# THE SEQUENCE OF *ESCHERICHIA COLI* RIBOSOMAL 16 S RNA DETERMINED BY NEW RAPID GEL METHODS

P. CARBON, C. EHRESMANN, B. EHRESMANN and J. P. EBEL
*Institut de Biologie Moleculaire et Cellulaire du CNRS, Laboratoire de Biochimie, 15, rue René Descartes, 67084 Strasbourg, France*

Received 28 August 1978

## 1. Introduction

Ten years ago, we began to determine the sequence of ribosomal 16 S RNA from *Escherichia coli* MRE 600 using the conventional sequencing technique of Sanger et al. [1]. As we pointed out in [2], a limit in the length of sequence that could be determined was reached, due partly to the intrinsic limitations of the technique itself, and partly to the extreme difficulty of preparing RNA fragments from exposed regions which are very susceptible to limited ribonuclease T1 digestion.

Recently a new ribonuclease, extracted from the venom of cobra *Naja oxiana* [3], which is specific for double-stranded or base-stacked regions [4] allowed us to prepare new RNA fragments different from those previously obtained. Furthermore, the recent development of rapid RNA sequencing techniques, using specific digestions of 5′-$^{32}$P-labelled RNA fragments and subsequent fractionation of the digests on polyacrylamide gels [5,6], allowed us to make rapid progress.

Using the new methodology, we have examined the entire 16 S RNA molecule: the sequences of previously non-ordered regions have been resolved; and several corrections have been made in areas already sequenced using the conventional RNA sequencing technique [1].

We report here the complete sequence of the 16 S RNA encompassing 1542 nucleotides.

## 2. Materials and methods

### 2.1. *Preparation of the RNA fragments*

In order to obtain fragments covering the entire molecule, three different enzymatic partial hydrolyses have been used. Subfragments were prepared:

(i) By dissociation of the 12 S and 8 S RNAs obtained by ribonuclease T1 (Sankyo) digestion of 16 S RNA in reconstitution buffer as in [7];
(ii) By ribonuclease digestion of the ribosomal protein S20–16 S RNA complex as in [8];
(iii) By digestion of 30 S subunits with the cobra venom ribonuclease under conditions that will be described [9].

In all cases, the 16 S RNA was weakly uniformly $^{32}$P-labelled in order to facilitate isolation and purification of fragments. The RNA subfragments were fractionated on preparative polyacrylamide gel slabs ($400 \times 300 \times 4$ mm$^3$) containing 8 M urea and eluted from the gel, in the absence of carrier RNA, using the elution technique in [5].

### 2.2. *Dephosphorylation of the RNA fragments*

Unlike the ribonuclease T1 fragments, the cobra venom ribonuclease fragments have a phosphate at their 5′-end. The latter RNA fragments were incubated in 25 μl of a solution containing 2.5 mM Tris/HCl (pH 8.0) and $25 \times 10^{-3}$ units of alkaline phosphatase from calf intestine (Boehringer) at 55°C for 25 min. The phosphatase activity was destroyed

by making the solution 5 mM in nitrilotriacetic acid [10] and the mixture was incubated 1 min at 100°C and dried.

2.3. *5′-Terminal labelling of the RNA fragments*

The fragments were labelled using T4 polynucleotide kinase prepared as in [11]. Labelled [$\gamma$-$^{32}$P]ATP (300–500 Ci/mmol) was prepared by a modification of the procedure in [12] or purchased from Amersham (3000 Ci/mmol). A minimum of 80 pmol highly-labelled ATP, or a 2-fold molar excess relative to pmol RNA 5′-ends present, was generally used. Each RNA fragment was incubated in the presence of ATP and 4 units polynucleotide kinase in 50 $\mu$l 10 mM Tris/HCl (pH 8.0); 10 mM $MgCl_2$; 6 mM mercaptoethanol, for 30 min, at 37°C. The RNA fragment was precipitated with ethanol and the pellet was dissolved in 20 $\mu$l 20 mM Na citrate (pH 5.0); 1 mM $Na_2$ EDTA; 7 M urea; 0.025% xylene cyanol; 0.025% bromophenol blue. The resulting 5′-$^{32}$P-labelled fragments were then fractionated on polyacrylamide gel slabs (400 × 300 × 2 mm$^3$) containing 8 M urea and eluted as in [5] in the presence of 20 $\mu$g tRNA carrier.

