Volume 9 Number 10 1981 Nucleic Acids Research

The sequence of the ribosomal 16S RNA from Proteus vulgaris. Sequence comparison with E. coli 16S RNA and its use in secondary structure model building

P.Carbon+, J.P.Ebel and C.Ehresmann

Laboratoire de Chimie Biologique, Institut de Biologie Moleculaire et Cellulaire du CNRS, 15, rue R.Descartes, 67000 Strasbourg, France

Received 27 March 1981

ABSTRACT

The complete nucleotide sequence of the 16S RNA from *Proteus vulgaris* has been determined. The molecule (1544 nucleotides) shows 93% homology with the sequence of *E.coli* 16S RNA. Six methylated nucleotides have been localized in positions homologous to those observed in the *E.coli* RNA molecule. Both *E.coli* and *P.vulgaris* 16S RNA chains can be folded up into a common secondary structure scheme. Comparative sequence analysis of the two molecules has provided a valuable contribution to 16S RNA secondary structure model building.

INTRODUCTION

One of the more powerful criteria used for the study of the higher order structure of RNAs is undoubtedly the phylogenetic comparison of sequence data. A significant example is provided by tRNAs and 5S RNAs for which secondary structures have been deduced only after comparative sequence analyses. Woese *et al.* (1) have recently made full use of the catalogs of T_1 RNase-generated oligonucleotides from over one hundred true bacteria as well as a partial sequence of *Bacillus brevis*, to derive a secondary structure model for the *E.coli* 16S RNA. In this paper we present the complete sequence of the ribosomal 16S RNA from an enterobacteria *Proteus vulgaris*, determined by the rapid sequencing gel electrophoresis methods . *Proteus vulgaris* was chosen because previous work in our laboratory showed the existence of a high degree of homology between the sequences of *E.coli* and *P.vulgaris* 16S RNAs. This conclusion was based on the comparative analysis of the T_1 RNase-generated oligonucleotides (2). The 16S RNA from *P.vulgaris* is also able to bind *E.coli* S4 protein with a high affinity (3). In such closely related sequences, compensatory changes should be detected very easily and thus provide helpful information in testing secondary structure models.

MATERIALS AND METHODS

1. Preparation of the RNA fragments

The strain of *P. vulgaris* was kindly provided by Dr. Monteil. In order to obtain fragments covering the entire molecule, two different enzymatic partial hydrolyses were used. In all cases the 16S RNA molecule was uniformly ^{32}P-labeled at low specific activity in order to facilitate isolation and purification of fragments, ^{32}P was incorporated as described in (4).

RNA fragments from 12S and 8S RNAs. The 16S RNA was prepared as described in (5) and digested with T_1 RNase (Sankyo) in the reconstitution buffer using the conditions previously described for *E. coli* 16S RNA (6). The resulting 12S and 8S RNAs were fractionated on a sucrose gradient in the reconstitution buffer. The subfragments were dissociated on preparative polyacrylamide slab gels (40 x 30 x 0.4 cm) containing 8M urea (6), 10% and 15% composite slab gels were used.

RNA fragments from 30S subunits treated with cobra venom ribonuclease. The 30S subunits resuspended in 20 mM Tris HCl (pH 7.5), 10 mM Mg acetate, 60 mM NH_4Cl, 6 mM 2-mercaptoethanol were preincubated for 30 min at 37°C and chilled on ice. The ribonuclease extracted from the cobra *Naja oxiana* (7) was added at an enzyme to substrate ratio of 1 unit : 1 mg 30S subunits. Incubation was carried out at 4°C for 10 min and 1h. The digest was treated with phenol in the presence of 0.1% dodecylsulfate and the extracted RNA was precipitated with ethanol. The RNA subfragments were fractionated on preparative slab gels (40 x 30 x 0.4 cm) containing 8M urea, 6% and 10% composite slab gels were used. In each case the RNA fragments were repurified by an additional electrophoresis cycle and eluted according to the procedure described in (8).

