Volume 5 Number 1 January 1978 Nucleic Acids Research

Application of a rapid gel method to the sequencing of fragments of 16S ribosomal RNA from Escherichia coli

Alexander Ross and Richard Brimacombe

Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, GFR

Received 3 November 1977

ABSTRACT

A gel sequencing method has been applied to two 5' end-labelled fragments of the 16S ribosomal RNA from E. coli. The procedure involves partial enzymatic hydrolysis by ribonucleases T_1, U_2 or A, in order to generate series of end-labelled sub-fragments terminating in guanine, adenine, or pyrimidine residues, respectively. The two fragments concerned were approximately 75 and 90 nucleotides in length, and both arose from the 3' region of the 16S RNA. The sequences deduced are compared with the published sequence of 16S RNA, and contribute information to the final ordering of the ribonuclease T_1 oligonucleotides in the latter, as well as revealing some probable errors.

INTRODUCTION

The application of high-resolution polyacrylamide gel electrophoresis has revolutionized nucleic acid sequencing technology. Up to now this new approach has been applied mostly to the sequencing of DNA, based either on the "copying" method of Sanger and Coulson (1), or on the method of chemical degradation of ^{32}P end-labelled DNA developed by Maxam and Gilbert (2). In the case of RNA, an elegant copying method involving reverse transcriptase was recently published by Brownlee and Cartwright (3), but the most direct approach relies simply on partial degradation by different endonucleases with an end-labelled RNA molecule as substrate, in order to generate series of labelled sub-fragments each terminating in a known nucleotide. This method has been used with oligonucleotides by Gupta and Randerath (4), and has more recently been applied to the sequencing of yeast 5.8 S RNA (5) and a tRNA from B. stearothermophilus (6). In both these latter cases, a 5' end-labelled RNA substrate was used, and the partial digestion products were

separated by polyacrylamide gel electrophoresis, the sequence being subsequently read off from the gel.

The experiments of Donis-Keller et al (5) involved hydrolysis with ribonucleases U_2 and T_1, to localize adenine and guanine residues respectively, the number of intervening pyrimidine residues being deduced from a partial alkaline hydrolysis, which gave a "ladder" of degradation products with a cleavage at every base. Simoncsits and Brownlee (6) used ribonuclease A and ribonuclease I from Physarum polycephalum (7) in addition to ribonucleases T_1 and U_2, and a "ladder" was generated by heating the RNA in formamide; ribonuclease A is specific for pyrimidines, whereas the P. polycephalum enzyme cleaves at every base except cytosine (8). We have been simultaneously developing a very similar sequencing method, using ribonucleases A, U_2 and T_1, with a view to helping the completion of the sequence of 16S ribosomal RNA from E. coli. While our method does not distinguish between pyrimidines, it gives (as was also pointed out by Donis-Keller et al, (5)) sufficient information to complete the ordering of T_1-oligonucleotides obtained from conventional sequencing experiments.

A partial sequence of 16S RNA was first published in 1972 by Ehresmann et al (9), and a substantially revised and improved version appeared in 1975 (10). Almost 25 % of the ribonuclease T_1 digestion products were still not finally ordered within this revised sequence, although progress has subsequently been made (C. Ehresmann, personal communication). It is precisely at this last difficult stage in sequencing of a long RNA molecule that a gel method of the type just described could be most usefully applied. Furthermore, during the course of our research into RNA-protein interactions within the E. coli 30S ribosome, we have accumulated a catalogue of reproducibly obtainable and highly specific RNA fragments from known regions of the 16S molecule (11,12), several of which would be ideal substrates for this sequencing method. In this paper, we present sequence data obtained in this manner, from two such fragments approximately 75 and 90 nucleotides in length, arising from the 3'-half of the 16S RNA molecule.

MATERIALS AND METHODS

Isolation of RNA fragments

Ammonium chloride washed 30S ribosomal subunits from E. coli MRE 600 were prepared as described (13), and were labelled in the RNA moiety with ^{3}H-uridine to a specific activity of 5 x 10^{6} counts/min/mg. The subunits (ca. 2 mg) were hydrolysed with ribonuclease T_1 in the presence of 2 M urea exactly as described previously (11), and the hydrolysis products were separated by electrophoresis on a 5 % polyacrylamide gel at pH 6 (14). The gel was sliced and analysed for ^{3}H-radioactivity, and gel slices containing ribonucleoprotein fragment "III" (11) or the small RNA fragment "A1" (12) were subjected to a second electrophoresis step on gels of 7 % or 10 % polyacrylamide, respectively, together with 7 M urea and 0.1 % dodecyl sulphate. After electrophoresis, these gels were sliced and analysed for ^{3}H-radioactivity, and appropriate RNA bands were extracted with phenol exactly as described (11,12), with the exception that no carrier RNA was added. The isolated RNA fragments were precipitated with ethanol, taken up in 50 mM Tris-HCl pH 7.8, 1 mM EDTA, 6 mM 2-mercaptoethanol, and stored at -80° until required.

