Nucleotide sequences of the 5.8S rRNAs of a mollusc and a porifer, and considerations regarding the secondary structure of 5.8S rRNA and its interaction with 28S rRNA

Dominique Ursi, Antoon Vandenberghe and Rupert De Wachter*

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

Received 27 July 1983; Revised and Accepted 21 October 1983

ABSTRACT

We report the primary structures of the 5.8 S ribosomal RNAs isolated from the sponge ${\it Hymeniacidon\ sanguinea}$ and the snail ${\it Arion\ rufus.}$ We had previously proposed (Ursi et al., Nucl. Acids Res. 10, 3517-3530 (1982)) a secondary structure model on the basis of a comparison of twelve 5.8 S RNA sequences then known, and a matching model for the interaction of 5.8 S RNA with 26 S RNA in yeast. Here we show that the secondary structure model can be extended to the 25 sequences presently available, and that the interaction model can be extended to the binding of 5.8 S RNA to the 5'-terminal domain of 28 S (26 S) RNA in three species.

INTRODUCTION

To our knowledge, the primary structure of 5.8 S rRNA has been investigated in 28 species, among which 8 vertebrates, 6 arthropods, 4 protozoa, 2 slime molds, 4 fungi, 3 higher plants, and one alga. Twenty of these sequences can be found in a recent review 1 . However, some of these sequences have already been corrected $^{2-6}$ and new ones have been added to the list $^{7-13}$. One of the objectives that can be served by the availability of such a sequence collection is the search, on a comparative basis, for a universal secondary structure model. We have therefore sought to diversify the sequence collection by examining two 5.8 S RNAs in hitherto uninvestigated animal phyla, namely those of the snail Arion rufus (Mollusca) and the sponge Hymeniacidon sanguinea (Porifera).

A number of models $^{14-16}$ for 5.8 S RNA secondary structure have been proposed, representing the molecule with the two terminal sequences joined by base-pairing. These models can at best reflect the structure of isolated 5.8 S RNA, since the present evidence $^{17-21}$ indicates that precisely the terminal sequences are engaged in base-pairing to 28 S RNA sequences in the 5.8 S - 28 S RNA complex. Some authors 13 , 19 , 22 , 23 therefore represent their sequences in models that borrow a number of helical areas from the Nazar model 14 but leave terminal sequences free for 28 S RNA binding. We have

recently proposed²⁴ a similar model in which, however, one of the helical regions postulated by Nazar et al.¹⁴ is replaced by a different one, judged more acceptable on the basis of an alignment of 12 sequences. We show here that this model remains applicable to 13 additional sequences determined in the meantime. We have also amended²⁴ the model of Veldman et al.²¹ for 5.8 S - 26 S RNA interaction in Saccharomyces carlsbergensis by assuming that 5 helical areas exist in the 5.8 S RNA moiety of the complex, as compared to 4 helices in the Veldman model. We show here that our interaction model, but also that of Veldman et al., fits to the newly published sequences of the large rRNAs of the mouse and of Physarum polycephalum.

MATERIALS AND METHODS

The collection of the sponge *Hymeniacidon sanguinea* and of the snail *Arion rufus*, as well as the isolation of their ribosomes, have already been described 25,26 . The preparation of 5.8 S RNA was as previously documented except that it started from complete ribosomes, the fractionation into 40 S and 60 S subunits being omitted. As a rule, about 3.6 $0D_{260}$ units of 5.8 S RNA were obtained by preparative gel electrophoresis starting from 225 $0D_{260}$ units of ribosomal RNA.