2. Dephosphorylation of the RNA fragments

The fragments yielded by T_1 RNAse and cobra venom ribonuclease have phosphate groups, respectively, at their 3'- or 5'-ends and they therefore required dephosphorylation prior to labeling at their 3'- or 5'-extremities. *E. coli* alkaline phosphatase (Boehringer) was used in the presence of 0.1% dodecylsulfate in order to inhibit the endonuclease activity present in the enzyme preparation, as described in ref. (9).

3. 5' and 3'-terminal labeling of the RNA fragments

The RNA fragments were labeled at their 5'-terminus using T_4 polynucleotide kinase in the presence of 100-200 μCi of {γ-^{32}P} ATP (Amersham 3000 Ci/mole) as described in (10). The RNA fragments were labeled at the 3'-terminus using T_4 RNA ligase (P.L. Biochemicals) in the presence of {5'-^{32}P}-pCp (Amersham,

2000-3000 Ci/mole) in the conditions described in (11). The labeled fragments were then fractionated in polyacrylamide slab gels (40 x 30 x 0.05 cm) containing 8M urea.

4. Sequencing procedure

The sequencing of each 3'-labeled fragment was analyzed using the chemical cleavage technique described by Peattie (12). Parallel determination using the enzymatic cleavage methodology of Donis-Keller et al. (13) were also made in most cases. The sequence of several fragments has also been analyzed from the 5'-extremity by the enzymatic sequencing technique. In this latter case, hydrolysis with the RNase from *Neurospora crassa* (Boehringer) was also used to discriminate between pyrimidines (14). The polyacrylamide slab gels (40 x 30 x 0.05 cm) were pre-electrophoresed and the electrophoresis was carried out at 2000 V. Running time and gel concentration were chosen according to the length of the fragments to be analyzed. The sequences of nucleotides 1-10, 10-20, 30-50, 100-150 and 150-200 (from the labeled end) were read, respectively, on 25%, 20%, 15%, 10% and 8% polyacrylamide gels.

RESULTS AND DISCUSSION

1. Sequencing strategy

Fragments covering the entire molecule were obtained by the two different enzymatic digestions as described in Material and Methods. We were largely aided in this purpose by the fact that the *P.vulgaris* and *E.coli* patterns of hydrolysis were very similar. In particular it was possible to prepare by mild T_1 RNase digestion the two well defined 12S and 8S RNAs. In order to rapidly select the useful fragments, an aliquot of each fragment was labeled either at its 5'-end (in the case of T_1 RNase fragments) or at its 3'-end (in the case of cobra venom RNase fragments) and the first 10 to 20 nucleotides were sequenced using the technique of Donis-Keller *et al.* (13). It was thus generally possible to localize the analyzed fragments by reference to the *E.coli* 16S RNA sequence (10,15). Fragment sizes were estimated by their electrophoretic mobilities compared to known *E.coli* 16S RNA fragments used as position markers. The selected RNA fragments were then dephosphorylated if necessary (in the case of T_1 RNase fragments) and labeled at their 3'-extremity. Each fragments was usually submitted to both the chemical (12) and enzymatic procedures (13) for sequence determination. A diagram of the various RNA fragments that have been analyzed is given in Fig. 1, with the sequenced sections and the direction of sequence determination indicated.