5'-end-labelling of RNA fragments

The fragments were labelled using a slight modification of the procedure of Maxam & Gilbert (2), as follows. An aliquot (20 µl) of the appropriate RNA fragment solution at a concentration of ca. 5 µM was made 0.1 mM in spermidine and heated to 90° for 5 min., then rapidly cooled. This solution was added to a previously lyophilized sample of γ^{32}P-labelled ATP (20 µCi, specific activity 2500 Ci/mmol, Amersham, UK) and the concentration of ATP was adjusted to 5 µM by addition of a suitable amount of unlabelled ATP. The solution was made 10 mM in magnesium, and then incubated at 37° for 30 min with 3-5 units of polynucleotide kinase (15).

The labelled fragments were purified by electrophoresis on a 20 % polyacrylamide gel 40 cm long and 0.25 cm thick, using the buffer system of Maxam & Gilbert (2), the gel being allowed to run until a xylene cyanol marker had run 40 cm. The RNA bands were localised by autoradiography, and appropriate parts

of the gel were then sliced and analysed for radioactivity; RNA was extracted from the gel slices in the presence of phenol as described above, and the 5'-labelled products were stored under ethanol until required.

Sequence analysis

Aliquots of 5'-labelled RNA (ca. 200,000 counts/min) were centrifuged off, washed with ethanol and dried under vacuum. The pellet was dissolved in an appropriate volume (ca. 0.5 ml) of a solution containing 1 mM EDTA and 1 mg/ml unlabelled carrier RNA, and the solution was divided into three equal portions. Two of these were made 10 mM in Tris-HCl pH 7.8 (for hydrolysis with ribonucleases A or T_1), and the third was made 3 mM in triethanolamine-acetate pH 4.5 (for hydrolysis with ribonuclease U_2). Each solution was heated to 90° for 5 min, followed by rapid cooling, and was then divided into aliquots of 10 µl for the enzymatic hydrolyses.

Normally, six different concentrations of each of the three ribonucleases were used, covering the following ranges: ribonuclease A, 0.005 - 0.5 µg/ml; ribonuclease U_2, 0.1 - 10.0 µg/ml; ribonuclease T_1, 1 - 100 units/ml. 2 µl of the appropriate enzyme solution was added to each 10 µl RNA sample, and the reaction mixtures were incubated at 37° for 1 hr. Next, the samples were pooled in pairs, so that for each enzyme the two lowest, the two middle, and the two highest enzyme concentrations were combined. The resulting nine samples (three for each enzyme), together with a control sample minus enzyme, were treated with an equal volume of 10 M urea solution containing gel reservoir buffer, and were applied to a 20 % polyacrylamide gel 40 cm long and 0.1 cm thick, according to Maxam and Gilbert (2). Three or four duplicate gels were run for each sequence analysis. The gels were run at 300 or 600 volts, until a xylene cyanol blue marker had run respectively 15 cm, 30 cm, or the equivalent of 45 cm or 60 cm. The gels were then covered with parafilm and subjected to autoradiography.

RESULTS

The RNA fragments selected for this study were firstly "Band III, sub-fragment 5" (11), which is approximately 75 nucleotides long and is part of section K of the 16S RNA (10). This is a region of the sequence in which about 50 % of the T_1 oligonucleotides have not been finally ordered (10). The second RNA substrate was fragment "A1" (12), which is approximately 90 nucleotides long and arises from section A of the 16S molecule, a region which (in contrast to section K) is almost completely sequenced (10). For convenience these RNA pieces will be referred to as fragment "5" and fragment "A" respectively in the ensuing discussion. Both fragments are obtained by limited ribonuclease T_1 digestion of intact 30S ribosomal subunits, and they were isolated by the published procedures (11, 12), with the exception that 30S subunits were used in which the RNA moiety was labelled with ^{3}H-uridine instead of ^{32}P-orthophosphate. The isolation procedure is summarized briefly in the Methods section.

Since the RNA fragments are internal products of a partial ribonuclease T_1 digest of 16S RNA, they have free 5'-hydroxyl groups, and it follows that 5' end-labelling with γ-^{32}P-ATP and polynucleotide kinase (15) could be achieved without prior phosphatase treatment. This end-labelling was carried out as described in the Methods section, and the reaction mixture was purified by a long electrophoresis run on a 20 % polyacrylamide gel (2), which served the dual purpose of removing unreacted ^{32}P-ATP and also of ensuring that the end-labelled fragments were completely homogeneous. Autoradiographs and a ^{32}P-profile of the 5'-labelled fragments on these gels are presented in Fig. 1, which shows that in both cases the RNA fragment runs as a clean sharp band with very little contaminating material. Appropriate gel fractions were homogenized and the RNA extracted with phenol, for use in the sequencing experiments.

