Proc. Natl. Acad. Sci. USA
Vol. 75, No. 10, pp 4801–4805, October 1978
Biochemistry

Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*

(recombinant plasmids/DNA sequence analysis/*rrnB* cistron)

JÜRGEN BROSIUS, MARGARET L. PALMER, POINDEXTER J. KENNEDY, AND HARRY F. NOLLER

Thimann Laboratories, University of California, Santa Cruz, California 95064

Communicated by Robert L. Sinsheimer, July 26, 1978

ABSTRACT **The complete nucleotide sequence of the 16S RNA gene from the *rrnB* cistron of *Escherichia coli* has been determined by using three rapid DNA sequencing methods. Nearly all of the structure has been confirmed by two to six independent sequence determinations on both DNA strands. The length of the 16S rRNA chain inferred from the DNA sequence is 1541 nucleotides, in close agreement with previous estimates. We note discrepancies between this sequence and the most recent version of it reported from direct RNA sequencing [Ehresmann, C., Stiegler, P., Carbon, P. & Ebel, J. P. (1977) *FEBS Lett.* 84, 337–341]. A few of these may be explained by heterogeneity among 16S rRNA sequences from different cistrons. No nucleotide sequences were found in the 16S rRNA gene that cannot be reconciled with RNase digestion products of mature 16S rRNA.**

rRNA is becoming increasingly important in our current perception of the mechanism of action of ribosomes. In particular, we wish to understand more fully the functional role of 16S rRNA, the major molecular component of the small ribosomal subunit of prokaryotes. 16S rRNA has been directly implicated in discrimination of mRNA initiation sites (1, 2), tRNA binding (3, 4), and association of the two ribosomal subunits (5, 6). Full understanding of the workings of the ribosome consequently are becoming limited by our lack of knowledge of rRNA structure. Such information will also be essential for elucidation of the process by which these complex structures assemble themselves. Finally, we expect to gain insight into such diverse problems as protein–nucleic acid recognition and the evolutionary origin of the coding process from a more thorough knowledge of rRNA structure.

Partial nucleotide sequences for 16S RNA have been published (7–9). However, numerous discrepancies in oligonucleotide sequences (6, 10, 11) and in the ordering of oligonucleotides (6, 12–14) have been reported by other investigators. Additional evidence for sequence errors comes from restriction endonuclease mapping of a 16S RNA gene: some cleavage sites predicted from the published sequences were not found (unpublished data). Finally, chemical (6, 12, 15) and enzymatic (7, 8, 16) probes have been used to detect single-stranded regions of 16S RNA; the lack of agreement of these findings with the proposed secondary structure derived from the published sequences further suggests error in the primary structure. In view of the possible misinterpretation of various biochemical studies involving 16S RNA, we were prompted to reinvestigate its primary structure.

The availability of rapid DNA sequence methods made possible the derivation of the nucleotide sequence of 16S RNA by direct sequencing of the 16S rRNA gene from the *rrnB* cistron of *Escherichia coli*. We present here the complete sequence of the portion of the cistron corresponding to mature 16S RNA. Reported discrepancies with the previously published sequence have been confirmed, and additional errors have been found involving oligonucleotide sequences, ordering of oligonucleotides, and, in one instance, the location of a larger section of the primary structure. No nucleotide sequences were found that cannot be accounted for from the RNase digestion products of mature 16S rRNA.

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.

METHODS

Cloning and Mapping of DNA. The 16S rRNA gene from the *rrnB* cistron of *E. coli* was cloned from two *Eco*RI restriction fragments of λ*rif*^{d}18 (17, 18) in the ColE1 plasmid vector. Determination of the location of the 16S rRNA sequences and restriction enzyme cleavage sites will be described elsewhere. The small *Hin*dIII fragment from pER24 was excised and reinserted into the vehicle pBR322 (19) to give the recombinant plasmid pKK115 (see Fig. 1). All recombinant DNA experiments were carried out under P1, EK1 conditions, as specified in the National Institutes of Health guidelines.

Generation of DNA Restriction Fragments. Plasmid DNA was isolated and digested on a milligram scale with *Eco*RI or *Hin*dIII, and the inserted sequence was resolved from the cloning vector by sucrose gradient centrifugation (20). The purification of restriction enzymes *Alu* I, *Hpa* II, *Eco*RI, *Sma* I, *Bgl* II, and *Hin*dIII will be described elsewhere; *Hae* III, *Hha* I, *Hin*fI, and *Mbo* II were purchased from New England Biolabs; *Ava* II and *Taq* I were from Bethesda Research Laboratories. The cloned fragments (20–40 pmol) were digested with the appropriate restriction endonuclease to obtain DNA fragments of suitable size for sequence determination and isolated as described below.