The major part of each sequence could be derived by Peattie's 27 method. In order to establish the sequences near the 5'-termini and to confirm them in other areas, four other combinations of gel sequencing were used, as summarized in Table 1. The only method that has not been described elsewhere is that making use of a 5.8 S RNA dimer. When 5.8 S RNA is incubated with $[5'-^{32}P]$ pCp in the presence of RNA ligase under the usual conditions 27 , the pCp serves as a donor and the RNA as an acceptor in the ligation reaction. However, the RNA can simultaneously serve as a donor for a second RNA molecule, in which case a labeled product containing two concatenated 5.8 S RNAs results:

 $p(5.8 \text{ S RNA})_{OH} + pCp \longrightarrow p(5.8 \text{ S RNA})pCp$ 2 $p(5.8 \text{ S RNA})_{OH} + pCp \longrightarrow p(5.8 \text{ S RNA})p(5.8 \text{ S RNA})pCp$

The labeled dimer is detected as a slower moving band when the ligation products are purified by gel electrophoresis as illustrated in Fig. 1. Although it contains less label than the monomer, it can be profitably used to detect the sequence near the 5'-terminus. Indeed, on a normal sequencing gel of 3'-labeled RNA, the slowest moving bands, which must reveal the 5'-terminal sequence, are usually obscured by a very heavy band of undegraded material. On a sequencing gel starting from the labeled dimer, the undegraded

Combination	Labeling method (a), reference	Degradation method, reference	Sequence segments identified	
Comp		i ei ei eile	H. sanguinea	A. rufus
1	Ligation of DCp at 3'-end (27)	Chemical (27)	9-156	12 - 157(b)
2	Ligation of BCp at 3'-end (27)	RNAses (28)	2- 59 76-113 122-143	
3	Ligation of DCp at 3'-end of dimer (this paper)	RNAses (28)	1- 54	
4	Dephosphorylation with phosphatase; kinase labeling with pppA at 5'-end (24)	RNAses (28)	2- 54 95-134 138-153	2- 44
5	Ligation of (Ap) ₄ A at 5'-end; kinase labeling with pppA at extended 5'-end (28)	RNAses (28)		1- 70

Table 1. Combinations of gel sequencing methods used.

(a) Abbreviations for labeled compounds : β Cp, [5'-32P] pCp; β ppA, [γ -32P] ATP.

(b) The sequence at positions 118-120 and 132-134 gave rise to band compression but could be resolved on gels⁶ run at 65°C.

material runs much slower, and one can read, as it were, through the 5'-end into the 3'-end of the sequence.

Ribose-methylated residues and pseudouridine residues were detected from the band pattern that they generate on Peattie gels as explained previously²⁴. Length heterogeneity was found only in *Hymeniacidon sanguinea* 5.8 S RNA, which contains two components differing by one residue at the 5'-terminus.

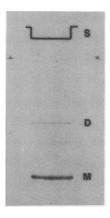


Fig. 1. Ligase catalyzed formation of a labeled 5.8 S RNA dimer.

10 μ g of Hymeniacidon sanguinea 5.8 S RNA was incubated for 16 h at 4°C with 5 nmol pppA, 60 x 10⁶ dpm [5'-³²P] pCp, and 4 units RNA ligase in 5 μ l of a solution containing 0.05 M HEPES (pH 8.3), 0.02 M MgCl₂, 0.003 M dithiothreitol and 20% dimethyl sulfoxide. The mixture was separated by electrophoresis on an 8% polyacrylamide gel of the same composition as used for sequencing. S, start; M, labeled 5.8 S RNA; D, labeled 5.8 S RNA dimer.

The two components, labeled by method 4 (Table 1) were separated on an 8% gel prior to 5'-terminal sequence analysis. The end-groups of each 5.8 S RNA were determined by alkaline hydrolysis of the 5'- and 3'-labeled RNA and identification of the labeled pNp or Np by paper electrophoresis.