For the sequence analysis, partial digests were made with ribonuclease A, T_1 and U_2, to give series of end-labelled subfragments terminating in pyrimidines, guanines and adenines, respectively (cf. refs. 5, 6). In order to obtain a wide spec-

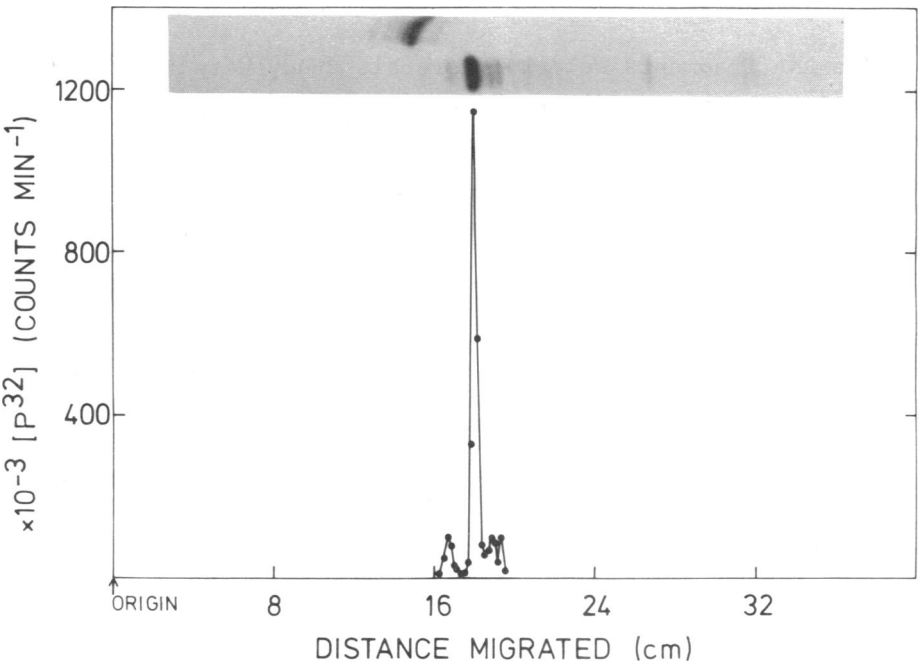

Figure 1: Isolation of 5' end-labelled fragments "A" and "5". The top part of the figure shows an autoradiograph of the two fragments on a 20 % gel (Fragment "A" above, "5" below). Direction of electrophoresis is from left to right. Fragment "A" ran rather skew, as the slot containing this fragment was rather near to the edge of the gel in this experiment. The lower part of the diagram shows the radioactivity found in fragment "5", after slicing the appropriate part of the gel. (Fragment "A" was simply excised in this particular experiment, with the help of the autoradiograph).

trum of hydrolysis products, several aliquots of RNA (mixed with a standard amount of unlabelled carrier RNA) were digested with a range of concentrations of each enzyme, as detailed in the Methods section. Before adding enzyme, the reaction mixtures were heated briefly to 95°, to randomize the secondary structure as far as possible. Hydrolsed samples were combined in pairs (see Methods section) and loaded onto 20 % polyacrylamide gels (2), together with a control sample minus enzyme. Triplicate or quadruplicate sets of samples were prepared in each case, and these were run on separate gels for different times in order to obtain optimal separations of the various degradation products. Subsequently the gels were auto-

