Primary and secondary structure of dinoflagellate U5 small nuclear RNA*

Mei-Hua Liu, Ram Reddy, Dale Henning, David Spector and Harris Busch

Department of Pharmacology, Baylor College of Medicine, Houston, TX 77030, USA

Received 2 June 1983; Revised and Accepted 12 December 1983

ABSTRACT

U5 RNA is one of the six capped small nuclear RNAs present in most eukaryotic cells. Like Ul, U2, U4 and U6 RNAs, U5 RNA is associated with hnRNP particles and is thus probably involved in some, as yet undefined, aspects of pre-messenger RNA processing. In this study, the complete nucleotide sequence of U5 RNA of a dinoflagellate, Crypthecodinium cohnii was determined. The analysis of this dinoflagellate U5 RNA sequence showed that a) the sequence homology between human, rat and chicken U5 RNA sequences and dinoflagellate U5 RNA sequence is 64%; b) the extent and the position of post-transcriptional modifications are similar to those found in U5 RNA of higher eukaryotes; c) although the dinoflagellate U5 RNA is shorter in length (108 nucleotides long vs 117 long in human, rat and chicken cells), the RNA fits well into the same secondary structure proposed for U5 RNA of higher eukaryotes (Krol et al. (1981) Nucl. Acids Res. 9 , 769); and $\,$ d) the AUn nucleoti $\,$ de sequence protected by the Sm-antigen and the tight secondary structure found near the 3'-end of other U-RNAs was also found in dinoflagellate U5 RNA. The high order of homology observed between dinoflagellate U5 RNA and U5 RNA of higher eukaryotes indicates that dinoflagellates are more closely related to metazoans than to early eukaryotes.

INTRODUCTION

There are six capped small nuclear RNAs (designated Ul to U6 snRNAs) present in all eukaryotic cells thus far examined $(1,2)$. Of these, five snRNAs Ul, U2, U4, U5 and U6 are localized to nucleoplasm and in part, are associated with hnRNPs (3,4). The U-snRNPs like Ul and U2 snRNPs have been proposed to function in the splicing of premessenger RNAs (5-7). The U5 RNPs associated with hnRNPs were found to be completely protected when these complexes were digested extensively by micrococcal nuclease (8). This result suggests that U5 RNP is localized inside the hnRNP complex and may play an important role in packaging hnRNPs and/or in processing of hnRNAs.

The primary sequences of human (11) , rat $(1,11)$, mouse (12) , and chicken (11) U5 RNA were previously determined and secondary structures were proposed (11,12). In this study, the complete nucleotide sequence of a dinoflagellate U5 RNA was determined and compared with other U5 RNA sequences. The dinoflagellate U5 RNA was found to be highly conserved and fits into a secondary structure model very similar to those proposed for other U5 RNAs. Since U5 RNA in a ribonucleoprotein particle has been implicated in premessenger RNA processing, the understanding of the structure of U5 RNA from evolutionarily distant species may be of use in understanding the mechanism of action of U5 RNP particles in eukaryotic cells.

MATERIALS AND METHODS

Preparation of dinoflagellate RNA. The non-photosynthetic dinoflagellate Crythecodinium cohnii was grown in Gold and Baren's medium (13). For preparation of labeled RNA dinoflagellates were incubated in phosphate-free medium (13) to which $\binom{32}{P}$ -orthophosphate was added at room temperature and without shaking for 24 hours. A typical preparation of 2 grams of dinoflagellates was incubated with 50 mCi $[^{32}P]$ -orthophosphate in 500 ml medium.

Preparation of U5 RNA. The U5 RNA used in this study was obtained from whole cell RNA extracts of the labeled dinoflagellates. When uniformly-labeled U5 RNA was isolated, 4-5 mg of yeast tRNA was added to the cells prior to the SDS-phenol extraction (14) to reduce degradation of RNA. When the U5 RNA was to be end-labeled, the yeast tRNA was not added during the extraction. The 4-8S RNA was precipitated from sucrose density gradients (14), separated on 10% polyacrylamide-7M urea gels, pH 8.3 (15) stained with methylene blue, excised and extracted (16).

Chemicals. RNase A, T1 RNase and U2 RNase were from Calbiochem-Behring; T_2 RNase was from Sankyo. P₁ Nuclease was from Sigma and polynucleotide kinase and RNA ligase were from P-L Biochemicals. $[\gamma^{-32}P]$ ATP and $[5'-^{32}P]$ pCp were from Amersham. Polyethyleneimine-cellulose plates were obtained from Brinkmann and DEAE-cellulose plates were from Analtech.