RESULTS AND DISCUSSION

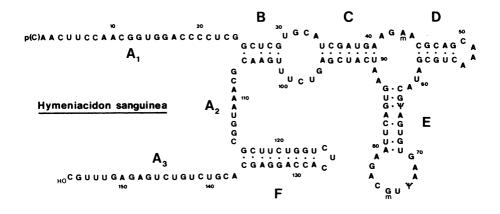
Primary structure and secondary structure model of Arion rufus and Hymeniacidon sanguinea 5.8 S RNAs

The structures of the 5.8 S RNAs from the two examined species are represented in Fig. 2. The *Hymeniacidon sanguinea* 5.8 S RNA shows length heterogeneity at the 5'-terminus, with only a fraction of the molecules possessing the pC $_1$ residue, the others having pA at the 5'-end. It contains 2 methylated riboses (positions 43 and 75) and 2 pseudouridines (positions 64 and 73). The *Arion rufus* 5.8 S RNA is homogeneous, contains a single 2'-0-methylguanosine at position 75 and a single pseudouridine at position 73.

We have previously derived 24 the secondary structure model of Fig. 2 on the basis of an alignment of twelve 5.8 S RNA sequences, and stated in detail why we prefer it to other models. The base-pairing scheme is the same as, or similar to, that assumed by Nazar et al. 14 for other sequences as far as areas C, D, E, and F are concerned, but it is different for area B. Moreover, any potential base-pairing between areas A_1 and A_3 is disregarded since in the ribosome these sequences probably interact with 28 S RNA according to the evidence cited in the introduction.

Applicability of the secondary structure model to the alignment of known 5.8 S RNA sequences

An alignment of 25 presently known 5.8 S RNA sequences is shown in Fig. 3. Among the vertebrates, only 3 species are represented although 8 have been examined. As explained previously 24 , the vertebrate sequences published until 1978 probably all contain the same sequencing error, which was revealed when the $\textit{Xenopus taevis}^2$, 3 rat 4 and mouse 5 sequences were reexamined at the DNA level. We therefore included only these corrected vertebrate sequences. The alignment contains minor adjustments with respect to the previous edition 24 which contained 12 sequences. The main differences reside in the widening of the gap separating the complementary strands of hairpin F, and a few gaps in the 3'-terminal sequences, all of which were necessary to accomodate the extra long Crithidia fasciculata sequence 29 . Beyond position 125 alignment becomes difficult anyway, due to the limited homology between sequences.



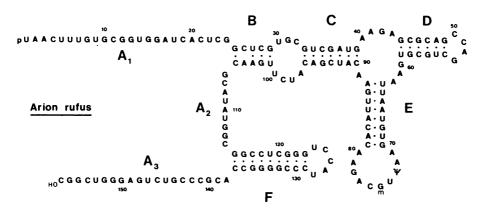
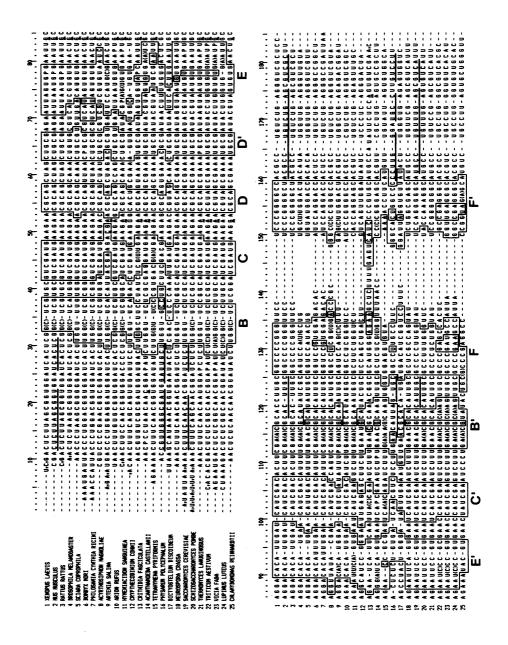


Fig. 2. Secondary structure models for the sequences of 5.8 S rRNA from Hymeniacidon sanguinea and Arion rufus.