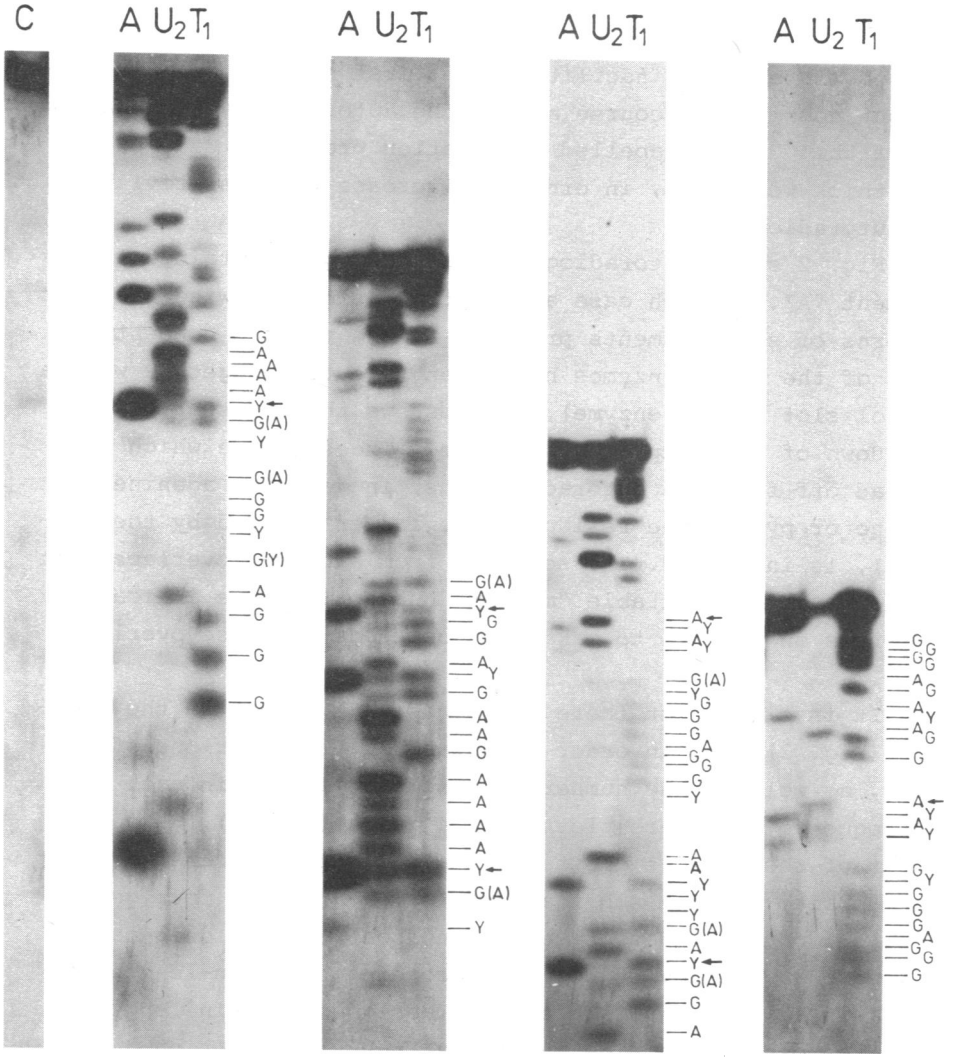

Figure 2: Sequence analysis of fragment "A" on 20% gels.

Thw first column shows a control slot minus enzyme, and the subsequent columns are sets of three hydrolysates (ribonuclease A, U2 or T1 as indicated) from gels run for different times (see Methods section). Direction of electrophoresis is from top to bottom, and the sequence is read (5' to 3') from the bottom of the gel. The sequence read off each gel is indicated (Y denotes pyrimidine), and the arrows give a point of reference showing how each sequence overlaps that of the next gel.

radiographed, and typical results are presented in Figs. 2 and 4. The fact that the RNA fragments were uniformly labelled with tritium (in order to facilitate the first stages of their isolation) does not of course affect the autoradiographic analysis of the ^{32}P end-labelled degradation products. Gels only 1 mm thick were used, in order to increase the sensitivity of the autoradiography.

Fig. 2 shows autoradiographs of four gels obtained from fragment "A". In each case adjacent gel slots showing the best patterns of sub-fragments produced by the various concentrations of the three enzymes have been selected, together with a control slot (minus enzyme), which shows that no spurious breakdown of the sample has occurred. The sequence which can be read off from the autoradiographs, in terms of adenine, guanine or pyrimidine residues (the latter denoted by the letter Y), is indicated. The sequence from each gel overlaps that of the next, and suitable "marker" nucleotides are indicated by arrows on Fig. 2, to make the positions of these overlaps clear.

It is clear that there are occasional "gaps" in the sequence (Fig. 2), and in order to establish the positions of these, as well as the number of residues missing in each gap, the sequence data were plotted on a graph (Fig. 3). Here the number of residues is plotted (on a logarithmic scale) against the mobility of each sub-fragment for each of the four gels. Since oligonucleotides of less than about ten residues cannot be read because they do not run according to chain-length on 20 % gels (2), we have arbitrarily assigned a chain-length of 25 to the oligonucleotide running at the same position as a xylene cyanol marker (cf. ref. 5) in order to plot this graph. This figure may of course be one or two nucleotides in error, but this affects neither the deduction of the sequence nor the linearity of the data, which fit reasonably well to a straight line in each case, with the positions of the gaps being clearly visible.

Similar results from fragment "5", in this case from three gels instead of four, are presented in an identical manner in Figs. 4 and 5, and the combined sequence data obtained