Fingerprinting of dinoflagellate U5 RNA. Complete digestion of U5 RNA with RNase A or T_1 RNase and two-dimensional separation

Figure 1. Fingerprinting of $32P-$ labeled Dinoflagellate U5 RNA. Uniformly labeled U5 RNA was digested with T_1 RNase (A), or RNase A (B) and fingerprinted. The first di- (A) , or RNase A (B) and fingerprinted. mension was carried out on cellulose acetate at pH 3.5 and the second dimension was carried out by homochromatography on PEI-cellulose plates (17,18). The sequences shown next to each oligonucleotide were derived from the analysis of products after secondary enzymatic digestions, sequencing gels or from wandering spot analysis of end-labeled oligonucleotides.

of oligonucleotides was as described by Brownlee et al. (17) . The second dimension was on polyethyleneimine-cellulose plates using C15 homo-mixture (18).

Sequencing of U5 RNA. 3'-end labeling of U5 RNA with pCp and RNA ligase was done according to the method of England and Uhlenbeck (19); the intact U5 RNA was then subjected to chemical modification and cleavage according to Peattie (20). Sequence gels were 90 cm long containing 10% polyacrylamide-7M urea, pH 8.3 and were usually run at 50°C (21).

Modified Nucleotides. Uniformly labeled U5 RNA was digested completely with T_2 RNase; the products were separated in the first dimension by electrophoresis on Whatman 3MM paper, pH 3.5 and in the second dimension by chromatography using isopropanol:HCl:water (22). The pseudouridine residues were localized to specific positions based on the absence of hydrazine reactivity in pyrimidine

Figure 2. Wandering spot analysis of T_1 RNase oligonucleotides. The 5'-end labeled oligonucleotides T-9 to T-12 were partially digested with formamide (16) and fingerprinted as described in Fig. 1. The 5'-end nucleotides were identified after complete digestion of these fragments with nuclease Pl and separation of the nucleotides on 3 MM paper by electrophoresis. Consecutive nucleotides in each case are identified representing the shift in mobility.

residues (23).

Sequences of Large T_1 Fragments. 1 µg of unlabeled U5 RNA was digested completely with T_1 RNase and the large T_1 fragments precipitated with ethanol (ethanol/sample, $2:1, V/V$). These large fragments were 5'-end labeled by the method of Donis-Keller (21), the labeled fragments were partially digested with formamide (16), and subjected to wandering spot analysis to confirm the sequences deduced from the sequencing gels.

RESULTS

Fig. 1 shows the T_1 RNase and RNase A fingerprints of dinoflagellate U5 RNA. The $[32p]$ -labeled U5 RNA was obtained by immunoprecipitating snRNP particles with anti-Sm antibodies. The oligonucleotide T-8 in the T_1 RNase fingerprint (Fig. 1A) and P-11 in the RNase A fingerprint (Fig. 1B) contained the 'cap' structure 6 m3GpppWp. The characterization of cap structures of dinoflagellate snRNAs was reported earlier (24). Oligonucleotide T-13 had the composition 4Cp,lAp and did not contain any Gp. Therefore, T-13 is the 3'-end oligonucleotide of U5 RNA. All the oligonucleotides were digested secondarily with U2 RNase, T_1 -RNase or RNase A and the products were analyzed on DEAE-cellulose paper. The sequences of some of the oligonucleotides were deduced from these products. For determination of sequences of larger T_1 RNase oligonucleotides, the 5'-end labeled oligonucleotides were partially digested with formamide (16) and subjected to wandering spot analysis. The results obtained with T-9, T-10, T-ll and T-12 are shown in Fig. 2. In addition, the 3'-end labeled U5 RNA was partially digested and subjected to wandering spot analysis. The sequence -G-C-C-C-C-A-C-p*-Cp at the 3'-end was confirmed by this experiment (data not shown).