Sequence complementarities corresponding to the postulated double-stranded areas B, C, D, E, and F can be found in all hitherto determined 5.8 S RNA sequences, as indicated by the boxes in Fig. 3. Single-stranded areas A_1 , A_2 and A_3 may play a role in interaction with 28 S rRNA, by analogy with the model proposed for this interaction in three other species (Fig. 4).

The boxes delimiting the complementary areas of the secondary structure model can be extended from the 12 sequences for which they were originally postulated ²⁴ to the 25 sequences presently known. This includes complementarity B-B', which assumes a base-pairing scheme different from the Nazar model. The largest shifts in the position of double-stranded areas on the alignment are found in the case of the slime molds *Physarum polycephalum* and *Dictyostelium discoideum*. Among the five postulated double-stranded



regions, hairpin structure E is probably the least convincing on a comparative basis, as already pointed out by others 22 . Not only is the position of the participating bases poorly conserved, but also is the length of the hairpin variable and is it energetically unfavourable in some cases. Obviously, unless one assumes that all hitherto proposed 5.8 S RNA base pairing schemes are completely wrong and that the true one remains to be discovered, one must conclude that the secondary structure of 5.8 S RNA is much less rigorously conserved than is the case with 5 S RNA 1,30 .

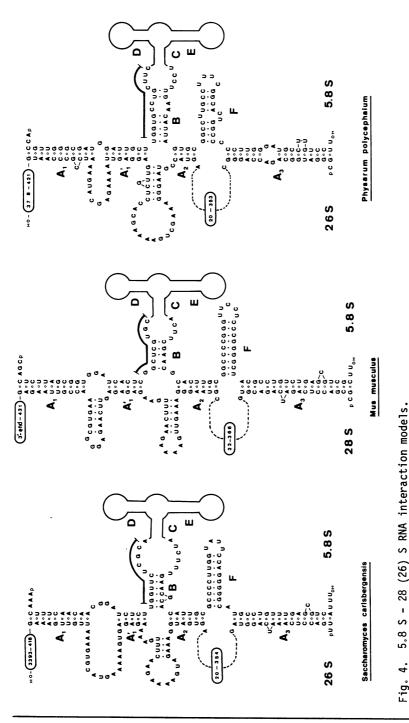
We have previously shown 24 that the secondary structure as defined in Fig. 3 is, in the majority of cases, energetically more favourable than the model proposed by Nazar et al. 14 . An extension of the free energy calculations (results not shown) to the 25 sequences compared here proved that this conclusion remains valid. Only in the case of the vertebrate and higher plant 5.8 S RNAs does Nazar's base pairing scheme give better free energy estimations than the alternative scheme proposed in Fig. 3.

5.8 S RNA - 28 S RNA interaction model

There now seems to grow a consensus that the 5.8 S - 28 S RNA interaction involves base pairing between the 5.8 S RNA 3'-terminal and the 28 S RNA 5'-terminal sequence on the one hand, and on the other hand the 5.8 S RNA 5'-end with a sequence situated some 400 nucleotides from the 28 S RNA 5'-terminus. This assumption rests on experimental probing of the 5.8 S RNA secondary structure as it exists in the complex 17,19,20 and on the fact that the two halves of insect 5.8 S RNAs bind independently to 28 S RNA¹⁸. It is corroborated by an interaction model postulated by Veldman et al. 21 for the 5.8 S - 26 S RNA complex in Saccharomyces carlsbergensis, which is analogous to a secondary structure model for the 5'-terminal domain of Escherichia coli 23 S

Fig. 3. Alignment of 5.8 S rRNA sequences.