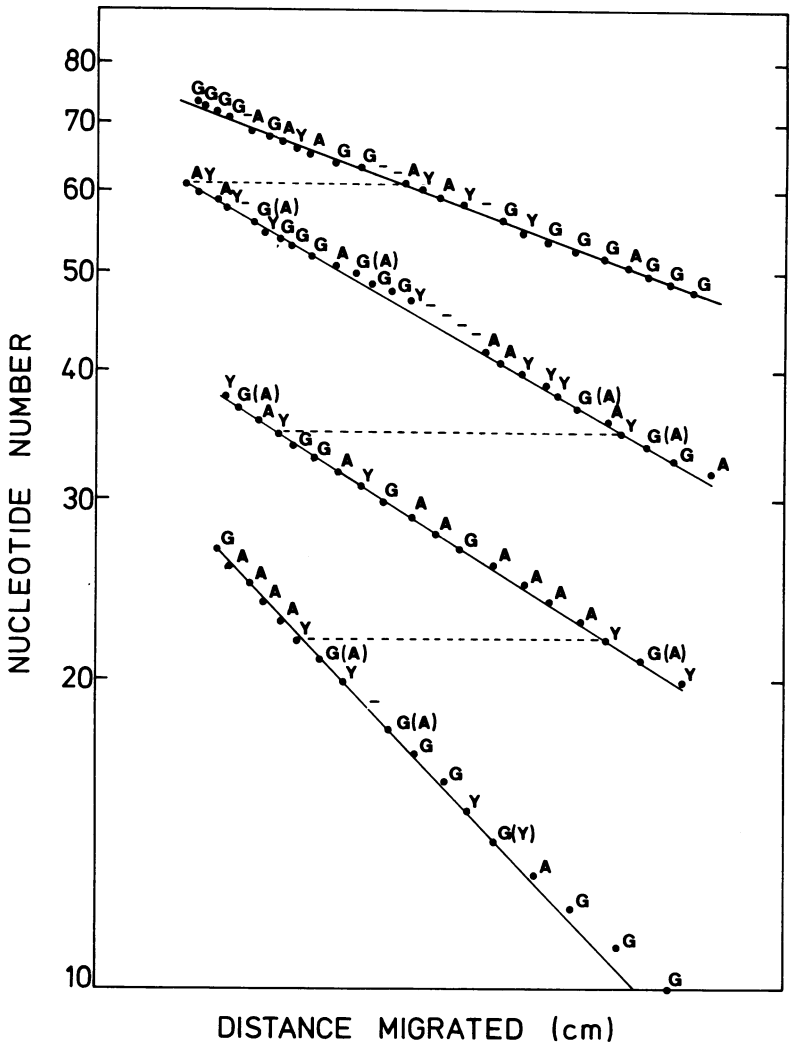

Figure 3: Logarithmic plot of sequence data from Fig. 2.

Chain length is plotted against distance migrated for each of the gels in Fig. 2, assuming that the position of a xylene cyanol marker corresponds to a chain length of 25 residues (see text). The positions and numbers of "gaps" in the sequence are indicated by dashes, the dotted lines corresponding to the arrows in Fig. 2. No scale is given on the abscissa, as each set of points has for convenience been moved to left or right so as to save space on the diagram.

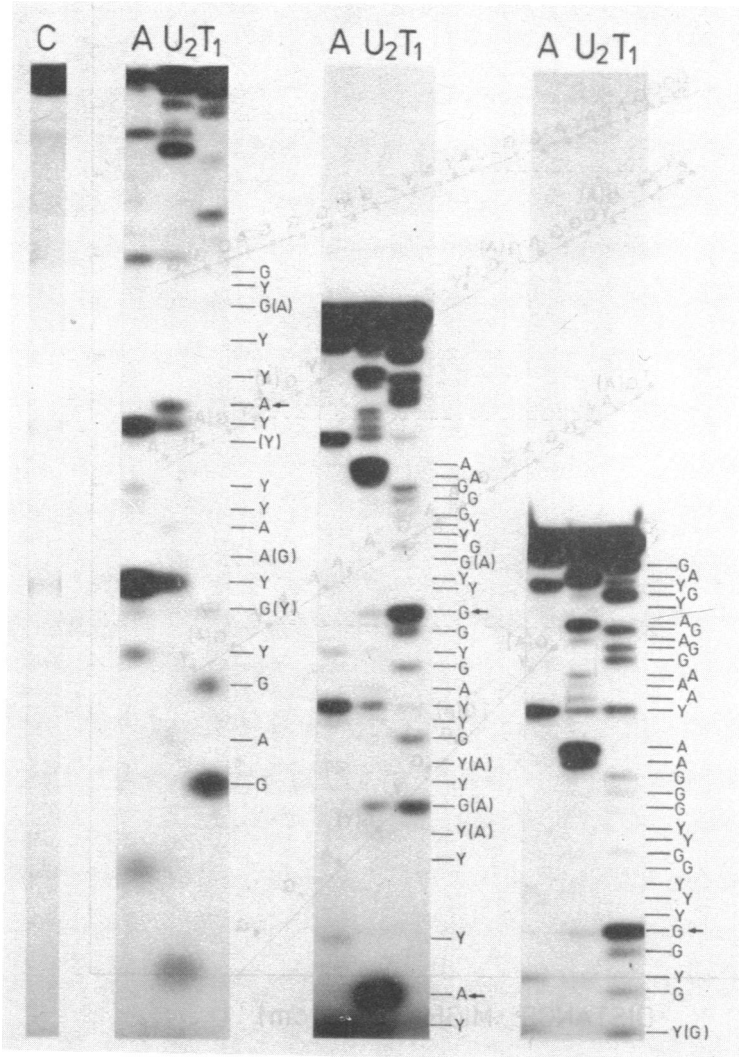

Figure 4: Sequence analysis of fragment "5" on 20% gels.

The first column shows a control slot minus enzyme, and the subsequent columns are sets of three hydrolysates as in Fig. 2. In this case only three gels were run, the longest of the usual four runs being omitted. The sequence is denoted as in Fig. 2. (Some very faint bands in the ribonuclease A slot are not visible in the photograph, but could be seen on the original autoradiograph).

