NUCLEOTIDE SEQUENCE OF 5 S RNA FROM *TORULOPSIS UTILIS*

Kazuya NISHIKAWA and Shosuke TAKEMURA

Institute of Molecular Biology, Faculty of Science, Nagoya University, Chikusa, Nagoya, Japan

Received 5 December 1973

1. Introduction

In recent years, the complete sequences of 5 S RNAs from *Escherichia coli* [1], human KB cells [2], *Pseudomonas fluorescens* [3], and *Saccharomyces carlsbergensis* [4] have been determined by the use of ^{32}P-labelled RNA and two-dimensional ionophoresis (Sanger's method). The total sequence of *Xenopus laevis* 5 S RNA [5] has also been derived by arranging the sequenced end-products of RNase digestions along the sequence of human 5 S RNA. Several models for the secondary structure of 5 S RNA have been proposed on the basis of many experimental and theoretical works [1,3,6–15]. However, none of these models have expressed general features of 5 S RNA conformation. Additional information on the sequences of 5 S RNAs from other species would be of much use in making a more generalized model.

We have purified non-radioactive 5 S RNA from *Torulopsis utilis* on a large scale, and determined its primary sequence using several column chromatographic procedures. We wish to report briefly the established primary structure, and propose a possible model for the secondary structure of *T. utilis* 5 S RNA which is also applicable to all other 5 S RNAs of known sequence.

2. Materials and methods

From a crude mixture of low-molecular-weight RNA from *T. utilis* supplied by Jūjō Paper Co. Ltd., 5 S RNA was isolated by successive chromatographic procedures on columns of DEAE-Sephadex A-50 and Sephadex G-100. The purified material was completely devoid of amino acid acceptor activity, sedimented as a single component with an $s_{20,w}$ of 4.8 S, and gave a single band by polyacrylamide gel electrophoresis.

The purified 5 S RNA was completely digested with pancreatic RNase and with RNase T_1. The digestion products were separated and sequenced by column chromatographic procedures as described previously [16] with a little modification. For the determination of the sequence U–U–C–U–C–C–G, it was modified with *N*-cyclohexyl-N′-(β-morpholinyl-(4)-ethyl)carbodiimide-methyl-*p*-toluene sulfonate before pancreatic RNase digestion. The nucleosides and nucleotides were identified by their ultraviolet-absorption spectra at pH 2 and 12, and by their chromatographic behavior on Dowex-1 columns. The extent of phosphorylation of the 5′-terminal residue was determined by quantitative phosphorus analysis as described in [17].

To obtain larger oligonucleotides, partial digestions with RNase T_1, pancreatic RNase and RNase U_2 were performed in separate experiments. The products were separated by chromatography on DEAE-cellulose columns in 7 M urea at pH 7.6 and then further at pH 3.5. The isolated fragments were sequenced by analysis of their complete digests with RNase T_1 and pancreatic RNase, as in [18].

3. Results and discussion

By overlapping of the fragments obtained by the complete and partial RNase digestions, the overall nucleotide sequence of *T. utilis* 5 S RNA was established as shown in fig. 1. This RNA is composed of

North-Holland Publishing Company – Amsterdam

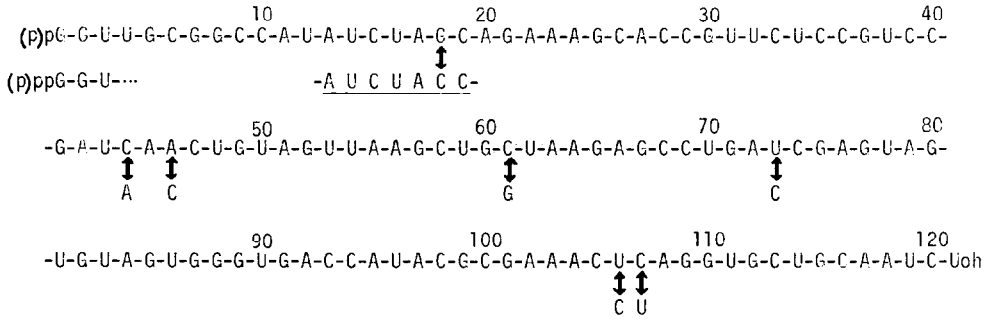

Fig. 1. Nucleotide sequence of *T. utilis* 5 S RNA (upper), showing the positions of nucleotides different from *S. carlsbergensis* (lower). The underlined sequence of the latter has not yet been determined [4].

121 nucleotide residues, and contains no modified nucleotide as do the other 5 S RNAs so far sequenced.