2.4. *Partial digestions of the 5′-$^{32}$P-labelled fragments*

2.4.1. Partial hydrolysis with ribonucleases T1 and U2

The same reaction buffer and incubation conditions were used as in [5]. The ribonuclease T1 (Sankyo)/RNA ratio used was generally $2 \times 10^{-2}$–$2 \times 10^{-3}$ unit/$\mu$g and the ribonuclease U2 (Sankyo)/RNA ratio was 1 unit/$\mu$g.

2.4.2. Partial hydrolysis with ribonuclease A

The same reaction buffer was used as for T1 and U2 hydrolysis, for 30 min, at 55°C. The ribonuclease A/RNA ratio was generally 0.2–2 ng/$\mu$g.

2.4.3. Partial hydrolysis with ribonuclease Phy I

The same reaction buffer was used as in [6]. The enzyme/RNA ratio was $2 \times 10^{-3}$ unit/$\mu$g. The mixture was incubated for 0.5–3 min depending on the size of the RNA fragment to be analysed.

2.4.4. Unspecific partial hydrolysis

This was carried out using the alkaline hydrolysis conditions in [5] or by incubating the RNA fragment in 10 $\mu$l double-distilled water for 10–30 min at 100°C, in a sealed capillary, as in [13]; the incubation time was chosen according to the size of the fragment to be analysed. After incubation, 5 $\mu$l 10 M urea, 0.025% xylene cyanol, 0.025% bromophenol blue were added.

2.5. *Fractionation of the digestion products*

The various samples were loaded onto thin polyacrylamide gel slabs (400 × 300 × 0.5 mm$^3$ or 900 × 300 × 0.5 mm$^3$) according to [14]. The acrylamide concentration in the gel (8–20%) was chosen according to the length of the fragments to be analysed. Electrophoresis buffer was the same as in [5]. The gels were generally freshly prepared and pre-electrophoresed for 1 h at 800 V. The electrophoresis was carried out at 1200–1500 V at constant voltage or wattage. In our first experiments, 2 mm thick gels were used.

## 3. Results and discussion

As already described by Donis-Keller et al. [5], we found that ribonucleases T1 and U2 cleave respectively GpN and ApN quite uniformly and almost exclusively. Ribonuclease A does not cleave after all pyrimidines. Ribonuclease Phy I is a useful tool for discriminating between C and U residues in numerous cases. We found that ribonuclease Phy I generally cleaves UpN with a higher efficiency than it cleaves CpN; e.g., UpG is much more susceptible than CpG. We also found that the cleavage rate depends on the 3′-neighbour of the cleavage site. In our hands, the most susceptible phosphodiester bonds were UpA, ApA and GpG, while CpU, ApU and CpC were the most resistant; generally the NpU bonds were resistant, and therefore it was rather difficult to discriminate between UpU and CpU in pyrimidine stretches. However, since RNA fragments were in non-denaturing conditions their secondary structures may have played a non-negligible role. All these observations agree reasonably well with the cleavage specificity studies of Bargetzi et al. [15].

In most cases non-enzymatic cleavages occurred between a pyrimidine and an adenine residue, which were also clearly observable in the control. Evidently this phosphodiester bond is fragile under the conditions used: heating in 7 M urea at 55°C.

