Volume 11 Number 5 1983 Nucleic Acids Research

Sequences of the 5S rRNAs of *Azotobacter vinelandii, Pseudomonas aeruginosa* and *Pseudomonas fluorescens* with some notes on 5S RNA secondary structure

Erna Dams, Antoon Vandenberghe and Rupert De Wachter*

Departement Celbiologie, Universiteit Antwerpen, Universiteitsplein 1, B-2610 Wilrijk, Belgium

Received 17 January 1983; Revised and Accepted 15 February 1983

ABSTRACT

Recently published alignments of available 5 S rRNA sequences have shown that a rigid base pairing pattern, pointing to the existence of a universal five-helix secondary structure for all 5 S RNAs, can be superimposed on such alignments. For a few species, the alignment and the base pairing pattern show distortions with respect to the large majority of sequences. Their 5 S RNAs may form exceptional secondary structures, or there may just be errors in the published sequences. We have examined such a case, *Pseudomonas fluorescens*, and found the sequence to be in error. The corrected sequence, as well as those of the related species *Azotobacter vinelandii* and *Pseudomonas aeruginosa*, fit perfectly in the 5 S RNA sequence alignment and in the five-helix secondary structure model. There exists comparative evidence for the frequent presence of non-standard base pairs at several points of the 5 S RNA secondary structure.

INTRODUCTION

5 S rRNA is becoming one of the most intensively studied macromolecules, with about one hundred sequences appearing in a recent review[1], and several additional ones being published each month. These data can be used for evolutionary studies[2,3] and for the investigation of secondary structure on a comparative basis. Although an early proposal[4] for 5 S RNA secondary structure stressed the similarity between eukaryotic and bacterial molecules, many authors subsequently favoured a four helix model for bacterial 5 S RNAs[5,6,7] as opposed to a five helix model for eukaryotic 5 S RNAs[8,9,10]. We have shown[11] that a careful alignment of available sequences, as well as energetic considerations, point to the presence of 5 helical areas not only in eukaryotic and eubacterial 5 S RNAs, but also in those present in archaebacterial and organelle ribosomes. The idea that all 5 S RNAs fold essentially according to the same pattern, with minor differences among large taxa, now seems to be gaining ground again[12,13].

Some of the initial confusion about 5 S RNA secondary structure can probably be traced to the presence of sequencing errors in the limited col-

lection of data that was available for the first comparisons[5,14]. A notorious case is the *Chlorella pyrenoidosa* sequence, originally derived[15] by fingerprinting methods on uniformly labeled RNA, which was thoroughly rearranged after reexamination[9] by gel sequencing of terminally labeled RNA. This is not a unique case, however. Five 5 S RNA sequences[1] have been corrected upon a second examination and other sequencing errors may remain undetected. We previously pointed out[11] that the sequences published for *Pseudomonas fluorescens*[16] and for *Thermus aquaticus*[17] can only be fitted in a 5 S RNA alignment under the assumption that a series of unique insertions and deletions distinguish them from other bacterial sequences. Also, secondary structure models can only be drawn for these species at the expense of distortions with respect to the uniform model[11]. We have reexamined and corrected the sequence of the 5 S RNA of *Pseudomonas fluorescens* and determined those of the related species *Pseudomonas aeruginosa* and *Azotobacter vinelandii*. The three sequences align very well with those of other organisms and can be folded in the standard five helix secondary structure model[11].

MATERIALS AND METHODS

Azotobacter vinelandii, strain NCIB 8789, was cultured at 30°C on 1.5% agar plates containing the growth medium described by Sweeney[18]. *Pseudomonas fluorescens* ATCC 13525 was grown at 27°C and *Pseudomonas aeruginosa* CCEB 481 (Culture Collection of Entomogeneous Bacteria, Prague; identical to type strain ATCC 10145) was grown at 37°C, both in liquid LB medium[19]. After harvesting of the cells, ribosomes were prepared by a simplified version of Kurland's[20] procedure. The number of fractional $(NH_4)_2SO_4$ precipitations was reduced from three to one, a single ultracentrifugation was performed to pellet the ribosomes, and the dialysis step was omitted. The 5 S RNA was obtained from the ribosomes as previously described[21]. As a rule, some 5 OD_{260} units of 5 S RNA were necessary for the sequencing experiments, and this corresponded with about 300 OD_{260} units of ribosomes obtainable from 3 g of cells.