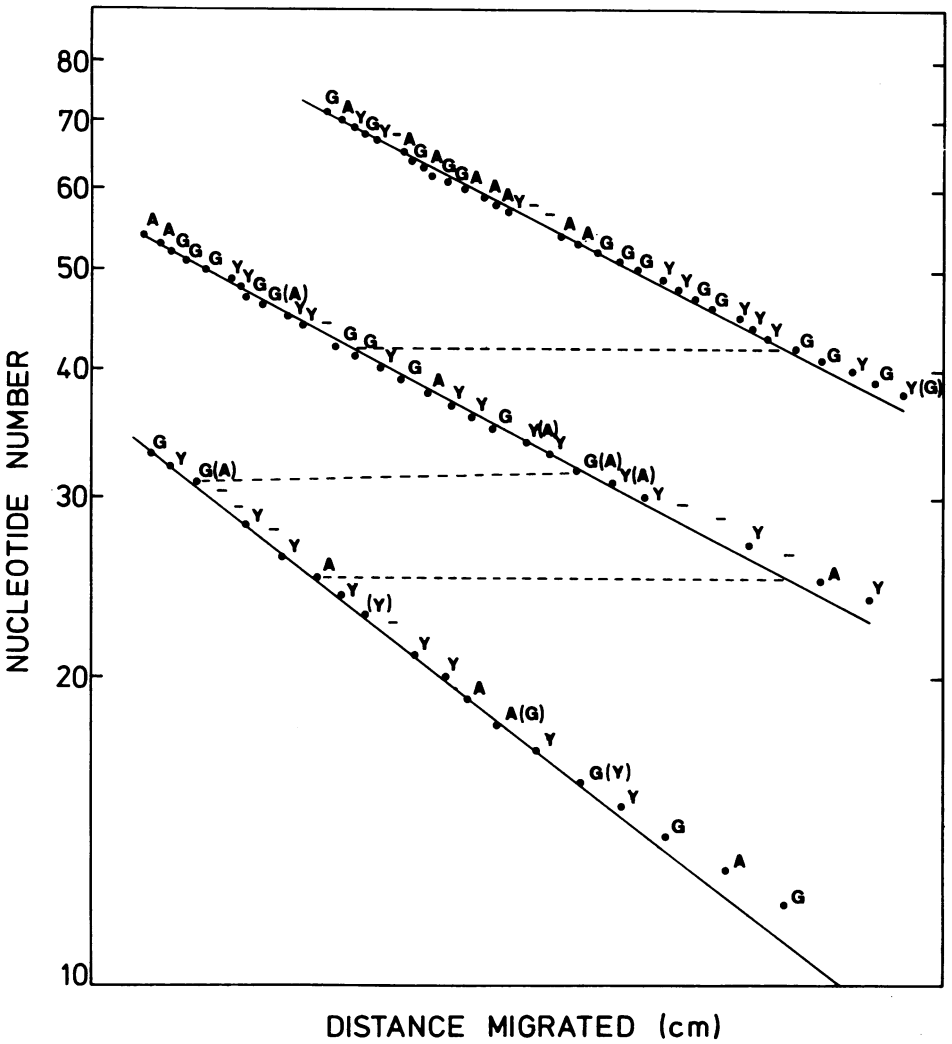

Figure 5: Logarithmic plot of sequence data from Fig. 4.

See legend to Fig. 3 for explanation. (There is a discrepancy of one base between the upper part of the first gel and the lower part of the second gel, in the region between the two dotted lines; the sequence from the second gel is taken as correct, being better resolved in this region).

from these experiments are shown in Fig. 6. In the case of fragment "5", a total of 62 bases could be read, starting from approximately the 12th residue, and continuing to the 3'-end

of the fragment, which is presumed to terminate in a guanine. This gives an excellent agreement with the estimated chain length of 75 for this fragment. In the case of fragment "A", 66 bases were read, starting from approximately the 10th residue, and it is clear from Fig. 2 that there are several bases between the last G-G-G-G sequence read and the 3'-terminus of the fragment, again giving a reasonable agreement with the estimated chain length of 90. For both fragments, the sequence deduced was reproducible in three independent experiments.

DISCUSSION

Fig. 6 compares the data which we have obtained here with the published sequences of the appropriate regions of 16S RNA (sections K and A (10)). In the case of fragment "A", there is a long stretch of 33 residues, starting with the C-A-A-A-A-G sequence in the middle of the region indicated, which is in complete agreement with the published sequence. The data also serve to show that the unlocated C-G dinucleotide cannot be in the left-hand position (between U-A-G and C-U-U-A-A-C-C-U-U-C-G), of the two possibilities suggested for it. However, on either side of this 33-nucleotide region, the agreement with the published sequence is poor. At the 3'-end, it may be that resolution has been lost on our fourth gel (Figs. 2 and 3) at the extreme high molecular weight end, although it is difficult to reconcile the "ladder" of four guanine residues (very clearly visible on a shorter exposure of the gel) near the 3'-end of the sequence (Fig. 2) with this conclusion. Further work will obviously be necessary to resolve this point.