Sequencing by Primed Synthesis with DNA Polymerase. For sequencing by either of the two methods developed by Sanger and coworkers (21, 22), restriction fragment primers were isolated by electrophoresis on 1% agarose gels (20 × 20 × 0.3 cm) in 0.02 M Na acetate/0.04 M Tris acetate, pH 8.3/2 mM EDTA or on 8% acrylamide/0.26% bisacrylamide gels (20 × 40 × 0.15 cm) in 0.09 M Tris borate, pH 8.3/2.5 mM EDTA. DNA bands were located by ethidium bromide staining, excised, crushed, and eluted by shaking overnight at room temperature in 0.5 M NaCl/0.1 M Tris·HCl, pH 8.0/5 mM EDTA and recovered by ethanol precipitation. Template strands were prepared from pER18 DNA by digestion with *Bgl* II, which cleaves at a single site in this plasmid, followed by cesium chloride density gradient centrifugation after heating and quick cooling in the presence of poly(U,G) (Miles) as described (23). Deoxynucleoside triphosphates labeled in the α position with ^{32}P (200–300 Ci/mmol) were obtained from New England Nuclear. Dideoxynucleoside triphosphates were obtained from P-L Biochemicals; DNA polymerase I and its proteolytic fragment lacking 5′ exonuclease activity were from Boehringer. Sequence determinations by the "plus and minus" method or by the "terminator" method were carried out as described (21, 22).