The boxes B-B' to F-F' delimit potential base-paired areas, as illustrated in the models of Fig. 2 and Fig. 4. Nested boxes correspond with bulges or small interior loops within these areas. Bases put between brackets correspond with non-standard base-pairs (see references 28, 30, 32 for a discussion of non-standard base pairs in RNA secondary structure). The insect sequences (nos 4 and 5) are interrupted between areas F and F'. Pseudouridine residues are indicated by P, methylated riboses by the subscript m. In some sequences examined on the DNA level the presence of modifications is unknown. In sequences showing length heterogeneity the different termini are indicated by commas. Underlined areas in 5.8 S RNA sequences nos. 2, 16 and 19 are those involved in intermolecular base pairing with 28 S (26 S) RNA sequences according to the models drawn in Fig. 4. The sequence of Saccharomyces carls-bergensis 5.8 S RNA³⁴ is identical to that of Saccharomyces cerevisiae.



Helix B in the 5.8 S RNA moiety The latter models, on the contrary, assume base-pairing A similar complementarity can be observed between Physarum polycephalum and intermolecular helix A2 are specific for our interaction model and do not appear in the models of between the 5.8 S RNA sequences defined by the line segment alongside area B, and residues 329 to 338 Intramolecular base-pairs are denoted by dots, intermolecular pairs by circles. RNA) or 341 to 350 (mouse 28 S RNA). A similar co 5.8 S RNA and residues 331 to 342 in its 26 S RNA.

RNA 31 . The model of Veldman et al. 21 implies that the sequences preceding area C, and those following area F' in Fig. 3 are involved in base pairing with 28 S (26 S) RNA sequences. In other words, it leaves room for the existence of helices C, D, E, and F, but not helix B, in the 5.8 S RNA moiety of the complex, and it implies that a relatively long sequence, comprising residues 107 to 125 in the alignment, remains single stranded. We have published an amended interaction model, illustrated in Fig. 4a, which differs from that of Veldman et al. in two respects. It assumes that helix B does exist in the 5.8 S moiety of the complex, and that the sequence separating helices B and F pairs with a complementary sequence in 26 S RNA, giving rise to area A_2 in Fig. 4. In Veldman's model, helix B is sacrificed to allow pairing between its 5'-terminal strand and a sequence around residue 340 in 28 S RNA.

In the meantime, a sufficient portion of the mouse 28 S $\mbox{RNA}^{\mbox{\scriptsize 5}}$ and the entire Physamum polycephalum 26 S RNA³³ have been sequenced, so it is possible to see which model withstands the test of these new data. In Fig. 4 b and c we show that our interaction model can be transposed to the newly determined sequences. However, Michot et al. 5 have shown that the Veldman model can be transposed to the mouse sequence, and we found that it is also applicable to Physarum polycephalum, at least as far as the intermolecular sequence complementarities are concerned. An argument in favour of the Veldman model is that it bears a closer similarity to the structures postulated for the 5'terminal domains of E. coli and maize chloroplast 23 S RNAs 31 . On the other hand the fact that complementarity B-B' (Fig. 3) exists in all hitherto sequenced 5.8 S RNAs argues for the model shown in Fig. 4. We think that the present evidence is not sufficient to tip the balance in favour of either of the two possibilities, and that the scheme illustrated in Fig. 4 should be given consideration until more definitive conclusions can be drawn from additional comparative sequencing.