On the other hand, at the 5'-end of the 33-nucleotide region there is a clear discrepancy with the published sequence (10). Although the U-G adjacent to the C-A-A-A-A-G sequence would fit our data, the best fit seems to be several nucleotides further to the 5'-end, as indicated by the lines drawn in Fig. 6. While it must be remembered that sequences of less than about 10 residues in length cannot be accurately read from 20 % gels (2), this region is in the 15-20 nucleotide range, which is reliable (cf. ref. 5, and cf. the sequence of fragment "5" discussed below). At all events there is

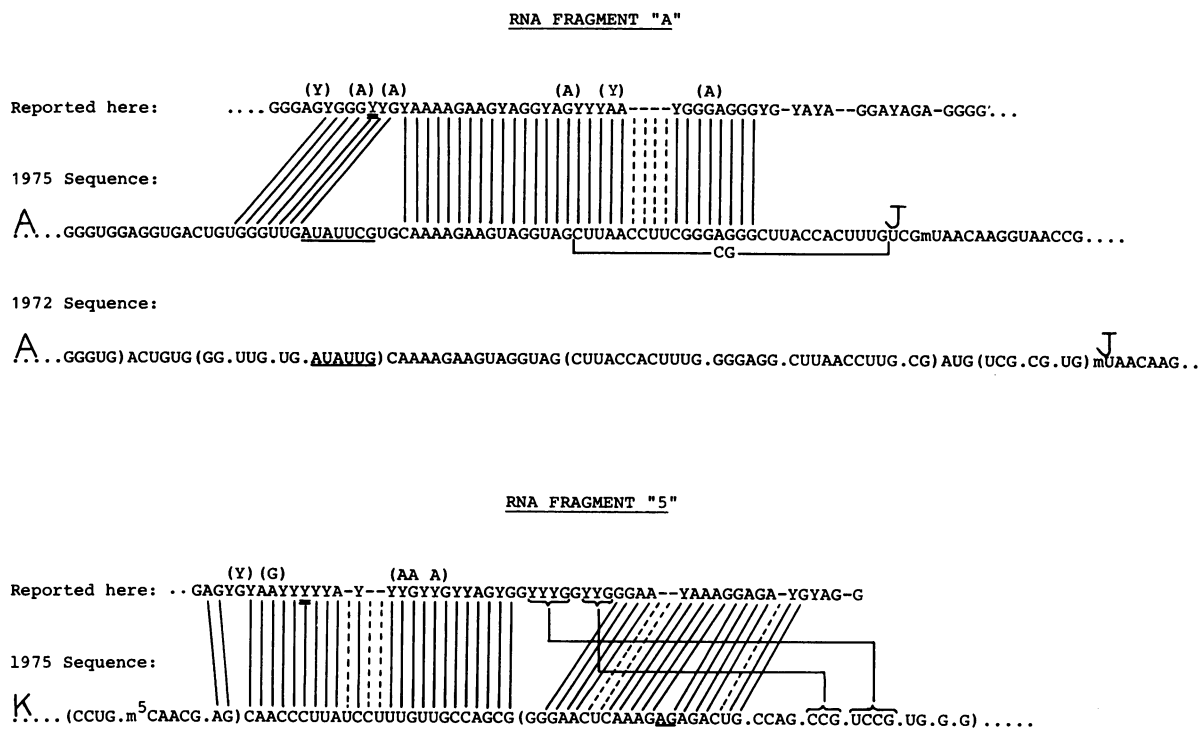

Figure 6. Comparison with the published sequence of 16S RNA.

The combined data from Figs. 2 and 3 (Fragment "A") are compared with the sequence of part of section A in the 1975 sequence (10), and the 1972 sequence of Ehresmann et al (9) is also included for comparison (see text). Similarly the combined data from Figs. 4 and 5 (Fragment "5") are compared with the sequence of part of section K from the 1975 sequence (10). Artefacts seen on the gels are indicated by the nucleotides in brackets above the sequence, and pyrimidine residues underlined twice are those which were only observed in other experiments with the same fragments. Gaps in our sequence are denoted by dashes, and correspondence with the published data is indicated by the vertical lines. Sequences read in a 5′ to 3′ direction.

no sign in our results of the intervening A-U-A-U-U-C-G sequence (underlined in Fig. 6). This latter oligonucleotide was written as A-U-A-U-U-G in the 1972 sequence (9), but according to Uchida et al (16) it is A-U-U-C-A-U-G, and Ehresmann et al (personal communication) now agree with this latter sequence. It is possible that an error in the sequencing of this oligonucleotide could have led to a false ordering of the other T_1 oligonucleotides in this region (compare the 1972 sequence (9), which is included in Fig. 6), if the ordering was based on a partial ribonuclease A digestion product purporting to originate from the middle of this oligonucleotide (10). Again, further work will be necessary to resolve this problem.