Although *T. utilis* belongs to a different class, *Fungi imperfecti*, from *S. carlsbergensis* which is classified into *Ascomycetes*, there is a close resemblance between these two RNA molecules. As shown in fig. 1, only seven nucleotide substitutions can be found, regardless of the undetermined part of the latter sequence. Recently, Soave et al. [19] have shown the existence of three forms of 5 S RNA, ending at their 5′-terminus with mono-, di-, or triphosphate groups, to be a general case for eukaryotic organisms. So, the apparent difference in the extent of 5′-terminal phosphorylation between the RNAs from *T. utilis* and *S. carlsbergensis* [4] would probably be ascribed to the partial loss of the γ- (and, to a lesser extent, β-)phosphate during the preparation or the subsequent analysis. The 3′-terminal heterogeneity found in human KB cells and *Xenopus* 5 S RNAs [2,5] and the intramolecular sequence heterogeneity found in *E. coli* and *Ps. fluorescens* 5 S RNAs [1, 3] were not observed in *T. utilis* 5 S RNA.

In fig. 2, we propose a possible model for the secondary structure of *T. utilis* 5 S RNA, which is consistent with the results of analysis of the partial RNase digests. Although several secondary structure models so far proposed (for example, that for human KB cells [6]) cannot fit *T. utilis* 5 S RNA, our model is well applicable to other eukaryotic 5 S RNAs of known sequence, leading to a generalized base-pairing pattern shown in fig. 3.

In this pattern, sequences common to eukaryotic 5 S RNAs (shown by letters) are found particularly in the single-stranded region. As an exception, the sequences –A–U–A(44)–A–C(46)–C–U–G– and –A–A–A–C–C(106)–U(107)–A–G– in *S. carlsbergensis* 5 S RNA cannot be arranged in this pattern. However, *S. cerevisiae* 5 S RNA completely fits this pattern, for the latter has –A–U– C(44)–A–A(46)–C–U–G– and –A–A–A–C–U(106)–C(107)–A–G which are the same as *T. utilis* 5 S RNA (M. Miyazaki, unpublished results). This model is also consistent with the results obtained by Vigne et al. [20], who detected in eukaryotic 5 S RNAs two sites (around positions 40 and 90) equally accessible to Rnases T_1

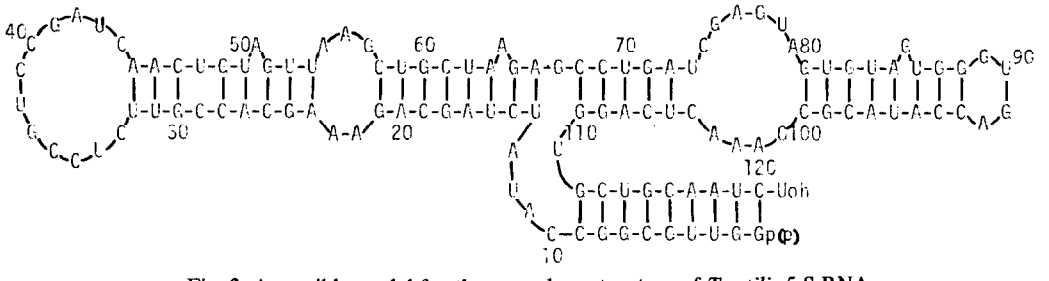

Fig. 2. A possible model for the secondary structure of *T. utilis* 5 S RNA.

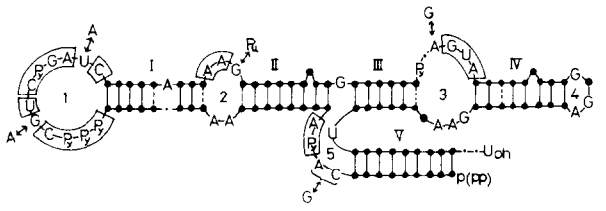

Fig. 3. A generalized base-pairing pattern for 5 S RNAs from eukaryotic organisms. Letters in loops indicate the nucleosides common to human KB cells, *X. laevis* and *T. utilis* 5 S RNAs. The secondary structure model of prokaryotic 5 S RNAs derived from this model (not shown) has somewhat shorter length of stem III and longer of stem V, and some base-substitutions at the positions indicated by arrows. The sequences in boxes are commonly found at the corresponding positions in both of the models for eukaryotic and prokaryotic 5 S RNAs.

and T_2. A slight modification of this model gives a very similar base-pairing pattern (not shown) common to prokaryotic 5 S RNAs. It resembles the model proposed by DuBuy et al. [3]. The main differences between the eukaryotic and prokaryotic models are; somewhat shorter length of stem III and longer of stem V in the latter, and the substitutions of some bases in the single-stranded region (indicated by arrows). The sequences enclosed in boxes are commonly found at the corresponding positions in both of the eukaryotic and prokaryotic models. Quite recently, the nucleotide sequence of *Bacillus stearothermophilus* 5 S RNA has been reported though it is partially undetermined [21]. Since this sequence well fits our model, it is possible to predict the sequence of the undetermined region.