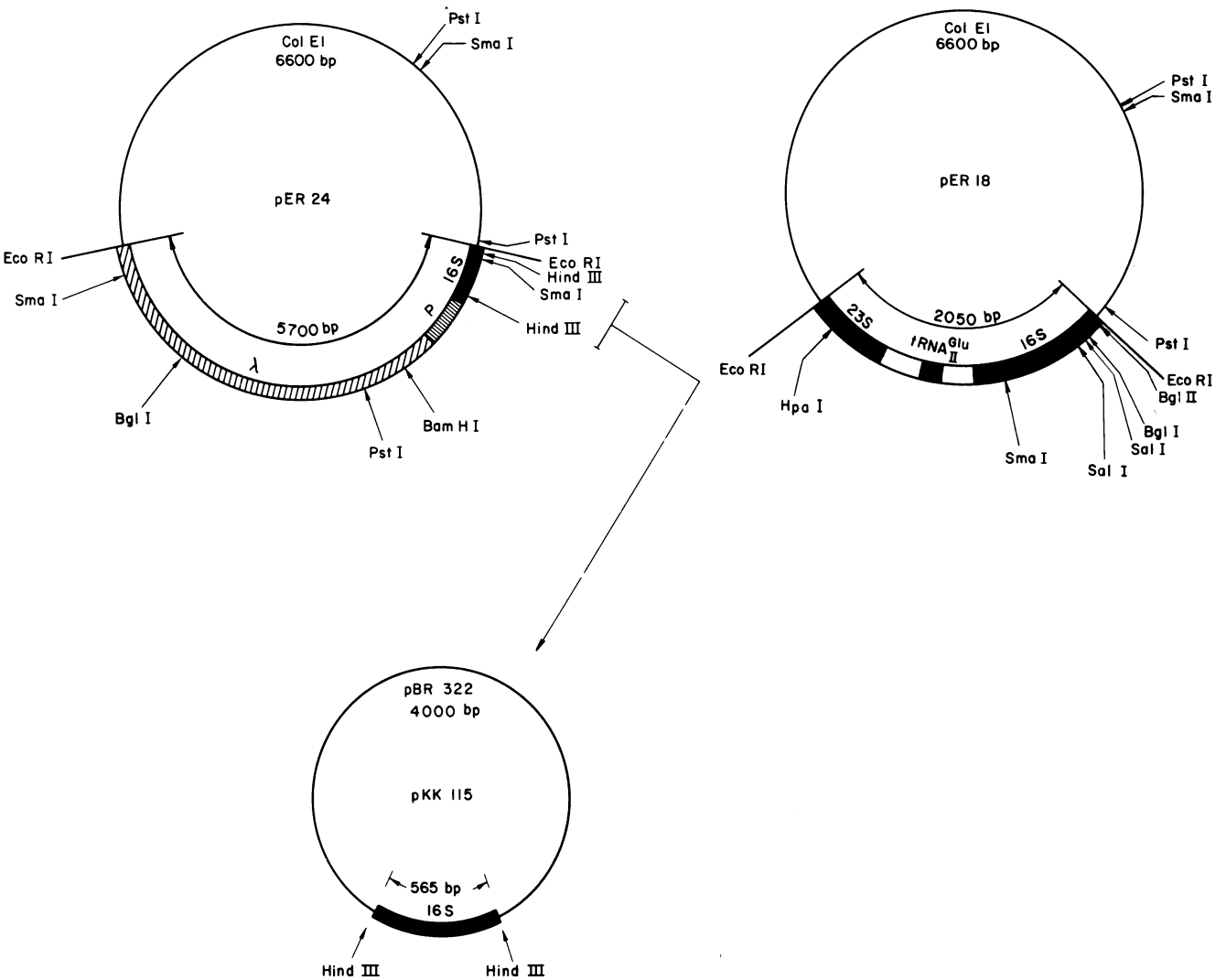

FIG. 1. Physical maps of recombinant plasmids containing portions of the 16S rRNA gene from the *rrnB* cistron of *E. coli*. The positions of 16S and 23S rRNA and tRNA sequences are shown by black segments. Sites of cleavage by restriction endonucleases are denoted by arrows. Wild-type λ sequences are shown by hatched segments, and the approximate position of the promoter for the *rrnB* cistron is indicated by the letter P. bp, base pairs.