The analyses of 5'-end oligonucleotide T-8 and P-12 after various enzyme digestions are shown in Fig. 3. Oligonucleotide T-8 was digested with RNase A (Lane 1), T_2 RNase (Lane 2) and U2 RNase (Lane 3) and subjected to electrophoresis on DEAE-cellulose paper at pH 3.5 (17) . The products obtained are consistent with the sequence m3Gppp $_{H-K}^{H}$ - $_{H-K}^{H}$ -A-C-A-Gp. The oligonucleotide P-12 was subjected to partial Pl-nuclease digestion and the partial products obtained (Lane 4) were analyzed on 3 MM paper by electrophoresis after complete Pl-nuclease digestion (Lanes 5 and 6). The sugar-methylated nucleotides were identified by descending chromatography (Fig. 3, part 7). These data clearly established the 5'-end sequence of dinoflagellate U5 RNA to be $mgGPPP_{A-H-}Q-A-$ C-A-Gp. The sequences of the oligonucleotides obtained from these and other data are shown in Fig. 1. Fig. 4 shows the autoradiogram of the sequencing gel obtained when 3'-end labeled U5 RNA was digested by the chemical method of Peattie (20) and fractionated on a 10% polyacrylamide gel. The sequence corresponding to nucleotodes ¹ to 90 is shown in panel 4A and the sequence up to nucleotide 108 is shown in panel 4B.

Four nucleotides at positions 66, 68, 71 and 85, indicated by asterisks in Fig. 4, panel A, showed considerable reactivity in the U channel. However, in other experiments these four residues showed considerably less reactivity in the U channel compared

Nucleic Acids Research

Figure 3. Analysis of 5'-end oligonucleotide (T-8) after various enzymatic digestions. Oligonucleotide T-8 obtained as described in Fig. 1A was digested with RNase A (Lane 1), T2 RNase (Lane 2) or U2-RNase (Lane 3). Oligonucleotide P-12 obtained as described in Fig. 1B was partially digested with Pl-nuclease (Lane 4). The digests were fractionated on DEAE-cellulose paper at pH 3.5. The partial digestion products obtained from oligonucleotide P-12 from Lane 4, were digested to completion with Pl-nuclease and the digests analyzed by electrophoresis on 3 MM paper at pH 3.5 (Lanes 5 and 6). The nucleotides designated pA, pCm and pUm obtained from Lanes 5 and 6 were subjected to descending chromatography on the Wyatt system (22). The authentic pCm and pUm marker nucleotides obtained from labeled U2 and U6 RNAs by the method of Silberklang et al (35) had mobilities identical to those of pCm and pUm obtained from the 5'-end oligonucleotide of dinoflagellate U5 RNA.

to adjacent authentic U cleavages (Fig. 4, panel C). These residues are designated as C residues based on wandering spot analysis shown in Fig. 2 and some, as yet unknown, modification of these residues has not been ruled out.

The 3'-ends of human, rat and chicken U5 RNAs were found to

Figure 4. Autoradiographs of sequencing gels of 3'-end labeled U5 RNA. The 3'-end labeled U5 RNA was digested by the method of Peattie (20) and fractionated on 10% polyacrylamide gels. The polyacrylamide gel was maintained at 250C during electrophoresis. Electrophoresis in Panel A was carried out for a longer duration compared to that in Panel B. Oligonucleotides 66, 68, 71 and 85, indicated by asterisks in' Panel A, showed greatly reduced U cleavages in other independent experiments (Panel C).

contain a tight secondary structure near the 31-end that interfered with resolution on sequencing gels (11). Similar secondary structural features were found in dinoflagellate U5 RNA. The se-

Figure 5. Comparison of sequencing gels run at 25 $^{\circ}$ C and 50 $^{\circ}$ C. 3'-end labeled U5 RNA was digested by the method of Peattie (20) and fractionated on 10% acrylamide gels at 25°C (A) or at 50°C (B). The electrophoretic runs were made in different experiments.

quence corresponding to nucleotides 93-97 was not resolved when the sequencing gel was run at 25° C (Fig. 5A). This region was resolved when the gel was run at 50° C (Fig. 5B) which indicates there is a tight secondary structure in this region.