In our first experiments we used relatively thick gel slabs (2 mm). In several cases, we then observed serious variations in the distances between consecutive nucleotide bands and we missed some nucleotides that had been unambiguously present in our previous analyses [2]. Probably this resulted from incomplete denaturation of very stable structured regions. These disadvantages could be easily overcome by using thinner gel slabs (about 0.4–0.5 mm) and electrophoresing them at high voltage, as described by Sanger and Coulson [14]. As an example, using 2 mm thick gels, we unambiguously read the sequence GGGUG instead of the correct sequence GGGCCUUG (positions 1385–1392, in section A); but using the thin gels at high temperature, the sequence was read correctly. Furthermore, the bands obtained in this way were sharper, thus increasing the resolving power of the gel.

A diagram of the various RNA fragments that have been sequenced using the new methodology is given in fig.1: the entire molecule is covered by these fragments. One can see how the ribonuclease from cobra venom has helped by providing new RNA fragments which could not be obtained by T1 hydrolysis. In particular, fragments containing the linking region between section C′1 and K′ (positions 682–691), which were regularly excised by ribonuclease T1, could be obtained using this ribonuclease. The cobra venom enzyme was also useful in order to verify overlapping between different sections (e.g., G–M, C″–C, C′2–S, O–O′, O′–D, E′–K, K–P).

The complete sequence deduced for 16 S RNA is shown in fig.2. This sequence takes into account our earlier results on cistron heterogeneity and on methylated nucleotides. Indeed, regions where two or three different related sequences were simultaneously present were difficult, if not impossible, to read on the gels (e.g., positions 267–275 in section R), where it was not possible to get one unique sequence. Furthermore, methylated nucleotides were not cleaved, or only poorly cleaved, by the nucleases used.

All the remaining ambiguities left by our previous results [2] from conventional RNA sequencing techniques [1] (regions where the ribonucleases T1 and A oligonucleotides were sequenced but not ordered) have now been resolved. In addition, we have made a number of corrections. Three main characteristic errors in the earlier results can be distinguished:

(i) The location of the common T1 dinucleotides UG and CG;
(ii) The ordering of G and A residues in large ribonuclease A oligonucleotides;
(iii) The inversion of oligonucleotides around a central sequence.

The two first classes of errors are directly related with the limit imposed by the conventional RNA sequencing technique; the third arose from the great difficulties in preparing enough of several fragments and in getting suitable partial subfragments.

The largest correction has been made in section I (positions 428–504): the sequence UAAAGU occurs not just once (in position 457–462) but also a second

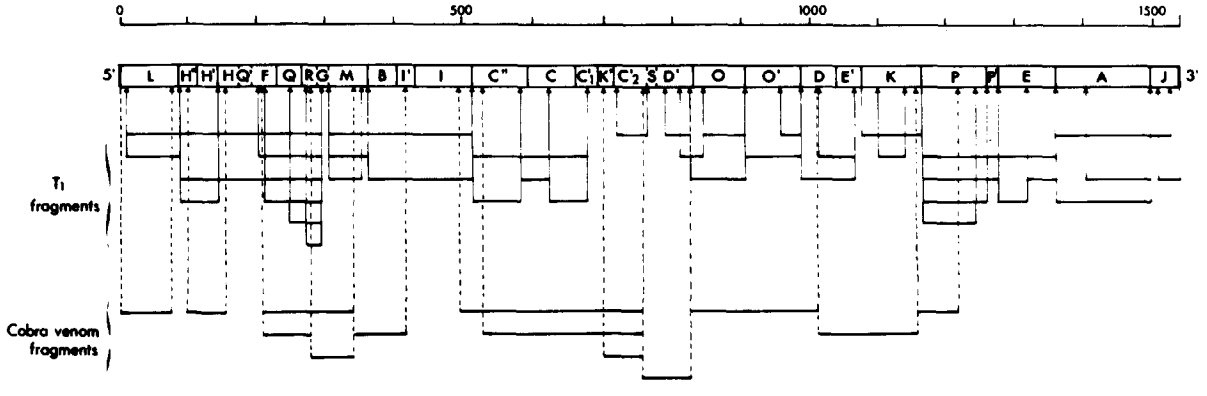


Fig.1. RNA fragments, obtained by digestion with ribonuclease T1 or with cobra venom ribonuclease, used in the sequencing of the 16 S RNA by the new rapid gel method. The letters indicate the various sections according to [2].