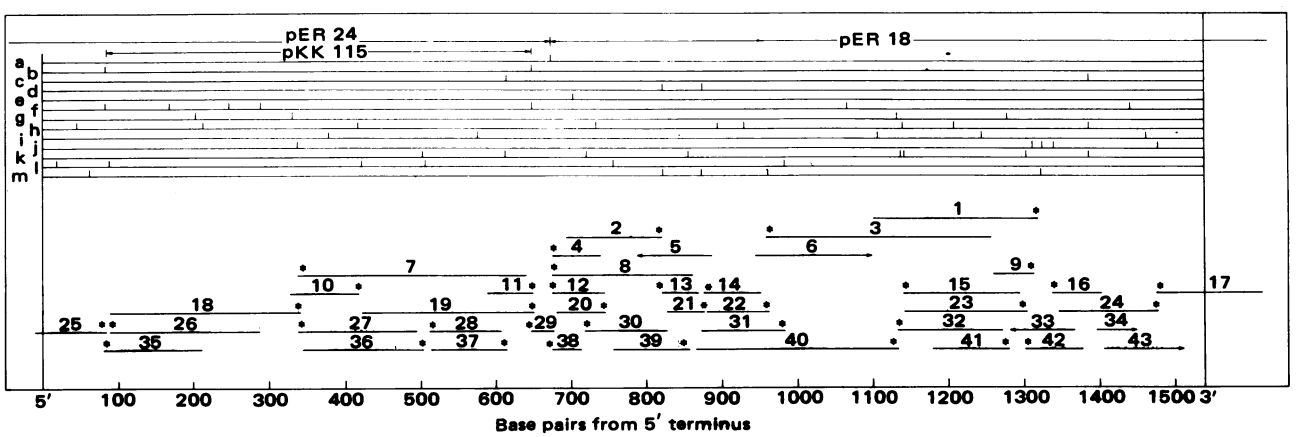

FIG. 2. Schematic map of the 16S rRNA gene, showing the detailed positions of restriction endonuclease cleavages, and the DNA fragments actually used in the sequence determination. Sequences obtained by primed synthesis (21, 22) are indicated by horizontal bars ending in arrows showing the direction of the chain extension. Sequences derived from partial chemical modification (24) are shown with asterisks at the 5′ end of the labeled chain. Restriction endonucleases: a, *Eco*RI; b, *Hin*dIII; c, *Sma* I; d, *Sal* I; e, *Bgl* II; f, *Alu* I; g, *Ava* II; h, *Hae* III; i, *Hha* I; j, *Hin*fI; k, *Hpa* II; l, *Mbo* II; m, *Taq* I. DNA fragments: 1, *Taq* I-3f; 2, *Taq* I-7f; 3, *Taq* I-3s; 4, *Alu* I-4f; 5, *Hae* III-13/L; 6, *Hae* III-13/H; 7, *Hin*fI-1s; 8, *Hin*fI-1s; 9, *Hin*fI-1f; 10, *Hae* III-2f; 11, *Hha* I-3f; 12, *Mbo* II-12s; 13, *Taq* I-10s; 14, *Taq* I-9f; 15, *Hpa* II-4f; 16, *Hin*fI-7f; 17, *Hin*fI-3f; 18, *Hin*fI-2f; 19, *Hae* III-1f; 20, *Mbo* II-12f; 21, *Taq* I-10f; 22, *Taq* I-9s; 23, *Hpa* II-4s; 24, *Hin*fI-7s; 25, *Hin*dIII/*Bam*HI-2; 26, *Hin*fI-2s; 27, *Hin*fI/*Hpa* II-2; 28, *Hpa* II-2f; 29, *Hin*dIII/*Bam*HI-1; 30, *Hpa* II-6s; 31, *Mbo* II-5s; 32, *Ava* II-4s; 33, *Sma* I/*Eco*RI-1/L; 34, *Sma* I/*Eco*RI-2/H; 35, *Hae* III-3f; 36, *Hpa* II/*Hin*fI-2; 37, *Hpa* II-2s; 38, *Hpa* II-13f; 39, *Hpa* II-6f; 40, *Ava* II/*Hpa* II-2; 41, *Ava* II-4f; 42, *Hpa* II-9f; 43, *Sma* I/*Eco*RI-2/H. The portions of the sequence covered by the plasmid DNA molecules in Fig. 1 are also shown.