Modified nucleotides of U5 RNA. Dinoflagellate U5 RNA was

Figure 6. Fractionation of a complete T_o RNase digest for the
identification of modified nucleotides. Electrophoresis was carried out in the first dimension on 3 MM paper, pH 3.5, and the second dimension was descending chromatography in isopropyl alcohol/HCl/H₂0 (680:176: 144, v/v) (6A). Autoradiogram of sequencing gel for the identification of pseudouridylic acid residues (Fig. 6B). Pseudouridylic acid residues were shown not to react with hydrazine in the chemical sequencing method of Peattie (20) and pseudouridylic acid residues found in U5 RNA were localized to positions 11, 30, 33, 42 and 45.

found to contain several post-transcriptional modifications (24). These modifications were further characterized in this study. A complete T_2 RNase digest of U5 RNA was fractionated on 3 MM paper by electrophoresis at pH 3.5 and by descending chromatography in the second dimension (Fig. 6A). In addition to the four major mononucleotides, U5 RNA contained pseudouridylic acid and several alkali-stable di and trinucleotides consisting of GmCp, CmVp, UmCp, UmUp and UmCmAp. The 'cap' containing oligonucleotide was not visualized in this system, possibly because the m_3^2 , 2 , 7 G was degraded due to high pH of triethylamine used in the preparation of RNA and the 2.1 M HC1 used for chromatography. The instability of m_qG containing cap to alterations in pH was observed previously (25). The pseudouridylic acid residues were found in oligonucleo-

Figure 7. Nucleotide sequence of U5 RNA. The T₁ RNase digestion fragments are shown below and RNase A digestion fragments are shown above the sequence. The numbers correspond to the numbered oligonucleotide shown in Fig. 1.

tides T-10 and T-12. They were further localized to positions 11, 30, 33, 42 and 45 (Fig. 6B).

The complete nucleotide sequence of dinoflagellate U5 RNA is shown in Fig. 7. This RNA is 108 nucleotides long and the large RNase A and T_1 RNase oligonucleotides are shown above and below the sequence, respectively. When this sequence was compared with chicken U5 RNA sequence, two regions of homology were found. The dinoflagellate U5 RNA sequence 1-65 was 77% homologous to chicken U5 RNA sequence and the region, 85-95 was over 90% homologous to the chicken U5 RNA sequence. The remainder of the sequence was less than 30% homologous.

DISCUSSION

Dinoflagellates are very primitive and perhaps the oldest eukaryotes; they may represent a link between prokaryotes and more advanced eukaryotes (9,10,26). Our earlier studies (24) showed that dinoflagellates contain six capped snRNAs corresponding to Ul-U6 RNAs found in higher eukaryotes (1) and the anti-Sm antibodies were shown to immunoprecipitate Ul, U2, U4, U5 and U6 RNA containing RNP particles (24). In this study the complete nucleotide sequence of U5 RNA was determined as part of a series of studies on conservation of primary and secondary structural features of U-snRNAs during evolution (11). Of the U-snRNAs, U5 RNA was the most uridylic acid rich and contains 36% uridylic acid (1). The dinoflagellate U5 RNA contained only 30% uridylic acid. The dinoflagellate U5 RNA was 108 nucleotides long, compared to 117 to 118 nucleotides in higher eukaryotes. Similar size variations in evolutionarily distant species are known for 5S, 5.8S, 18S and 28S ribosomal RNAs. However, it is noteworthy that this size variation has not affected markedly the secondary structure (see Fig. 9).

Figure 8. Homology between U5 RNA sequences of chicken and dinoflagellates. The dinoflagellate U5 RNA sequence is aligned with the chicken U5 RNA sequence (5). The alignment was done so as to maximize the homology between the two RNA sequences.

Secondary structure of dinoflagellate U5 RNA. A possible secondary structure was proposed for human, rat, mouse and chicken U5 RNAs (11,12). The sequences of U5 RNA from human to avain species are over 98% homologous and the proposed secondary structure for chicken U5 RNA is shown in Fig. 9. The dinoflagellate U5 RNA fits well into a similar secondary structure (Fig. 9). Stems ^I and II as well as loops A, B, C, D and E found in other U5 RNA secondary structures were also found in the dinoflagellate U5 RNA structure. The loops A, B and C were identical in size between dinoflagellate U5 RNA and chicken U5 RNA. The E loop is considerably smaller (3 nucleotides) in dinoflagellate U5 RNA compared to six nucleotide E loop in chicken U5 RNA (Fig. 9). A loop structure consisting of only three nucleotides is permissible according to Tinoco et al. (27).