Each of the three bacterial 5 S RNA preparations proved to be a mixture of RNAs with the same sequence, but different in length, and therefore special precautions had to be taken before sequencing gels were run. After labeling with [5'-^{32}P]pCp at the 3'-terminus, the undegraded RNA was fractionated on an 80 cm long 8% polyacrylamide gel. This yielded two bands in the case of *A. vinelandii* and *P. aeruginosa* and three main bands in the case of *P. fluorescens*. An autoradiogram is shown in Fig. 1. Peattie's[22] chemical degrada-

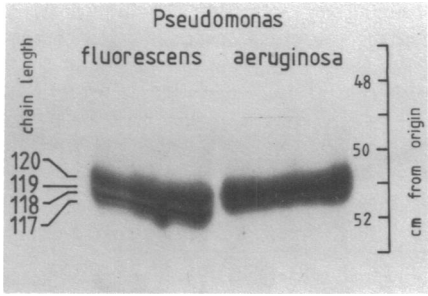

Fig. 1. Length heterogeneity in the 5 S RNA preparations. 5 S RNA labeled with ^{32}P at the 3'-terminus was run on high resolution polyacrylamide gels as used for sequencing. This gave two bands in the case of *A. vinelandii* (not shown) and *P. aeruginosa* and four in the case of *P. fluorescens*, with the chain lengths indicated. Each component was independently sequenced, except the shortest one from *P. fluorescens* which was present in too low a concentration. The scale shows the distance migrated from the origin.

tion method was then applied to each extracted band, which allowed to read the sequence except for the 5'-end group. Each of the three 5 S RNAs proved to contain components of chain length 120 and 119, the latter missing one nucleotide at the 3'-end. *P. fluorescens* 5 S RNA also contained a component of chain length 118, missing a nucleotide at both termini, and a very small amount of chain length 117 material, which was not sequenced.

The 5'-end groups were identified and the adjacent sequences confirmed by extending the RNAs with a pentaadenylic acid segment as follows. From each 5 S RNA preparation 10 μg was incubated overnight at 4°C with 10 μg $(Ap)_4A$ (Miles, Elkhart, IN, U.S.A.) and 5 units RNA ligase in 5 μl of a solution containing 1 mM ATP, 50 mM HEPES-NaOH (pH 8), 20 mM $MgCl_2$, in 20% DMSO. The elongated RNA was separated from the unmodified fraction by electrophoresis on a thin 8% polyacrylamide gel, detected by staining with 0.1% toluidine blue, and extracted. The oligo A "leader" was then 5'-terminally labeled with [γ-^{32}P]ATP in the presence of polynucleotide kinase. Individual components of different chain length were separated as in the case of 3'-terminally labeled RNA. Gel electrophoresis after partial hydrolysis with RNases PhyM, T_1, U_2, Chicken liver 3, and *B. cereus*, revealed the 5'-terminal sequences. The RNases were obtained in a kit from B.R.L. (Gaithersburg, MD, U.S.A.) and used under the conditions advised by the manufacturers.

RESULTS AND DISCUSSION

The sequence of 120 nucleotides forming the longest 5 S RNA component in each of the three bacteria is shown in Fig. 2a. The heterogeneity found at the termini is defined in Fig. 2b, which illustrates how the different components can result from ambiguity in the processing of the ribosomal RNA precursor. Each component bears a 5'-terminal phosphate group and has an un-