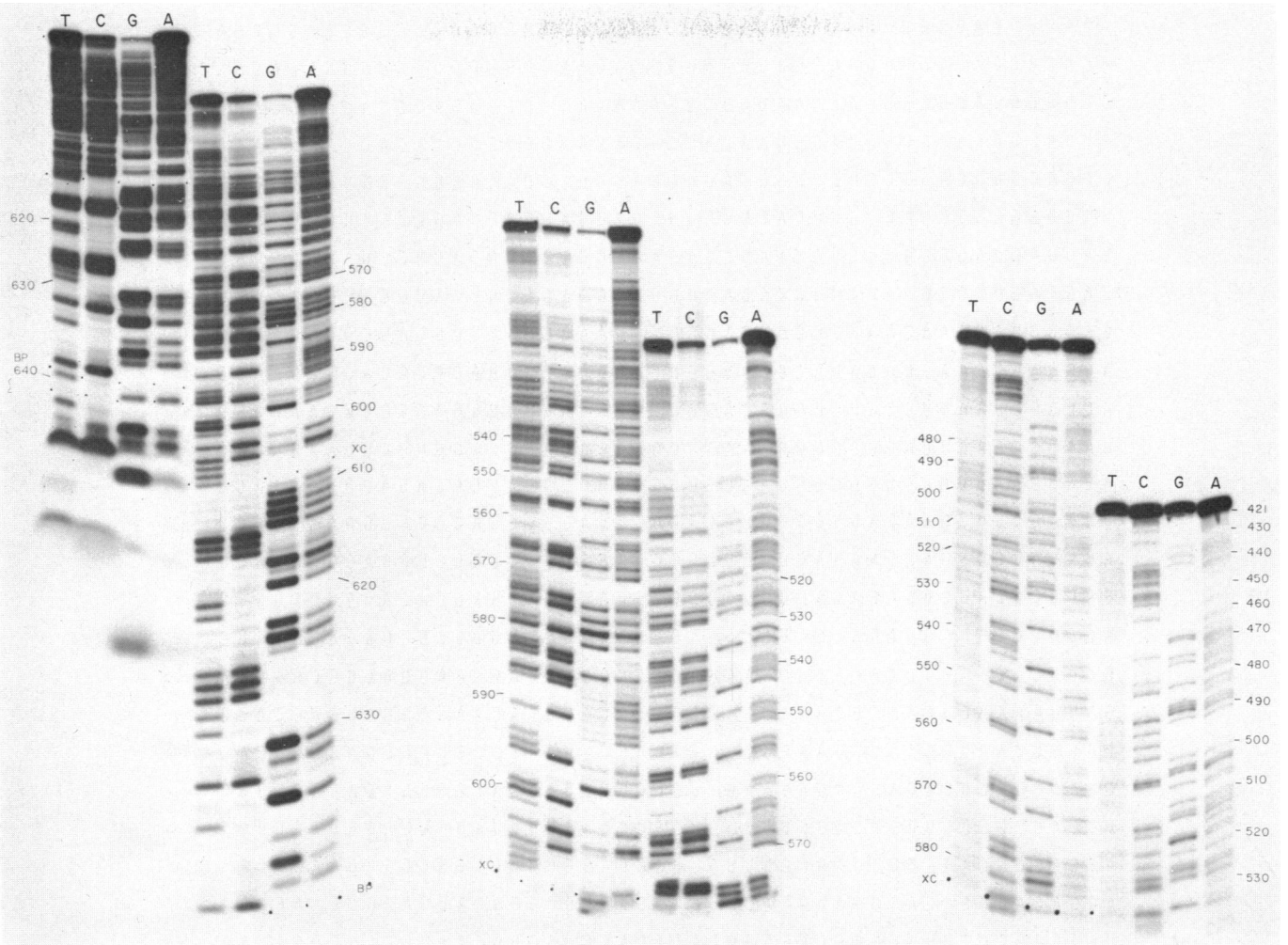

FIG. 3. Autoradiographs of sequence gels from the partial chemical modification method applied to the single-stranded DNA fragment *Hae* III-1f (Fig. 2). The sequence read from the gel is complementary to the 16S RNA sequence and covers the region from positions 421 to 650. The positions of nucleotides in the 16S RNA sequence are indicated by the numbers. The gels (0.7 × 20 × 40 cm) were 12% polyacrylamide/0.6% bisacrylamide run 1, 3, 9, and 12 hr, respectively, at 1000 V (*Left* and *Center*) and 8% acrylamide/0.4% bisacrylamide run for 7 and 10 hr, respectively, at 1000 V (*Right*).

Sequencing by the Partial Chemical Modification Method. For sequence determination according to the method of Maxam and Gilbert (24), the mixture of restriction fragments (20–40 pmol) was treated with bacterial alkaline phosphatase (1 μg/5 μg of DNA) for 1 hr at 37°, extracted four times with an equal volume of phenol saturated with 10 mM Tris·HCl, pH 8.0/1 mM EDTA, and precipitated with ethanol. The fragments were then labeled at their 5′ termini by using 5–10 units of polynucleotide kinase (P-L Biochemicals) and 0.5–2 mCi of [^{32}P]ATP (1500–2000 Ci/mmol) synthesized as described (25). Excess ATP and phosphate were removed by passing the reaction mixture through a column of Sephadex G-75 in 2.5 mM Tris·HCl/1 mM NaCl, pH 7.4, poured in a 1-ml disposable pipet. The peak containing the labeled DNA was lyophilized and taken up in 50 μl of 0.02 M Tris borate, pH 8.3/0.5 mM EDTA, and the labeled restriction fragments were resolved by electrophoresis on an 8% polyacrylamide gel (0.5 × 20 × 40 cm) as described above. DNA fragments were located by radioautography and eluted as described above. Singly end-labeled DNA was obtained by strand separation on 8% acrylamide/0.13% bisacrylamide gels (0.5 × 20 × 40 cm) (26) after the DNA sample was dissolved in 0.3 M NaOH/1 mM EDTA/10% (vol/vol) glycerol/0.05% xylene cyanol/0.05% bromophenol blue or by recleavage of the double-stranded DNA with a second restriction endonuclease.

Chemical treatment of the end-labeled DNA was performed essentially as described (24) except that a saturated solution of NaCl was used for cytidine suppression, 1 M instead of 0.5 M piperidine was used for the pyrimidine reactions, and piperidine was removed under reduced pressure at 60° instead of by lyophilization. Sequence gels were constructed and run essentially as described (27).

RESULTS AND DISCUSSION

Sequencing Strategy. The 16S RNA gene from the *E. coli rrn*B cistron was conveniently isolated in two *Eco*RI fragments from the transducing phage λ*rif*^{d}18 (17, 18). These fragments were cloned by insertion at the *Eco*RI site of ColE1 plasmid, as shown in Fig. 1. Because of the difficulty in growing large amounts of the recombinant plasmid pER24, presumably due to the presence of phage genes in the plasmid, the small *Hin*dIII fragment was excised from pER24 and reinserted into the vehicle pBR322 (19). Residues 1–80 and 648–673 and the sequence preceding the 5′ terminus of 16S RNA were obtained from pER24, residues 81–647 were from pKK115, and residues 674–1541 and the sequence following the 3′ terminus were from pER18. Fig. 2 summarizes the positions of cleavage by the restriction endonucleases used in this work and the fragments from which the sequences were derived.

