The EMBO Journal Vol.1 No.10 pp.1259–1265, 1982

U2 RNA shares a structural domain with U1, U4, and U5 RNAs

Christiane Branlant, Alain Krol, Jean-Pierre Ebel, Eliane Lazar[1], Bernard Haendler[1], and Monique Jacob[1]*

Institut de Biologie Moléculaire et Cellulaire du CNRS, 15, rue René Descartes, 67000 Strasbourg, and [1]Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, 11, rue Humann, 67085 Strasbourg Cédex, France

Communicated by J.-P. Ebel
Received on 26 August 1982

We previously reported common structural features within the 3′-terminal regions of U1, U4, and U5 RNAs. To check whether these features also exist in U2 RNA, the primary and secondary structures of the 3′-terminal regions of chicken, pheasant, and rat U2 RNAs were examined. Whereas no difference was observed between pheasant and chicken, the chicken and rat sequences were only 82.5% homologous. Such divergence allowed us to propose a unique model of secondary structure based on maximum base-pairing and secondary structure conservation. The same model was obtained from the results of limited digestion of U2 RNA with various nucleases. Comparison of this structure with those of U1, U4, and U5 RNAs shows that the four RNAs share a common structure designated as domain A, and consisting of a free single-stranded region with the sequence Pu-A-$(U)_n$-G-Pup flanked by two hairpins. The hairpin on the 3′ side is very stable and has the sequence Py-N-Py-Gp in the loop. The presence of this common domain is discussed in connection with relationships among U RNAs and common protein binding sites.

Key words: evolution/nucleotide sequence/small nuclear RNA/structural domain/U2 RNA

Introduction

The cell nucleus contains a set of metabolically stable small RNAs in the 4–10S range which were designated as U1 to U6 RNAs. U1, U2, U4, U5, and U6 RNAs were found in heterogeneous nuclear ribonucleoproteins (hnRNP) (Deimel *et al.*, 1977; Guimont-Ducamp *et al.*, 1977; Northemann *et al.*, 1977; Gallinaro and Jacob, 1979), in the nuclear matrix (Miller *et al.*, 1978; Zieve and Penman, 1976), and in small nuclear ribonucleoprotein (snRNP) particles (Raj *et al.*, 1975; Gallinaro and Jacob, 1979; Lerner and Steitz, 1979; Sri-Widada *et al.*, 1981), U3 being specifically located in the nucleolus (Prestayko *et al.*, 1971). The function of the U RNAs is largely unknown. The observation that snRNPs containing the nucleoplasmic RNAs U1, U2, U4, U5, and U6 are all recognized by an antibody from patients with autoimmune diseases, designated as anti-Sm (Lerner and Steitz, 1979), suggests that the five RNA species are all associated with at least one common protein. If U1, U2, U4, U5, and U6 RNAs share a binding site for the same protein, the existence of a common structural feature in all these RNAs is likely. We have shown that common structures are present in U1, U4, and U5 RNAs (Krol *et al.*, 1981a) and we have now tried to extend this observation to U2 RNA.

*To whom reprint requests should be sent.

 0261-4189/82/0110-1259$2.00/0.

The primary structure of U2 RNA from Novikoff hepatoma has already been published and different models of secondary structure were established by computer analysis (Reddy and Busch, 1981; Reddy *et al.*, 1981). However, the validity of the models obtained in this way is disputable since the present knowledge of the thermodynamic criteria governing RNA secondary structure are not sufficient to allow accurate determination of structure from nucleotide sequence data (Branlant *et al.*, 1981a). Additional information is required and we used two approaches to determine the secondary structure of U2 RNA: (1) an experimental study, namely the determination of the sensitivity of the RNA to nucleases under non-denaturing conditions as previously described (Branlant *et al.*, 1981a) to identify the single-stranded and base-paired sequences; (2) a comparative study of U2 RNA from different animal species which may help in the choice of possible models, since the secondary structure of a RNA molecule is generally highly conserved during evolution as shown for rRNA (Branlant *et al.*, 1981b; Stiegler *et al.*, 1981), and for U1, U4, and U5 RNAs (Krol *et al.*, 1981a).

The experiments allowed us to propose a unique model of the secondary structure for the 141 nucleotides at the 3′ end of the U2 RNA molecule, and comparison with the previously determined structures of U1, U4, and U5 RNAs showed that the four RNAs share a common domain.