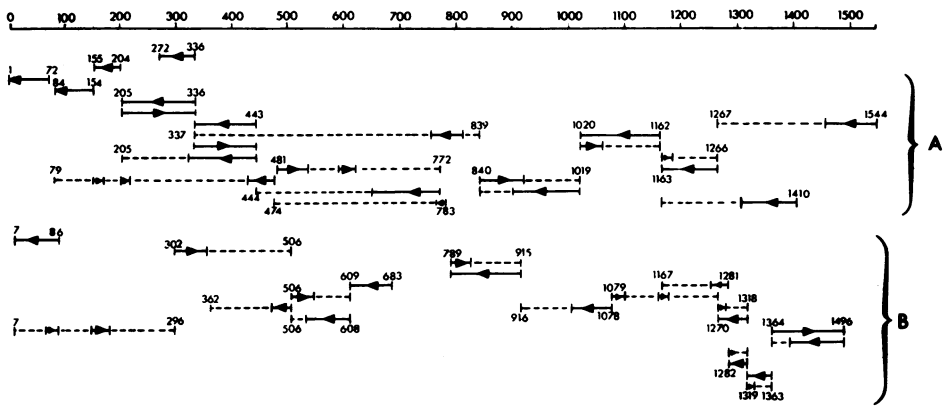

Figure 1 : Schematic map of the RNA fragments used in the sequencing of the 16S RNA from *P.vulgaris*. (A) Fragments obtained with cobra venom ribonuclease ; (B) fragments obtained with T_1 ribonuclease. The precise cleavage positions are indicated. The solid lines indicate the part of the sequence which has been sequenced and the direction of sequence determination is shown by arrows.

2. *Sequencing results*

The deduced sequence is shown in Fig. 2. The molecule contains 1544 nucleotides. It has been previously reported (2) that total T_1 RNase digestion yields seven oligonucleotides containing methylated nucleotides : m^2G, A-U-m^2G, m^5C-A-A-C-G, C-C-m^7-G-C-G, m^4C_m-C-C-G (and not m^4C_m-C-C-C-G as reported at that time), mU-A-A-C-A-A-G and m^6_2A-m^6_2A-C-C-U-G.

We localized six nucleotides whose reactivity towards chemical modification or nuclease cleavage strongly suggests the presence of modification. Some of these modifications are easily identifiable : m^7G is totally cleaved after reduction and aniline treatment, m^5C is not reactive to hydrazine modification (8), and 2'-O methyl renders the adjacent phosphodiester bond resistant to any cleavage involving the formation of a 2'-3' cyclic intermediate. All the nucleotides which exhibit an atypical behaviour have been found localized in the above mentioned sequences and most likely correspond to the reported methylated nucleotides. Furthermore, they are localized in positions homologous to those observed in *E.coli* 16S RNA. Nevertheless, it was not possible to detect the two m^2G-residues which exist in the *P.vulgaris* RNA, since they are indistinguishable from unmodified G. The C-residue at position 1409 in *P.vulgaris* 16S RNA is not methylated whereas the analogous nucleotide is present as m^5C in *E.coli* RNA. In addition, the residue at position 1518 shows a markedly reduced sensitivity to T_1 RNase and an ambiguous reactivity to dimethylsulfate and