Most of the sequence was derived by the partial chemical modification method of Maxam and Gilbert (24). The sequence was determined independently two to six times at each position,

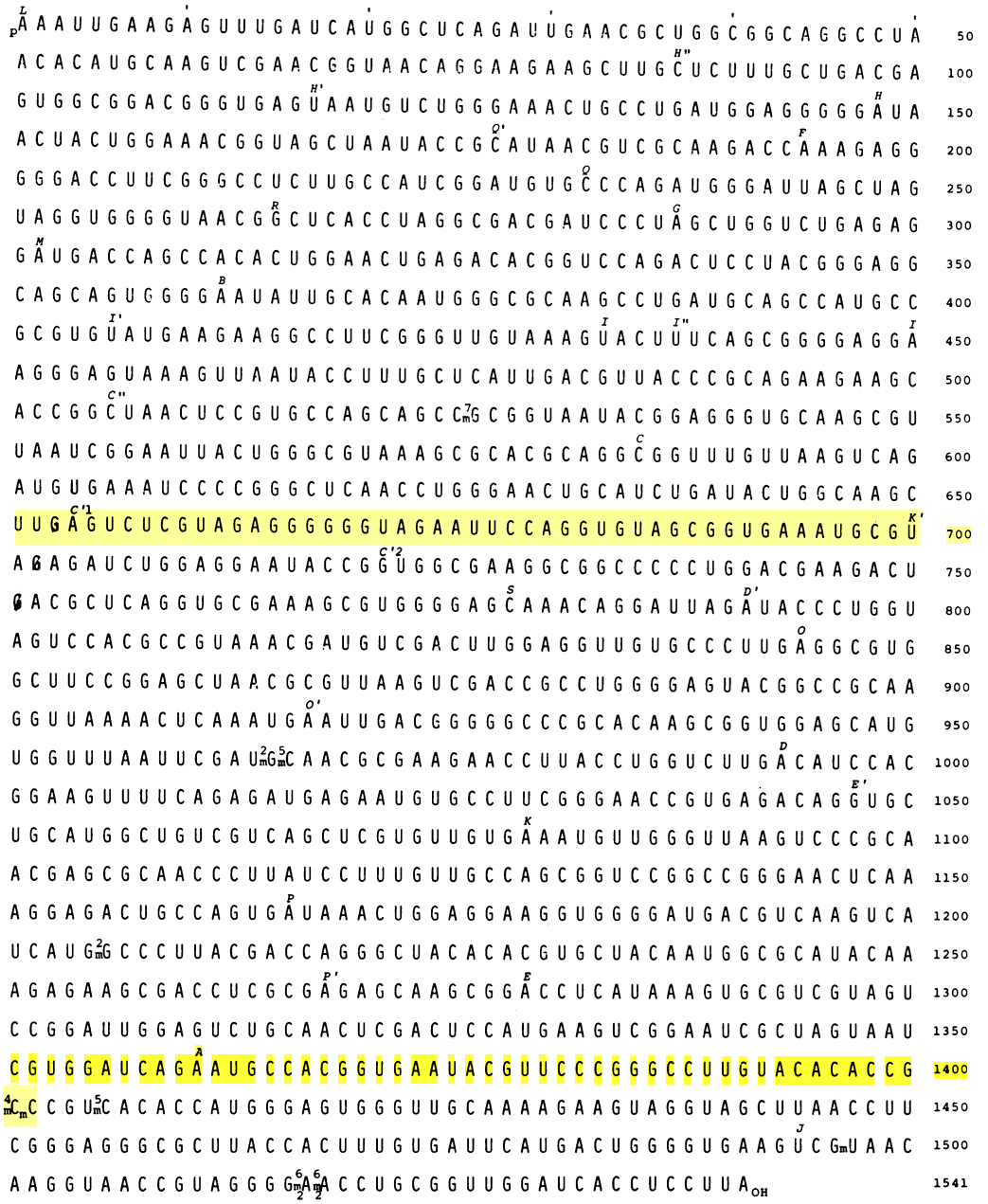

FIG. 4. The nucleotide sequence of 16S rRNA from the *rrnB* cistron of *E. coli*. The RNA sequence was inferred from the DNA sequence. The ends of the 16S rRNA and positions of methylated nucleotides were identified by comparison with RNA sequence results (1, 6–10, 28–31, and unpublished). For comparison with the sequences reported by Ehresmann *et al.* (7–9), the positions of their lettered sections is given. In some cases (e.g., section I″) the assignment of these sections was necessarily approximate.

and by sequencing both DNA strands, except for residues 1–80, 288–336, 652–673, and 1513–1541, which were sequenced on only one strand.