Results

Isolation of U2 RNA

The 4–12S RNA isolated from purified nuclei of brain or liver from chicken, pheasant, and rat was fractionated by polyacrylamide gel electrophoresis. A band with the mobility expected for U2 RNA was found in all cell types except in chicken liver where we detected two new bands in the 4.5S region. As will be shown below, one of them corresponds to the 3′ half of the U2 molecule, the other probably being the 5′ half. When RNA was extracted from total tissue without prior isolation of the nuclei, an intact U2 RNA molecule was found, suggesting a specific cleavage of the U2 molecule during the preparation of nuclei.

Sequence analysis of rat, chicken, and pheasant U2 RNAs

The U2 RNA molecules were 3′ end-labeled using [5′-^{32}P]pCp and T4 RNA ligase and analyzed using both the chemical and the enzymatic method for RNA sequencing. The results were verified by sequence analysis of two types of fragments: (a) U2 RNA was partially digested with T1 RNase in the presence of [γ-^{32}P]ATP and T4 polynucleotide kinase, so that the products of digestion were 5′ end-labeled immediately after their release. These products were analyzed by the enzymatic method for RNA squencing; (b) U2 RNA was hybridized to oligo$(dGp)_6$, the hybrid was digested with RNase H and the released RNA fragments were 3′ end-labeled and analyzed as described above for total U2 RNA.

U2 RNA from rat brain and liver. We have determined the nucleotide sequence of the 150 nucleotides at the 3′ end of the U2 RNA molecule (Figure 1). No difference was observed between the RNA from brain and liver suggesting that they are transcribed from the same gene. The sequence is also

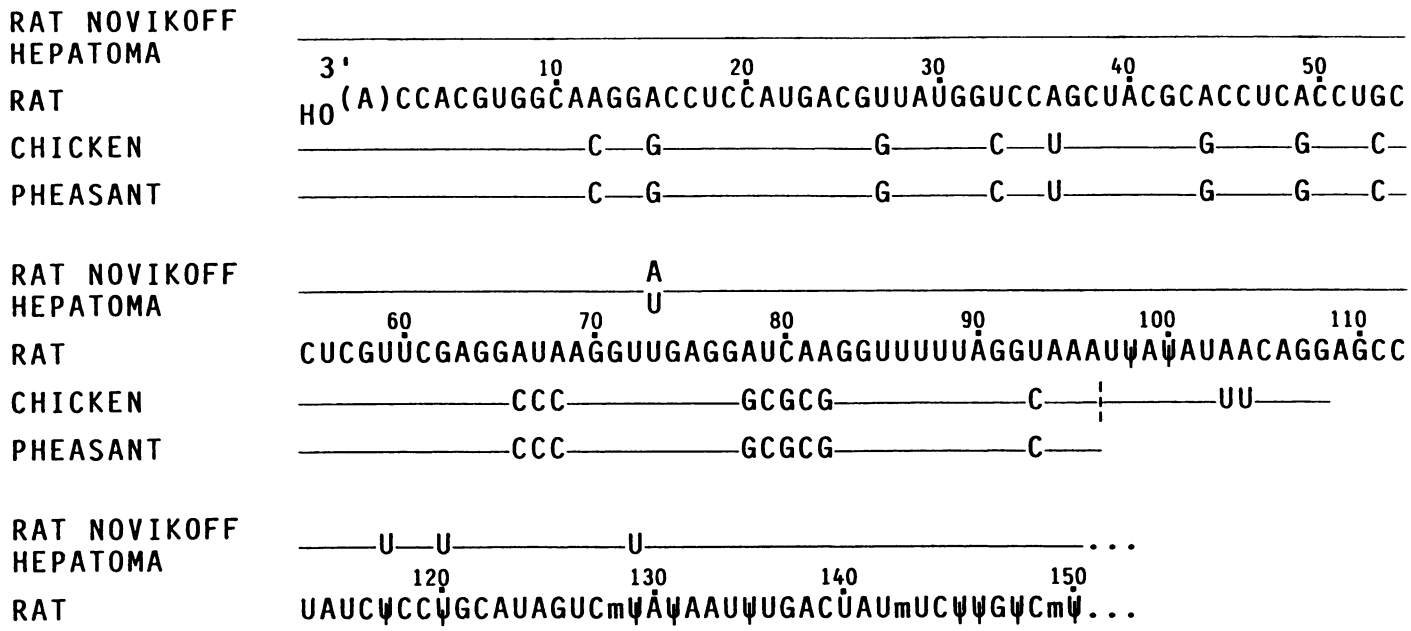

Fig. 1. The nucleotide sequences determined for rat and chicken U2 RNAs. The written sequence is that determined for rat brain and liver U2 RNA. The differences observed in the sequence of Novikoff hepatoma U2 RNA established by Reddy *et al.* (1981) and those in chicken and pheasant U2 RNA are indicated. The vertical dashed line in the chicken sequence indicates the 5′ end of cleaved U2 RNA.