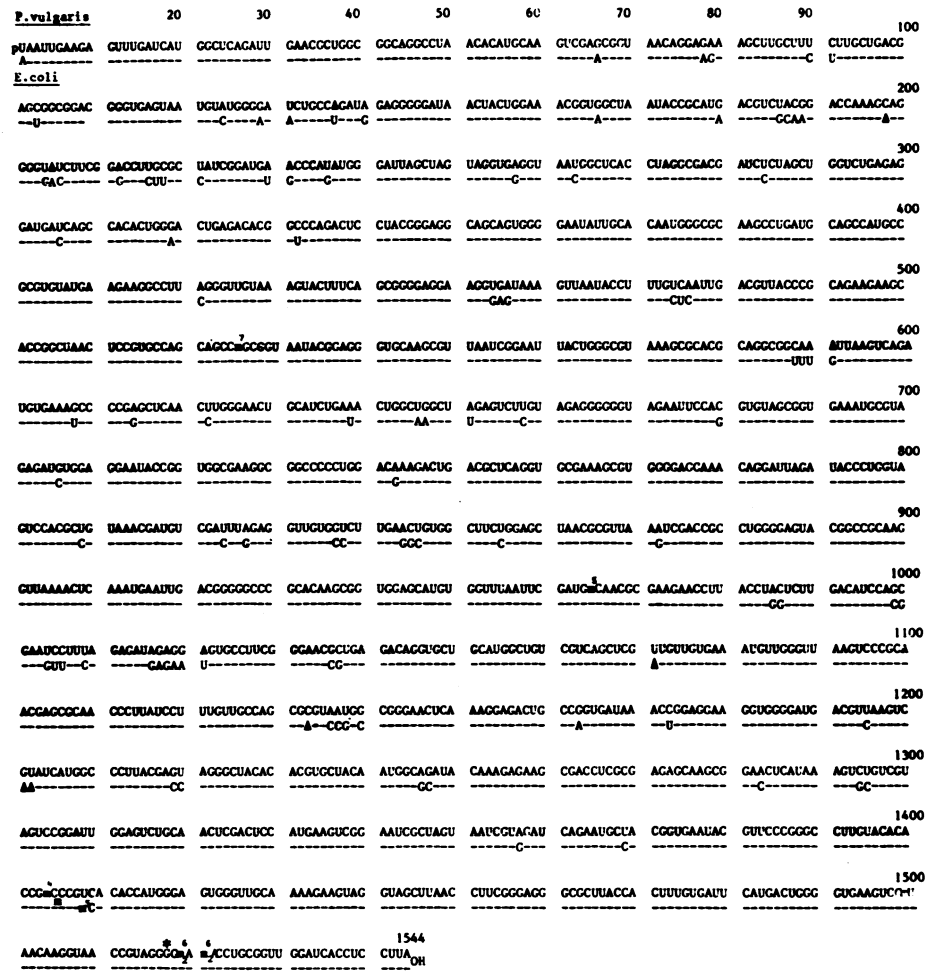

Figure 2 : The nucleotide sequence of the 16S RNA from *P.vulgaris*. The *P.vulgaris* sequence (top line) is compared with the *E.coli* sequence (10,15). Homologous nucleotides are denoted by dashes, the base replacements are indicated. (▲) symbolizes a deleted position.

diethylpyrocarbonate. We believe that this residue is therefore a modified G, but the nature of the modification is still unknown.

3. *Sequence homology*

The sequence of *P.vulgaris* 16S RNA shows 93% homology with the sequence of *E.coli* 16S RNA. One hundred and four nucleotides have been found different, 96

of them are point-variations but 3 residues are "deleted" and 5 "insered" when compared to the *E.coli* 16S RNA (see Fig. 2). These differences are not distributed uniformely along the RNA molecule. Variations are observed to a higher degree in 5 distinct areas of the molecule : nucleotides {120 to 140} : 25% ; {180 to 230} : 25% ; {585 to 660} : 16% ; {820 to 850} : 23% and {980 to 1040} : 27%. On the other hand, nucleotides {2 to 60}, {340 to 420}, {480 to 580}, {750 to 800}, {880 to 980}, {1050 to 1130}, {1300 to 1350} and {1370 to 1544} are unchanged. Woese *et al.* (16) has already suggested the existence of phylogenetically conserved regions among bacteria. There is a good correlation between the distribution of the conserved regions defined by Woese and coworkers and our results.

4. Secondary structure

A concensus on the folding of *E.coli* 16S RNA is now emerging (1,17,18,19,20). The secondary structure model we have proposed was essentially constructed on the basis of topographical data and comparative sequence analysis (19,20). Due to the large amount of homology between their sequences, both 16S RNA molecules from *E.coli* and *P.vulgaris* are expected to be folded into nearly identical secondary structure schemes. Therefore, the limited number of nucleotide changes observed in the *P.vulgaris* 16S RNA sequence could be easily correlated with compensing changes which allow the strict conservation of precise secondary structure motifs in both molecules. This is illustrated in Fig. 3 which shows the secondary structure model of the *E.coli* 16S RNA (19,20) and indicates the positions of the nucleotide replacements in the *P.vulgaris* sequence. Thus, a number of helices (indicated by heavy lines) are supported by compensating changes (only double base changes have been considered). These motifs have also been proposed independently by other authors (1,17,18). Nevertheless, one ambigous point remains in region {170-220} in which the changed nucleotides could not be correlated with other compensatory modifications. Nucleotides C-216, U-217 and U-218 are changed in *P.Vulgaris* to U, G and C, respectively. These changes are not compensated in the opposite strand of the stem proposed in Fig. 3. The addition of a C-residue between residues G-197 and A-198 is also not compensated by a coordinated change in the opposite strand, and the proposed helix is therefore weakened. These two stems, written in italics in Fig. 3, were only proposed as tentative structures for which alternative possibilities were not excluded (19,20). Neither were more satisfactory solutions proposed in the other models. Identical stems were proposed by Glotz and Brimacombe (17) and Brimacombe (18). Also, the helix proposed by Woese *et al.* (1) is not supported by coordinate nucleotide changes in *P.vulgaris*. The ques-