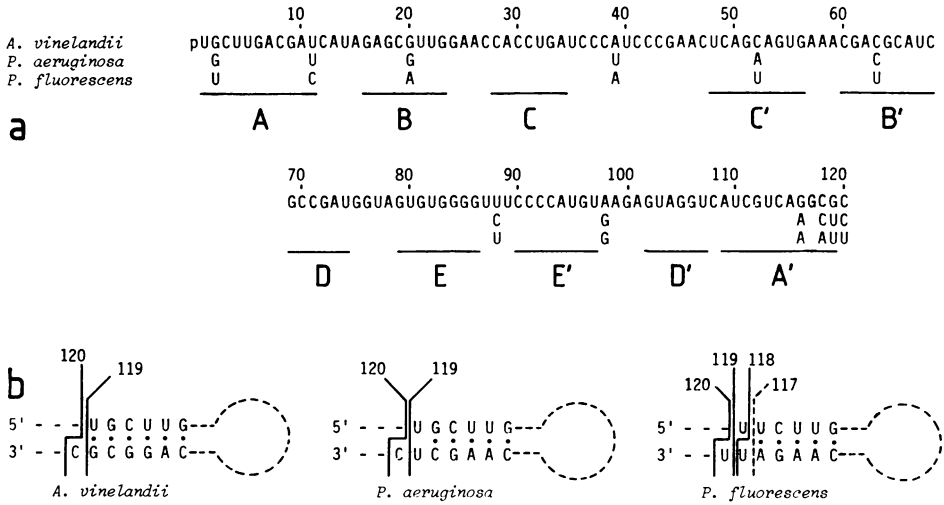

Fig. 2. Primary structure of the three 5 S RNAs.

a) Sequences of the 120 nucleotide components. The bases that are identical in the three species are written only once. The five double-stranded areas of the secondary structure model (Fig. 3) are labeled A-A' to E-E' in the order of appearance from 5'- to 3'-end.

b) Alternative processing schemes of ribosomal RNA precursor in base-paired area A-A' leads to 120 and 119 nucleotide components in the three species, plus a 118 nucleotide component in *Pseudomonas fluorescens*. The latter species contains a small amount of 117 nucleotide component (not sequenced), possibly having the termini indicated by the dotted line.

phosphorylated 3'-terminus, which is proven by the fact that $(Ap)_4A$ and pCp can be ligated at the 5'- and 3'-ends, respectively. This excludes the possibility that the shorter components would be generated by a contaminating RNase, which would yield 3'-phosphorylated termini.

A sequence for *Pseudomonas fluorescens* 5 S RNA, derived by complete and partial digestion of uniformly labeled material with pancreatic and T_1 RNases, has been formerly published[16]. It has exactly the same base composition as the sequence shown in Fig. 2a, but contains the following errors. Nine RNase T_1 oligonucleotides were assigned to a wrong position, viz. those occupying positions 10 to 18, 52 to 54, and 97 to 117 in the correct sequence. Moreover, errors were committed in the sequencing of the oligonucleotides now occupying positions 1-6, 25-33, 107-112, and 118-120. The same report[16] also mentions G/U_7 and G/A_{19} heterogeneity (numbering refers to the incorrect sequence). We did not detect any positional, but only length heterogeneity.

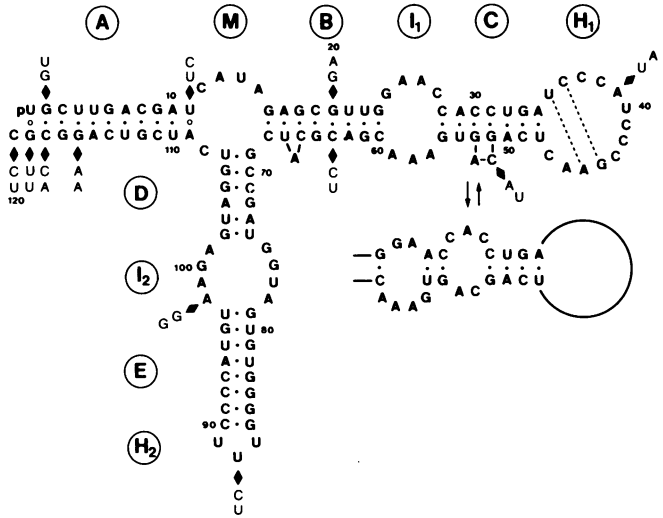

Fig. 3. Secondary structure model for the three 5 S RNAs.
The sequence of *A. vinelandii* is fitted in the universal 5-helix model[11] with the bulge-internal loop equilibrium always possible in area I_1-C. Positions not identical in the three species are indicated by a losenge, with the base found in *P. fluorescens* (outermost character) and *P. aeruginosa* (innermost character) shown. Base pairs that are non-standard (see text) in at least one of the *Pseudomonas* 5 S RNAs are indicated by circles. Broken lines in loop H_1 connect adjacent base pairs possible in most, but not all, 5 S RNAs.