The common sites in the proposed secondary structure might probably be related to a common role of 5 S RNA in ribosome structure. Among these common sites, the most interesting would be loop 1 which is rich in pyrimidine residues. Prokaryotic 5 S RNAs have the sequence G–A–A–C, and eukaryotic 5 S RNAs have G–A–U–C at the corresponding positions in this loop. The sequence G–A–A–C found in several 5 S RNAs has been considered to be a possible recognition site for tRNA binding on ribosomes, for it is complementary to the sequence G–T–ψ–C commonly found in tRNAs [22,23]. Eukaryotic initiator tRNAs contain the sequence G–A–U–C (complementary to the common sequence G–A–U–C in eukaryotic 5 S RNAs) in place of G–T–ψ–C [24]. Hence it is very tempting to suppose that eukaryotic 5 S RNAs may serve as one of the sites for initiator tRNA binding on ribosomes, and that the site for other tRNAs may be 5.8 S RNA or such another. The 5.8 S RNA is present in the large ribosomal subunit of *S. cerevisiae* together with 5 S and 28 S RNAs, and has the sequence G–A–A–C at positions 43–46 from the 5′-terminus [25].

Acknowledgement

We thank Dr. K. Hata of Jūjō Paper Co. Ltd. for the supply of *T. utilis* tRNA mixture, and also Drs. F. Egami and T. Uchida for RNase U_2. This work was supported in part by a grant from the Ministry of Education of Japan.

References

[1] Brownlee, G.G., Sanger, F. and Barrell, B.G. (1967) Nature 215, 735.
[2] Forget, B.G. and Weissman, S.M. (1967) Science 158, 1695.
[3] DuBuy, B. and Weissman, S.M. (1971) J. Biol. Chem. 246, 747.
[4] Hindley, J. and Page, S.M. (1972) FEBS Letters 26, 157.
[5] Brownlee, G.G., Cartwright, E., McShane, T. and Williamson, R. (1972) FEBS Letters 25, 8.
[6] Forget, B.G. and Weissman, S.M. (1969) J. Biol. Chem. 244, 3148.
[7] Boedtker, H. and Kelling, D.G. (1967) Biochem. Biophys. Res. Commun. 29, 758.
[8] Cantor, C.R. (1968) Proc. Natl. Acad. Sci. U.S. 59, 478.
[9] Lee, J.C. and Ingram, V.M. (1969) J. Mol. Biol. 41, 431. g[10]
[10] Lewis, J.B. and Doty, P. (1970) Nature 225, 5110.
[11] Vigne, R. and Jrdan, B.R. (1971) Biochimie 53, 981.
[12] Mirzabekov, A.D. and Griffin, B.E. (1972) J. Mol. Biol. 72, 633.
[13] Raacke, I.D. (1968) Biochem. Biophys. Res. Commun. 31, 528.
[14] Richards, E.G. (1969) Eur. J. Biochem. 10, 36.
[15] Lapidus, I.R. and Rosen, B. (1970) J. Theoret. Biol. 27, 417.
[16] Takemura, S., Mizutani, T. and Miyazaki, M. (1968) J. Biochem. 64, 827.
[17] Miyazaki, M. and Takemura, S. (1969) J. Biochem. 65, 159.

[18] Mizutani, T., Miyazaki, M. and Takemura, S. (1968) J. Biochem. 64, 839.
[19] Soave, C., Nucca, R., Sala, E., Viotti, A. and Galante, E. (1973) Eur. J. Biochem. 32, 392.
[20] Vigne, R., Jordan, B.R. and Monier, R. (1973) J. Mol. Biol. 76, 303.
[21] Marotta, C.A., Levy, C.C., Weissman, S.M. and Varrichio, F. (1973) Biochemistry 13, 2901.
[22] Jordan, B.R. (1971) J. Mol. Biol. 55, 432.
[23] Shimizu, N., Hayashi, H. and Miura, K. (1970) J. Biochem. 67, 373.
[24] Simsek, M., Petrissant, G. and RajBhandary, U.L. (1973) Proc. Natl. Acad. Sci. U.S. 70, 2600.
[25] Rubin, G.M. (1973) J. Biol. Chem. 248, 3860.