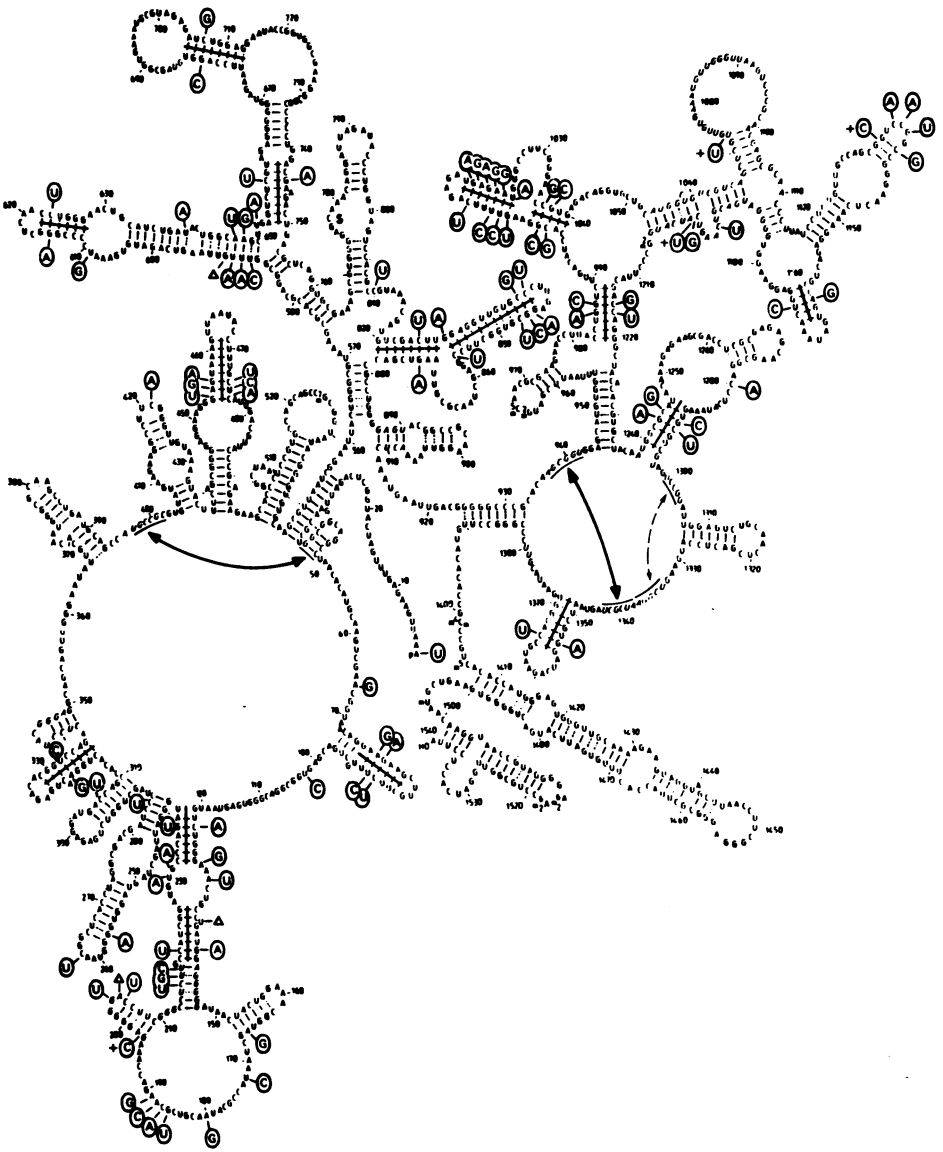

Figure 3 : Secondary structure model of the *E.coli* and *P.vulgaris* 16S RNAs. The secondary structure model of the *E.coli* 16S RNA is given (19,20) and base replacements in *P.vulgaris* 16S RNA are indicated (0). Nucleotide deletions are denoted by (Δ) and base addition by (+). The helices where compensating base changes exist between the two 16S RNA molecules are indicated by heavy lines. Double arrows indicate additional base bairing (20).