The Sm-antigen was shown to protect the sequences in U5 RNP, comprising loops D, E and the stem II (4). The sequence AAUUUUU was conserved between the dinoflagellate and chicken U5 RNAs in this region protected by Sm-protein. These results indicate that this AAUUUUU sequence may be important for the binding of Sm-antigen. The sequence corresponding to stem II and loop E was not conserved, but a very tight secondary structure was found in this portion of the U5 RNA of dinoflagellates and chicken (Figs. 8 and 9). There are several compensating nucleotide substitutions or deletions resulting in a conserved secondary structure (Fig. 9).

Several other notable similarities were observed between dinoflagellate U5 RNA and U5 RNAs of higher eukaryotes. These inlude: a) the cap structure containing trimethylaguanosine; b) sugar methylated nucleotides at positions 2, 34, 36, 40 and 44; c) pseudouridylic acid residues at positions 30, 33 and 45. In

Figure 9. Comparison of possible secondary structures of U5 ANA of chicken and dinoflagellates. A secondary structure for chicken U5 RNA proposed by Krol et al. (11) shown on the right. The dinoflagellate U5 RNA sequence fits well into a similar secondary structure (left).

addition, the overall homology between dinoflagellate U5 RNA and U5 RNA of higher eukaryotes was 64%. The sequence homology between U3-snRNA of rat (28) and Dictyostelium (29) was found to be 40% . Similar results were obtained by Hinnebusch et al. (30) who sequenced dinoflagellate 5S and 5.8S RNAs and evaluated the phylogenetic position of dinoflagellates. The dinoflagellate 5S RNA showed 75% homology to the 5S RNA sequences of higher animals and

Figure 10. Homology between ASB-viroid sequence and U5 RNA sequences of dinoflagellates and rat. The U5 RNA sequences of rat and dinoflagellates were compared with the ASBviroid sequence and the portions of sequences showing homologies shown. Asterisks indicate nucleotide iden-
tity. Periods in between the sequences indicate con-Periods in between the sequences indicate continuity in sequence and solid lines indicate breaks in the sequences being compared.

less than 60% homology to prokaryotic sequences. The results obtained in this study using U5 snRNA are similar to that of Hinnebusch et al. (30) and support the suggestion that dinoflagellates are more closely related to higher eukaryotes than to lower eukaryotes. In addition, the similarities observed between dinoflagellate and higher eukaryotic U5 snRNA primary and secondary structure suggest that this snRNA has remained relatively constant, implying that U5 snRNA has an important, highly conserved function.

Several studies suggested that viroids may have originated from cellular RNAs (31,32). Based on the homology between rat U5 RNA and ASB-viroid, Kiss and Solymosy (33,34) suggested small nuclear RNAs as possible candidates from which some viroid sequences have arisen. When the dinoflagellate U5 RNA sequence was analyzed for homology to ASB-viroid sequence, the sequence of dinoflagellate U5 RNA from nucleotide 26 to 101 was found to be 60% homologous to a portion of the viroid sequence (Fig. 10). It is also interesting that the extent of homology between the U5 RNA of dinoflagellates and the rat to ASB-viroid was about the same (Fig. 10). This conservation of homology would be consistent with the hypothesis that some viroids may have their origin in cellular sequences of small nuclear RNAs (31-34).

*This work was supported by the Cancer Center Grant CA-10893, P-3, awarded by the National Cancer Institute, DHEW

REFERENCES

- 1. Reddy, R. and Busch, H. (1981) Cell Nucleus, Academic Press, New York, Ed. H. Busch, 8, 261-306.
- 2. Busch, H., Reddy, R., Rothblum, L. and Choi, Y. C. (1982) Ann. Rev. Biochem. 51, 617-654.
- 3. Sekeris, C. E. and Niessing, J. (1975) Biochem. Biophys. Res. Commun. 62, 642-650.
- 4. Liautard, J. P., Sri-Widada, J., Brunel, C. and Jeanteur, P. (1982) J. Mol. Biol. 162, 623-643.
- 5. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980) Nature (Lond.) 283, 220-224.
- 6. Rogers, J. and Wall, R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1877-1879.
- 7. Ohshima, Y., Itoh, M., Okada, N. and Miyata, T. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4471-4474.
- 8. Sri-Widada, J., Liautard, J. P. Brunel, C., and Jeanteur, P. (1983) Nucl. Acids Res. 11, 6631-6646.
- 9. Schoff, J. W. (1968) J. Paleontol. 42, 651-688.
- 10. Bujak, J. P. and Williams, G. L. (1981) Can. J. Bot. 59,