Sequence Results. Examples of sequence gels are shown in Fig. 3. By use of thin gels (27) and other minor modifications, it was often possible to read sequences accurately to about 250 base pairs from the proximal end of a fragment. Fig. 4 shows the complete nucleotide sequence of the 16S RNA as derived from the gene sequence. The total length of the chain is 1541 nucleotides, in close agreement with earlier estimates based on chemical methods (9, 32) but somewhat less than values predicted from physicochemical methods (33–35).

Comparison with Previous Sequence Results. Two conflicting catalogs of RNase T1 oligonucleotides have been published for *E. coli* 16S RNA (8, 36). In the course of determining sequences around kethoxal-reactive sites in 16S RNA (6, 12), we have noted some discrepancies between our results and those of Ehresmann *et al.* (8), but our results are in agreement with the sequences reported by Woese and coworkers (11, 36). In the most recently published version of their sequences, Ehresmann *et al.* (9) reported much closer agreement with the latter catalog, but a few discrepancies remain. There are no differences between the final catalog of T1 oligonucleotides as determined by the Woese group and those in the present sequence (Fig. 4), with the apparent exception of two modified oligonucleotides, whose general forms are C-C-N-C-G (positions 524–528) and N-C-C-G (positions 1401–1404) (refs. 10 and 11; C. Woese, personal communication). The differences, however, are trivial in that they concern only the identification of the modified nucleotides itself. Woese and coworkers incorrectly assumed that N in the pentanucleotide corresponds to $^{4}_{m}C_{m}$ and therefore N in the tetranucleotide is $^{7}_{m}G$ (C. Woese, personal communication).

A serious source of discrepancy between the sequence presented here and that by Ehresmann *et al.* (9) is in the alignment of oligonucleotides. This has been previously noted in instances

in which kethoxal modification studies have provided ordering of adjacent oligonucleotides (6), in a recent study using rapid RNA sequencing techniques (13), and in other studies using DNA methods (14). A prominent example of this problem is the region 398–504 (Fig. 4). In the latest version of the sequence of Ehresmann *et al.* (9), 43 nucleotides originating from positions 462–504 are inserted at position 432/433.

The remaining differences between the sequence presented here and the previous one (9) concern small sections of the molecule, and many involve the insertion or deletion of mono- or dinucleotides. In all, we note discrepancies involving about 200 nucleotides. A few of these differences might be due to sequence heterogeneity among the various rRNA cistrons, for which evidence has been presented (7–10, 34). Our sequence is in agreement with much of the recently published corrections in the region 1103–1165 and 1414–1488 reported by Ross and Brimacombe (13) on the basis of a newly developed rapid RNA sequence method. However, these authors have apparently deleted pyrimidines at positions 1161, 1466, 1473, 1477, 1478, and 1480. Young and Steitz (14) used DNA sequence methods and reported sequences from the 5′ and 3′ ends of 16S RNA from *rrnD* and one other rRNA cistron, as yet unidentified. Their results for residues 1–18 and 1381–1541 are in complete agreement with our findings.

Consequences of Sequence Changes for Secondary Structure Prediction. Previously, we noted that a majority of the sites of kethoxal modification of 16S RNA, which must be single stranded (37), have been assigned to double-stranded structures in the secondary structure model for 16S RNA proposed by Ehresmann *et al.* (8). Furthermore, some of the sites of nuclease attack observed under partial digestion conditions reported by these authors (7, 8) are inconsistent with their model. As anticipated, the sequence in Fig. 4 gives rise to a very different secondary structure prediction, in which these conflicts are largely resolved. A secondary structure model based on this sequence as well as a discussion of the structural and functional implications of these findings will be presented elsewhere.

Possible Coding Functions of 16S RNA. The use of rRNA as messenger by the cell has often been considered. Examination of the 16S rRNA sequence in Fig. 4 shows that most AUG or GUG triplets are followed by in-phase termination codons after a short interval. One exception is the AUG triplet at position 1187, which has no in-phase terminator until position 1439, allowing for a translation of a sequence of 84 amino acids. Furthermore, preceding the initiator codon by a distance of nine nucleotides is the sequence G-G-A-G-G, which would allow a favorable base-paired interaction with the 3′-terminal sequence of 16S rRNA in the mRNA recognition mechanism proposed by Shine and Dalgarno (1). Whether this theoretically possible message is in fact translatable *in vivo* is open to question. We have searched the sequence for possible homology with known *E. coli* protein sequences, including the ribosomal proteins, without success. It is most likely that translation of this sequence is precluded by formation of stable secondary structure of the RNA and by binding of ribosomal proteins.

