1975) *Proc. Natl. Acad. Sci. USA* 72, 4790-4794.
7. Vögeli, G., Stewart, T.S., McCutchan, T. and Söll, D. (1977) *J. Biol. Chem.* 252, 2311-2318.
8. Maxam, A. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
9. Morgan, E.A., Ikemura, T. and Nomura, M. (1977) *Proc. Natl. Acad. Sci. USA* 74, 2710-2714.
10. Ginsburg, D. and Steitz, J.A. (1975) *J. Biol. Chem.* 250, 5647-5654.
11. Lund, E. and Dahlberg, J.E. (1977) *Cell* 11, 247-262.
12. Ikemura, T. and Nomura, M. (1977) *Cell* 11, 779-793.
13. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
14. Ikemura, T. and Dahlberg, J.E. (1973) *J. Biol. Chem.* 248, 5024-5032.
15. Boyer, H.W. and Roulland-Dussoix, D. (1969) *J. Mol. Biol.* 41, 459-472.
16. Boliver, F., Rodriguez, R.L., Betlach, M.C. and Boyer, H.W. (1977) *Gene* 2, 75-93.
17. Clewell, D.B. and Helinski, D.R. (1970) *Biochem.* 9, 4428-4439.
18. Yang, R.C.A., Van de Voorde, A. and Fiers, W. (1976) *Eur. J. Biochem.* 61, 101-117.
19. Allet, B. and Bukhari, A.I. (1975) *J. Mol. Biol.* 92, 529-540.
20. Takanami, M., Okamoto, T., Sugimoto, K. and Sugisaki, H. (1975) *J. Mol. Biol.* 95, 21-31.
21. Young, R.A., Macklis, R. and Steitz, J.A. (1979) *J. Biol. Chem.*, in press.

22. Ehresmann, C., Stiegler, P., Carbon, P. and Ebel, J.P. (1977) *FEBS Letters* 84, 337-341.
23. Yarus, M. and Barrell, B.G. (1971) *Biochem. Biophys. Res. Comm.* 43, 729-734.
24. Branlant, C., Sri Widada, A., Krol, A. and Ebel, J.P. (1977) *Eur. J. Biochem.* 74, 155-170.
25. Branlant, C. and Ebel, J.P. (1977) *J. Mol. Biol.* 111, 215-256.
26. Lund, E., Dahlberg, J.E., Kindahl, L., Jaskunas, S.R., Dennis, P.P. and Nomura, M. (1976) *Cell* 7, 165-177.
27. Young, R.A. and Steitz, J.A. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3593-3597.
28. Loewen, P.C., Sekiya, T. and Khorana, H.G. (1974) *J. Biol. Chem.* 249, 217-226.
29. Sekiya, T., Contreras, R., Küpper, H., Landy, A. and Khorana, H.G. (1976) *J. Biol. Chem.* 251, 5124-5140.
30. Altman, S. and Smith, J.D. (1971) *Nature N.B.* 233, 35-39.
31. Stark, B.C., Kole, R., Bowman, E.J. and Altman, S. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3717-3721.
32. Sekiya, T., Contreras, R., Takeya, T. and Khorana, H.G. (1978) *J. Biol. Chem.* in press.
33. Perry, R.P. (1976) *Ann. Rev. Biochem.* 45, 605-629.
34. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.E. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4801-4805.
35. Carbon, P., Ehresmann, C., Ehresmann, B. and Ebel, J.P. (1978) *FEBS Letters* 94, 152-156.
36. Ohmori, H., Tomizawa, J. and Maxam, A.H. (1978) *Nucleic Acids Res.* 5, 1479-1485.