ACKNOWLEDGEMENTS

We thank Fang Bao-Ling and Erna Dams for preparations of ribosomes from *Arion rufus* and *Hymeniacidon sanguinea*. Part of this work was carried out while D. Ursi received an I.W.O.N.L. scholarship and part of it was supported 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. Boseley, P.G., Tuyns, A., Birnstiel, M.L. (1978) Nucl. Acids Res. 5, 1121-
- 3. Hall, L.M.C., Maden, B.E.H. (1980) Nucl. Acids Res. 8, 5993-6005.
- 4. Subrahmanyam, C.S., Cassidy, B., Bush, H., Rothblum, L.I. (1982) Nucl. Acids Res. 10, 3667-3680.
- 5. Michot, B., Bachellerie, J.P., Raynal, F. (1982) Nucl. Acids Res. 10, 5273-5283.
- 6. Nazar, R.N., Wildeman, A.G. (1981) Nucl. Acids Res. 9, 5345-5358.
- 7. Jordan, B.R., Latil-Damotte, M., Jourdan, R. (1980) Nucl. Acids Res. 8, 3565-3573.
- 8. Fujiwara, H., Kawata, Y., Ishikawa, H. (1982) Nucl. Acids Res. 10, 2415-2418.
- Feng, Y.-X., Krupp, G., Gross, H.J. (1982) Nucl. Acids Res. 10, 6383-6387.
- 10. Fujiwara, H., Ishikawa, H. (1982) Nucl. Acids Res. 10, 5173-5182.
- 11. Otsuka, T., Nomiyama, H., Sakaki, Y., Tagaki, Y. (1982) Nucl. Acids Res. 10, 2379-2385.
- 12. Wildeman, A.G., Nazar, R.N. (1981) J. Biol. Chem. 256, 5675-5682.
- 13. Rafalski, J.A., Wiewiorowski, M., Söll, D. (1983) FEBS Lett. 152, 241-246.
- 14. Nazar, R.N., Sitz, T.O., Busch, H. (1975) J. Biol. Chem. 250, 8591-8597.
- 15. Ford, P.J., Mathieson, T. (1978) Eur. J. Biochem. 87, 199-214.
- 16. Luoma, G.A., Marshall, A.G. (1978) Proc. Natl. Acad. Sci. USA 75, 4901-4905.
- 17. Pace, N.R., Walker, T.A., Schroeder, E. (1977) Biochemistry 16, 5321-5328. 18. Pavlakis, G.N., Jordan, B.R., Wurst, R.M., Vournakis, J.N. (1979) Nucl. Acids Res. 7, 2213-2238.
- Walker, T.A., Johnson, K.D., Olsen, G.J., Peters, M.A., Pace, N.R. (1982) Biochemistry 21, 2320-2329.
- 20. Peters, M.A., Walker, T.A., Pace, N.R. (1982) Biochemistry 21, 2329-2335.
- 21. Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta, R.J., Branlant, C., Krol, A., Ebel, J.-P. (1981) Nucl. Acids Res. 9, 6935-6952.
- 22. Olsen, G.J., Sogin, M.L. (1982) Biochemistry 21, 2335-2343.
- 23. Schaak, J., Mao, J.-I., Söll, D. (1982) Nucl. Acids Res. 10, 2851-2864.
- 24. Ursi, D., Vandenberghe, A., De Wachter, R. (1982) Nucl. Acids Res. 10, 3517-3530.
- 25. Dams, E., Vandenberghe, A., De Wachter, R. (1982) Nucl. Acids Res. 10, 5297-5302.
- 26. Fang, B.L., De Baere, R., Vandenberghe, A., De Wachter, R. (1982) Nucl. Acids Res. 10, 4679-4685.
- Peattie, D.A. (1979) Proc. Natl. Acad. Sci. USA 76, 1760-1764.
- 28. Dams, E., Vandenberghe, A., De Wachter, R. (1983) Nucl. Acids Res. 11, 1245-1252.
- 29. Schnare, M.N., Gray, M.W. (1982) Nucl. Acids Res. 10, 2085-2092.
- 30. De Wachter, R., Chen, M.-W., Vandenberghe, A. (1982) Biochimie 64, 311-329.
- 31. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.-P., Edwards, K., Kössel, H. (1981) Nucl. Acids Res. 9, 4303-4324.
- 32. Ninio, J. (1979) Biochimie 61, 1133-1150.
 33. Otsuka, T., Nomiyama, H., Yoshida, H., Kukita, T., Kuhara, T., Kuhara, S., Sakaki, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 3163-3167.
- 34. de Jonge, P., Kastelein, R.A., Planta, R.J. (1978) Eur. J. Biochem. 83, 537-546.