We are grateful to the following friends and colleagues for supplying unpublished results, strains, technical advice, and encouragement at various stages of this work: J. Abelson, W. Barnes, B. Barrell, W. Gilbert, C. Guthrie, W. Herr, K. Katze, J. Kirschbaum, B. Lowe, A. Maxam, A. Otsuka, C. Panofsky, R. Rodriguez, F. Sanger, S. Stein, J. Steitz, C. Woese, and R. Young. J.B. is a postdoctoral fellow of the Fogarty International Center of the National Institutes of Health; M.L.P. is the recipient of a President's Undergraduate Fellowship from the University of California. This work was supported by National Institutes of Health Grant GMS-17129.

1. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
2. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4734–4738.
3. Noller, H. F. & Chaires, J. B. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3115–3118.
4. Schwartz, I. & Ofengand, J. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 3951–3955.
5. Santer, M. & Shane, S. (1977) *J. Bacteriol.* **130,** 900–910.
6. Chapman, N. M. & Noller, H. F. (1977) *J. Mol. Biol.* **109,** 131–149.
7. Ehresmann, C., Stiegler, P., Fellner, P. & Ebel, J. P. (1972) *Biochimie* **54,** 901–967.
8. Ehresmann, C., Stiegler, P., Fellner, P., & Ebel, J. P. (1975) *Biochimie* **57,** 711–748.
9. Ehresmann, C., Stiegler, P., Carbon, P. & Ebel, J. P. (1977) *FEBS Lett.* **84,** 337–341.
10. Uchida, T., Bonen, L., Schaup, H. W., Lewis, B. J., Zablen, L. & Woese, C. (1974) *J. Mol. Evol.* **3,** 63–77.
11. Magrum, L., Zablen, L., Stahl, D. & Woese, C. (1975) *Nature (London)* **257,** 423–426.
12. Noller, H. F. (1974) *Biochemistry* **13,** 4694–4703.
13. Ross, A. & Brimacombe, R. (1978) *Nucleic Acids Res.* **5,** 241–256.
14. Young, R. A., & Steitz, J. A. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3593–3597.
15. Hogan, J. J. & Noller, H. F. (1978) *Biochemistry* **17,** 587–593.
16. Santer, M. & Santer, U. (1973) *J. Bacteriol.* **117,** 63–77.
17. Kirschbaum, J. B. & Konrad, E. B. (1973) *J. Bacteriol.* **116,** 517–526.
18. Lindahl, L., Jaskunas, S. R., Dennis, P. P. & Nomura, M. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 2743–2747.
19. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. & Boyer, H. W. (1977) *Gene* **2,** 95–113.
20. Valenzuela, P., Bell, G. I., Venegas, A., Sewell, E. T., Masiarz, F. R., DeGennaro, L. J., Weinberg, F. & Rutter, W. J. (1977) *J. Biol. Chem.* **252,** 8126–8135.
21. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94,** 441–448.
22. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
23. Szybalski, W. (1968) *Methods Enzymol.* **12,** 330–360.
24. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
25. Glynn, I. M. & Chappell, J. B. (1964) *Biochem. J.* **90,** 147–149.
26. Szalay, A. A., Grohmann, K. & Sinsheimer, R. L. (1977) *Nucleic Acids Res.* **4,** 1569–1578.
27. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87,** 107–110.
28. Noller, H. F. & Herr, W. (1974) *Mol. Biol. Rep.* **1,** 437–439.
29. Ehresmann, C., Stiegler, P. & Ebel, J. P. (1974) *FEBS Lett.* **49,** 47–48.
30. Fellner, P. (1969) *Eur. J. Biochem.* **11,** 12–27.
31. Sprague, K. U. & Steitz, J. A. (1975) *Nucleic Acids Res.* **2,** 787–798.
32. Midgley, J. E. M. (1965) *Biochim. Biophys. Acta* **108,** 340–347.
33. Kurland, C. G. (1960) *J. Mol. Biol.* **2,** 83–91.
34. Stanley, W. M. & Bock, R. M. (1965) *Biochemistry* **4,** 1302–1311.
35. Ortega, J. P. & Hill, W. E. (1973) *Biochemistry* **12,** 3241–3243.
36. Woese, C. R., Fox, G. E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, B. J. & Stahl, D. (1975) *Nature (London)* **254,** 83–86.
37. Litt, M. (1969) *Biochemistry* **8,** 3249–3253.