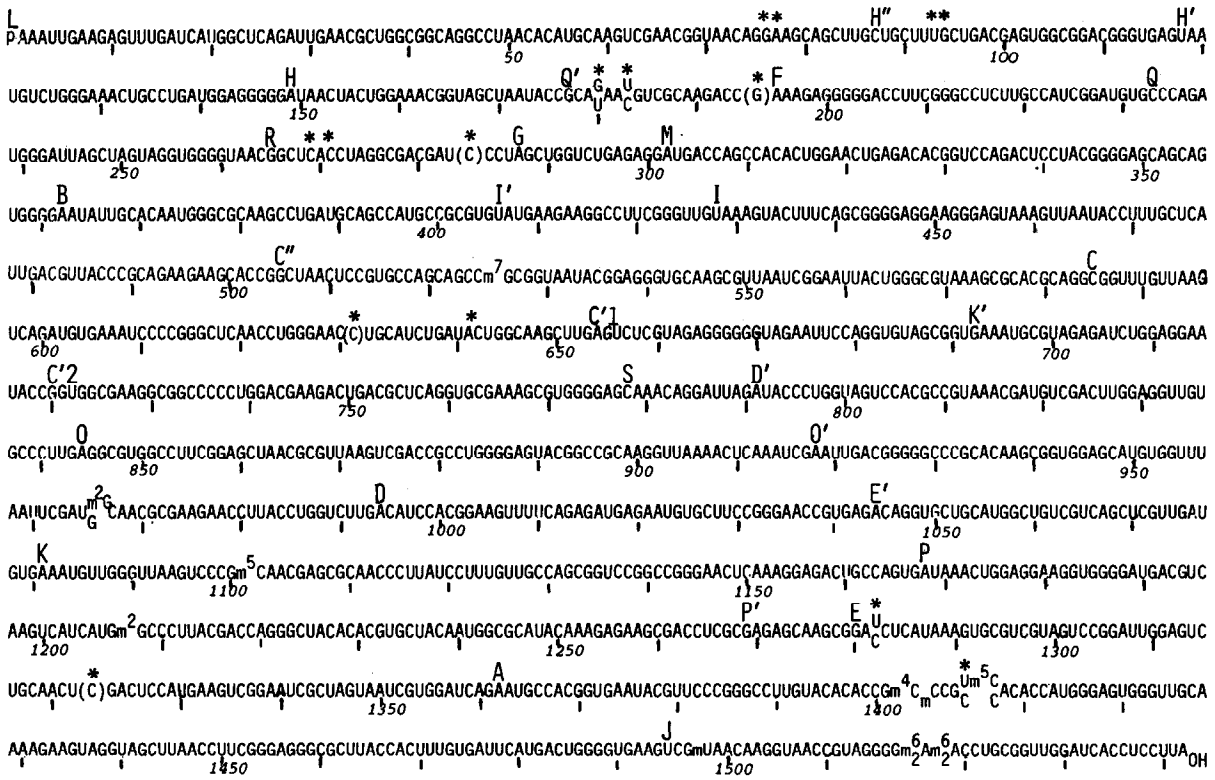

Fig.2. The sequence of the 16 S RNA of *E. coli*. The various sections are indicated by capital letters. Positional heterogeneities, as detected [2], are denoted by asterisks. Possible insertions are shown in parentheses with asterisks over the top.

time (in positions 428–433). Furthermore in the sequence previously reported the two blocks 433–456, and 457–504 were inverted. Some corrections have also been made in section C′2 (positions 741–768) where an isomer of the ribonuclease A oligonucleotide GAAAGC: GAAGAC has now been found to be present. Several other ordering errors concerning adjacent T1 oligonucleotides have been detected in sections B (positions 399–405), I (positions 481–490), C″ (positions 542–556), C′1 (positions 654–663), D′ (positions 824–837), O (positions 877–884). Other corrections concern the positions of the T1 dinucleotides UG and CG and the internal ordering of G and A residues in large ribonuclease A products.