In the case of *Pseudomonas aeruginosa*, an oligonucleotide catalog, but no sequence, has previously been published[23]. We found the RNase T_1 catalog to be correct except for sequencing errors in three oligonucleotides and incomplete data on two others.

Fig. 3 represents the *Azotobacter vinelandii* sequence in the five helix-model previously shown[11,1] to be adaptable to all hitherto sequenced 5 S RNAs. The equilibrium structure in area I_1-C is also universal[11,21], which we take as evidence that the position or function of 5 S RNA within the ribosome requires flexibility. Helix D contains the structure $G_{72}{\cdot}A_{104}$ which is observed at this position in many eubacterial RNAs. Originally[24], only G·C and A·U base pairs were thought to occur in RNA secondary structures. Since the formulation of the wobble hypothesis[25], the existence of G·U pairs is also admitted. Comparative studies on tRNA[26], 5 S RNA[11], and large rRNA[27] secondary structure are now making it increasingly likely that base pairs other than G·C, A·U and G·U occur occasionally within, or even at the ter-

mini of helices. The terms odd-, non-complementary-, and non-standard base pair have been used[26,27] to describe these structures. There is as yet no experimental evidence that the bases in such pairs form hydrogen bonds, or even that they form stacks with the neighbouring regular base pairs. However, the fact that non-standard pairs are observed within the stems of several tRNA cloverleafs would suggest that their presence does not disrupt the helix structure, at least if some universality for the tRNA spatial structure[28] is presumed. Structural models have been proposed[27,29] for A·C pairs with one of the bases in the imino form, and for purine-purine pairs with one or both bases in the rare tautomeric form or syn configuration. These structures are thought to form transiently in DNA undergoing spontaneous mutation[29] and it is conceivable that they can exist permanently at certain points in RNA secondary structures. Helix A consists of 11 base pairs in *A. vinelandii*, but the $U_{11}{\cdot}A_{109}$ changes to C·A in *P. fluorescens*, and the $U_1{\cdot}G_{119}$ becomes U·U in both *Pseudomonas* species. The conservative interpretation would be that helix A is shortened at the expense of a larger M-loop and unpaired termini in the latter species. An alternative possibility is that the helix length is maintained by the existence of non-standard base pairs. The pyrimidine pairs U·U and U·C are also found frequently[11,1] (but not in the species examined here) within helix E of bacterial 5 S RNAs.

The existence of tertiary interactions between loops H_1 and I_2 has been postulated by different investigators. When transposed to the sequence of *A. vinelandii*, the proposals consist of a parallel interaction[12] between CAU_{40} and GUA_{78}, antiparallel base pairing[30] between $CCAU_{40}$ and $AUGG_{76}$, or antiparallel pairing[7] between $(C)CCG_{44}$ and $UGG(U)_{77}$. The three possibilities are mutually exclusive and it is difficult to deduce which, if any, is the most likely. Because the sequences involved are rather conserved, the complementarities can be found, in whole or in part, in most 5 S RNAs. However, if there is a change in one of the sequences, one rarely observes a compensating substitution in the putative complement to support the existence of the interaction, as is the case for the secondary structure model.

Some 32 eubacterial 5 S RNA sequences are known at present. Apart from those reported here, the following four can be folded into a secondary structure model that is completely or nearly identical to the one represented in Fig. 3 : *Escherichia coli*, *Proteus vulgaris*, *Photobacterium phosphoreum*, and *Beneckea harveyi*. In all these cases a $C_{11}{\cdot}A_{109}$ non-standard pair is found at the end of helix A. Moreover, the $G_{69}{\cdot}U_{107}$ pair seen at the M-loop side of

helix D (Fig. 3), is substituted by G·G in *E. coli* and *P. vulgaris*, by G·A in *P. phosphoreum* and *B. harveyi*. The seven species with similar secondary structure pattern also formed a cluster in a phylogenetic tree constructed from some 140 presently known 5 S RNA sequences (unpublished work from our laboratory). This clustering is in agreement with the prokaryote phylogeny proposed by Fox et al.[31], where these species form one of the four branches of the taxon designated as purple photosynthetic bacteria and relatives.