In the case of fragment "5", there is a stretch of 25 residues which are in complete agreement with the published sequence (10), starting with the long T_1 oligonucleotide C-A-A-C-C-C... (Fig. 6). Further towards the 5'-end there is an A-G which fits, although the intervening Y-G dinucleotide cannot be accounted for. At the 3'-end of the 25-nucleotide stretch, our data enable the T_1 oligonucleotides from this partially sequenced region to be ordered; it is clear that the U-C-C-G sequence comes first, followed by C-C-G, with two single guanine residues interspersed. Next comes the 21-nucleotide sequence starting with G-G-G-A-A-C, with which our data are in good agreement, apart from one G residue missing at the beginning, and one A residue missing from the middle of the region. The C-C-A-G sequence cannot be placed, but it could correspond to the Y-A-G sequence at the 3'-end of our fragment if resolution between the two pyrimidine residues was lost on our gel (Fig. 4). The last two bases at the extreme 3'-end of our fragment could also correspond to the U-G dinucleotide which Ehresmann et al (10) find in this region. Recently Ehresmann et al (personal communication) have been able to order this part of the 16S sequence, and their latest results are in complete agreement with our data for fragment "5" as presented in Fig. 6, with the exceptions of the Y-G dinucleotide at the 5'-end of our sequence (already mentioned) and the A-G (also already mentioned, and underlined in Fig. 6), for which we find only a guanine residue.

The number of artefacts observed in our experiments is acceptably low, the most common being "overcutting" by ribonuclease U_2 at guanine residues (see Figs. 2, 4 and 6, and cf. ref. 6). The "gaps" in the sequence appear to occur exclusively at pyrimidine positions, but despite this tendency, ribonuclease A does contribute significant positive information to the sequence in our hands. This is in contrast to the results of Donis-Keller et al (5), who found that ribonuclease A was too sequence specific to generate useful partial digestion products, although Simoncsits and Brownlee (6) did not have this difficulty. The very fact that ribonuclease A tends to give a few rather strong bands makes it important to include a ribonuclease A hydrolysis. This guards against the danger of artefacts, which could otherwise be easily introduced by even a very minor ribonuclease A contamination of the other enzymes or of the RNA substrate itself. The result of such a contamination would be that a pyrimidine could be misread as a guanine or adenine, and two such artefacts can be seen in Fig. 4 (cf. Fig. 6, fragment "5"). The combination of a ribonuclease A digest and a chemical hydrolysis to give a "ladder" as used by Simoncsits and Brownlee (6) is almost certainly the best solution at the present time.

At all events, it is clear that this sequencing method, or variations of it, will prove very useful in completing the analysis of the ribosomal RNA molecules. Preliminary data, covering a total of a further 150 nucleotides, have already been obtained in a similar way with a number of other fragments of the 16S RNA (Ross and Brimacombe, unpublished data).

Note added:

In a further personal communication, Ehresmann et al now find that their sequence in the region of the 5'-end of our fragment "A" (see Fig. 6) is in agreement with our data.

ACKNOWLEDGEMENTS

The authors are very grateful to Drs. B. Griffin, M. Szekely, G. Brownlee and H.G. Wittmann for helpful discussions of this work, and to Dr. C. Ehresmann for making available her latest information on the 16S RNA sequence.

REFERENCES

1 Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. 94, 441 - 448
2 Maxam, A.M. and Gilbert, G. (1977) Proc. Nat. Acad. Sci. USA 74, 560 - 564
3 Brownlee, G.G. and Cartwright, E.M. (1977) J. Mol. Biol. 114, 93 - 117
4 Gupta, R.C. and Randerath, K. (1977) Nucleic Acids Res. 4, 1957 - 1978
5 Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527 - 2538
6 Simoncsits, A. and Brownlee, G.G. (1977) Nature, 269, 833-836
7 Braun, R. and Behrens, K. (1969) Biochim. Biophys. Acta 195, 87 - 98
8 Contreras, R. and Fiers, W. (1975) Analyt. Biochem. 67, 319 - 326
9 Ehresmann, C., Stiegler, P., Fellner, P. and Ebel, J.P. (1972) Biochimie 54, 901 - 967
10 Ehresmann, C., Stiegler, P., Fellner, P. and Ebel, J.P (1975) Biochimie 57, 711 - 748
11 Yuki, A. and Brimacombe, R. (1975) Eur. J. Biochem. 56, 23 - 34
12 Rinke, J., Ross, A. and Brimacombe, R. (1977) Eur. J. Biochem. 76, 189 - 196
13 Morgan, J. and Brimacombe, R. (1972) Eur. J. Biochem. 29, 542 - 552
14 Rinke, J., Yuki, A. and Brimacombe, R. (1976) Eur. J. Biochem. 64, 77 - 89
15 Richardson, C.C. (1965) Proc. Nat. Acad. Sci. USA 54, 158 - 165
16 Uchida, T., Bonen, L., Schaup, H.W., Lewis, B.J., Zablen, L. and Woese, C. (1974) J. Mol. Evolution 3, 63 - 77