Most of these results are clear and unambiguous; nevertheless, in a very limited number of cases, where the results were not completely free of doubt (especially in some pyrimidine stretches), we took into account our preceding data. The small linking region between sections E′ and K (positions 1073–1080) was not completely clear, in spite of numerous experiments. Region 177–204 was in places difficult to read, partly because of the presence of heterogeneities and partly because a relative resistance to the enzymes; the adjacent sequences were unambiguously deduced from shorter fragments.

Our present results agree with partial RNA sequence data in section K (positions 1105–1162) and in section A (positions 1415–1461) reported [16] and with the DNA sequence corresponding to the 160 nucleotides at the 3′-end of the 16 S RNA gene deduced by Young and Steitz [17].

The sequence of the 16 S RNA we report here is now complete and continuous. It encompasses 1542 nucleotides without counting the possible insertions. The implications of the sequence of 16 S RNA for its secondary structure will be discussed elsewhere.

## Acknowledgements

We are indebted to Dr G. Keith for his generous gift of T4 polynucleotide kinase, to Dr S. K. Vassilenko for providing us with the *Naja oxiana* venom ribonuclease, to Dr J. P. Bargetzi for providing us with Phy I ribonuclease, to Professor M. Daune for his generous gift of protein S20, and to Miss B. Muller for her excellent technical assistance. This work was financially supported by the Centre National de la Recherche Scientifique, the Délégation Générale à la Recherche Scientifique et Technique and the Commissariat à l'Energie Atomique.

## References

[1] Sanger, F., Brownlee, G. G. and Barrell, B. G. (1965) J. Mol. Biol. 13, 373–398.
[2] Ehresmann, C., Stiegler, P., Carbon, P. and Ebel, J. P. (1977) FEBS Lett. 84, 337–341.
[3] Vassilenko, S. K. and Ryte, V. C. (1975) Biokhimia 40, 578–583.
[4] Vassilenko, S. K. (1978) in preparation.
[5] Donis-Keller, H., Maxam, A. and Gilbert, W. (1977) Nucl. Acids Res. 4, 2527–2538.
[6] Simoncsits, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. and Guilley, H. (1977) Nature 269, 833–836.
[7] Ehresmann, C., Stiegler, P., Fellner, P. and Ebel, J. P. (1975) Biochimie 57, 711–748.
[8] Ungewickell, E., Garrett, R. A., Ehresmann, C., Stiegler, P. and Fellner, P. (1975) Eur. J. Biochem. 51, 165–180.
[9] Vassilenko, S. K., Ehresmann, C., Carbon, P. and Ebel, J. P. (1978) in preparation.
[10] Silberklang, M., Gillum, A. M. and RajBhandary, U. L. (1977) Nucl. Acids Res. 4, 4091–4108.
[11] Panet, A., van de Sande, J. H., Loewen, P. C., Khorana, H. G., Paae, A. J., Lillehaug, J. R. and Kleppe, K. (1973) Biochemistry 12, 5045–5050.
[12] Glynn, I. M. and Chappell, J. B. (1964) Biochem. J. 90, 147–149.
[13] Winter, G. W. and Brownlee, G. G. (1978) personal communication.
[14] Sanger, F. and Coulson, A. R. (1978) FEBS Lett. 87, 107–110.
[15] Pilly, D., Niemeyer, A., Schmidt, M. and Bargetzi, J. P. (1978) J. Mol. Biol. in press.
[16] Ross, A. and Brimacombe, R. (1978) Nucleic Acids Res. 5, 241–256.
[17] Young, R. A. and Steitz, J. A. (1978) personal communication.