ACKNOWLEDGEMENTS

The three bacterial strains examined were obtained from the Microbiological Laboratory of the University of Ghent. We thank Mr. R. De Baere for skilful microbiological and preparative work. Erna Dams received an I.W.O.N.L. scholarship. Our research was supported in part by F.K.F.O. grant nr. 2.0017.80.

*To whom correspondence should be addressed.

REFERENCES

1. Erdmann, V.A., Huysmans, E., Vandenberghe, A., De Wachter, R. (1983) Nucl. Acids Res. 11, r105-r133.
2. Küntzel, H., Heidrich, M., Piechulla, B. (1981) Nucl. Acids Res. 9, 1451-1461.
3. Hori, H., Osawa, S. (1979) Proc. Natl. Acad. Sci. USA 76, 381-385.
4. Nishikawa, K., Takemura, S. (1974) FEBS Lett. 40, 106-109.
5. Fox, G.E., Woese, C.R. (1975) Nature 256, 505-507.
6. Douthwaite, S., Garrett, R.A. (1981) Biochemistry 20, 7301-7307.
7. Pieler, T., Erdmann, V.A. (1982) Proc. Natl. Acad. Sci. USA 79, 4599-4603.
8. Hori, H., Osawa, S., Iwabuchi, M. (1980) Nucl. Acids Res. 8, 5535-5539.
9. Luehrsen, K.R., Fox, G.E. (1981) Proc. Natl. Acad. Sci. USA 78, 2150-2154.
10. Garrett, R.A., Olesen, S.O. (1982) Biochemistry 21, 4823-4830.
11. De Wachter, R., Chen, M.-W., Vandenberghe, A. (1982) Biochimie 64, 311-329.
12. Böhm, S., Fabian, H., Welfle, H. (1982) Acta Biol. Med. Germ. 41, 1-16.
13. Delihas, N., Andersen, J. (1982) Nucl. Acids Res. 10, 7323-7344.
14. Vigne, R., Jordan, B.R. (1977) J. Mol. Evol. 10, 77-86.
15. Jordan, B.R., Galling, G., Jourdan, R. (1974) J. Mol. Biol. 87, 205-225.
16. Du Buy, B., Weissman, S.M. (1971) J. Biol. Chem. 246, 747-761.
17. Nazar, R.N., Matheson, A.T. (1977) J. Biol. Chem. 252, 4256-4261.
18. Sweeney, W.V. (1981) J. Biol. Chem. 256, 12222-12227.
19. Miller, J.H. (1972) in Experiments in Molecular Genetics, p. 433, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
20. Kurland, C.G. (1966) J. Mol. Biol. 18, 90-108.
21. Fang, B.-L., De Baere, R., Vandenberghe, A., De Wachter, R. (1982) Nucl. Acids Res. 10, 4679-4685.
22. Peattie, D.A. (1979) Proc. Natl. Acad. Sci. USA 76, 1760-1764.
23. Sogin, S.J., Sogin, M.L., Woese, C.R. (1972) J. Mol. Evol. 1, 173-184.
24. Fresco, J.R., Alberts, B.M., Doty, P. (1960) Nature 188, 98-101.

25. Crick, F.H.C. (1966) J. Mol. Biol. 19, 548-555.
26. Ninio, J. (1979) Biochimie 61, 1133-1150.
27. Traub, W., Sussman, J.L. (1982) Nucl. Acids Res. 10, 2701-2708.
28. Kim, S.H., Suddath, F.L., Quigley, G.J., Mc Pherson, A., Sussman, J.L., Wang, A.H.J., Seeman, N.C., Rich, A. (1974) Science 185, 435-440.
29. Topal, M.D., Fresco, J.R. (1976) Nature 263, 285-289.
30. Hancock, J., Wagner, R. (1982) Nucl. Acids Res. 10, 1257-1269.
31. Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N., Woese, C.R. (1980) Science 209, 457-463.