tion remains open whether any of the secondary structure models proposed for this region of *E.coli* RNA are correct or if the 16S RNA from *P.vulgaris* has a slightly different conformation in this region. The high variability in terms of size, nucleotide sequence and secondary structure, of this region in the other rRNA molecules sequenced so far, must be noted (manuscript in preparation). A specific conformation may therefore exist in each RNA species.

About 71% of the variant nucleotides have been found localized in double-stranded regions. This finding corroborates the observation of Woese *et al.*(1) who have noted the extreme constancy of sequence in non-paired regions.

ACKNOWLEDGMENTS

Dr. R. Rhoads is thanked for careful reading of the manuscript. We thank B. Muller for her excellent technical assistance. This work was supported by the Centre National de la Recherche Scientifique and by the Délégation Générale à la Recherche Scientifique et Technique (contrat n° 78.7.1088).

+Present address : Laboratoire de Chimie Biologique, Université Libre de Bruxelles, 67, rue des Chevaux, 1640 Rhode St.Genèse (Belgium).

REFERENCES

1. Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J. and Noller, H.F. (1980) Nucleic Acids Res. 8, 2275-2293.
2. Fischel, J.L., Krol, A., Ehresmann, C., Fellner, P. and Ebel, J.P. (1975) Biochimie 57, 885-897.
3. Daya-Grosjean, L., Geiser, M., Stöffler, G. and Garrett, R.A. (1973) FEBS Lett. 37, 17-20.
4. Branlant, C., Krol, A., Machatt, A. and Ebel, J.P. (1981) Nucleic Acids Res. 9, 293-307.
5. Fellner, P. (1969) Eur. J. Biochem. 11, 12-27.
6. Ehresmann, C., Stiegler, P., Fellner, P. and Ebel, J.P. (1975) Biochimie 57, 711-748.
7. Vassilenko, S.K. and Ryte, V.C. (1975) Biokhimia 40, 578-583.
8. Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci., USA, 74, 560-564.
9. Shinagawa, M. and Padmanabhan, R. (1979) Anal. Biochem. 95, 458-464.
10. Carbon, P., Ehresmann, C., Ehresmann, B. and Ebel, J.P. (1978) FEBS Lett. 94, 152-156.
11. England, T.E. and Uhlenbeck, O.C. (1978) Nature 275, 560-561.
12. Peattie, D.A. (1979) Proc. Natl. Acad. Sci, USA, 76, 1760-1764.
13. Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527-2538.
14. Krupp, G. and Gross, H.J. (1979) Nucleic Acids Res. 6, 3481-3490.
15. Brosius, J., Palmer, M.L., Kennedy, P.L. and Noller, H.F. (1978) Proc. Natl. Acad. Sci., USA, 75, 4801-4805.
16. Woese, C.R., Fox, G.E., Zablen, L., Uchida, T., Bonen , L., Pechman, K., Lewis, B.J. and Stahl, D. (1975) Nature 254, 83-86.
17. Glotz, C. and Brimacombe, R. (1980) Nucleic Acids Res. 8, 2377-2395.
18. Brimacombe, R. (1980) Biochem. Int. 1, 162-171.

19. Stiegler, P., Carbon, P., Zuker, M., Ebel, J.P. and Ehresmann, C. (1980) C.R. Acad. Sci., Paris, Série D, 291, 937-940.
20. Stiegler, P., Carbon, P., Zuker, M., Ebel, J.P. and Ehresmann, C. (1980) submitted to Nucleic Acids Res.