identical to that described by Reddy *et al.* (1981) for U2 RNA from Novikoff hepatoma cells with the exception of three pseudouridines instead of uridines at positions 117, 120, and 129. These nucleotides were not cleaved by any of the reagents used in the chemical sequencing method. In the enzymatic method, they were cleaved by pancreatic ribonuclease but not by *Neurospora crassa* nuclease. This behaviour is characteristic of pseudouridine (Krol *et al.*, 1981a) and was also found with the pseudouridines at positions 98, 100, 131, 135, 145, 146, 148, and 150. A difference in the level of post-transcriptional modification of uridines between normal and hepatoma cells is possible, but a careful re-examination of the sequence of U2 RNA from the latter cells is required for its unambiguous demonstration. The sequence heterogeneity described by Reddy *et al.* (1981) at position 73 was not observed in our case.

U2 RNA from chicken brain and liver. As mentioned above, intact U2 RNA could be obtained from chicken brain nuclei but not from liver nuclei where two new bands were observed at 4.5S. We first determined the sequence of the 108 nucleotides at the 3′ end of brain U2 RNA. In this part of the molecule there are 19 substitutions as compared with rat U2 RNA, indicating a conservation of 82.5% (Figure 1). The base substitutions are spread along the sequenced region and they lead to an increased G-C content in the chicken RNA. Interestingly, such a G-C increase was also observed between chicken and rat U1 RNA (Branlant *et al.*, 1980).

Only one of the two new bands from liver nuclei could be 3′ end-labeled. We showed that it is 96 nucleotides long and has the exact sequence of the 3′ part of brain U2 RNA. The same sequence was also found in the 3′ part of intact U2 RNA extracted from total liver. It is likely that liver U2 RNA was cleaved between nucleotides A96 and U97, generating two fragments of 93 and 96 nucleotides. The 93-nucleotide fragment, corresponding to the 5′ end, probably bears a 3′ phosphate that does not allow 3′ end-labeling. The identity of sequences of the 3′ ends parts of liver and brain RNA suggests that they are transcribed from the same gene. The recent finding of an autosplicing of the 26S pre-rRNA from *Tetrahymena* which involves the specific cleavage of two U-Ap bands (Zaug and Cech, 1982), emphasizes the need to understand the mechanism of cleavage of these bonds.

U2 RNA from pheasant liver. To determine whether the cleavage of U2 RNA in chicken liver was also a feature of avian liver, we studied pheasant liver nuclei. We found an intact U2 RNA which indicates that the cleavage of U2 RNA is not common to all birds. The sequence of the 96 nucleotides at the 3′ end of pheasant U2 RNA was determined and found to be identical to that of chicken U2 RNA, suggesting a high conservation of this RNA in birds (Figure 1).

The secondary structure of U2 RNA

The U2 RNAs from rat liver and chicken brain, as well as the cleaved U2 RNA (3′ half) from chicken liver, were 3′ end-labeled. A fraction of each labeled RNA was digested by chemical reagents as for RNA sequencing. The other fractions were digested with enzymes under mild conditions preserving the secondary structure of the RNA such that T1 and S1 nucleases cleaved only single-stranded regions, and *Naja oxiana* nuclease cleaved only base-paired regions. All the digests were then fractionated simultaneously by polyacrylamide gel electrophoresis. Figure 2 illustrates the fractionation obtained with chicken U2 RNA. Similar results were obtained for rat U2 RNA, indicating a similar distribution of single- and double-stranded sequences in the two RNAs. The results are summarized in Figure 3.

The comparison of intact and cleaved chicken U2 RNA is of particular interest. The preferential sites of enzymatic digestion are the same up to nucleotide 80 but differ in the fragment 80–96. First, the segment 93–96 is hydrolyzed by S1 nuclease in cleaved U2 RNA but not in intact U2 RNA, where it is hydrolyzed by *N. oxiana* nuclease. Second, the segment 80–92 is more strongly hydrolyzed by S1 nuclease in cleaved than in intact U2 RNA. These observations indicate

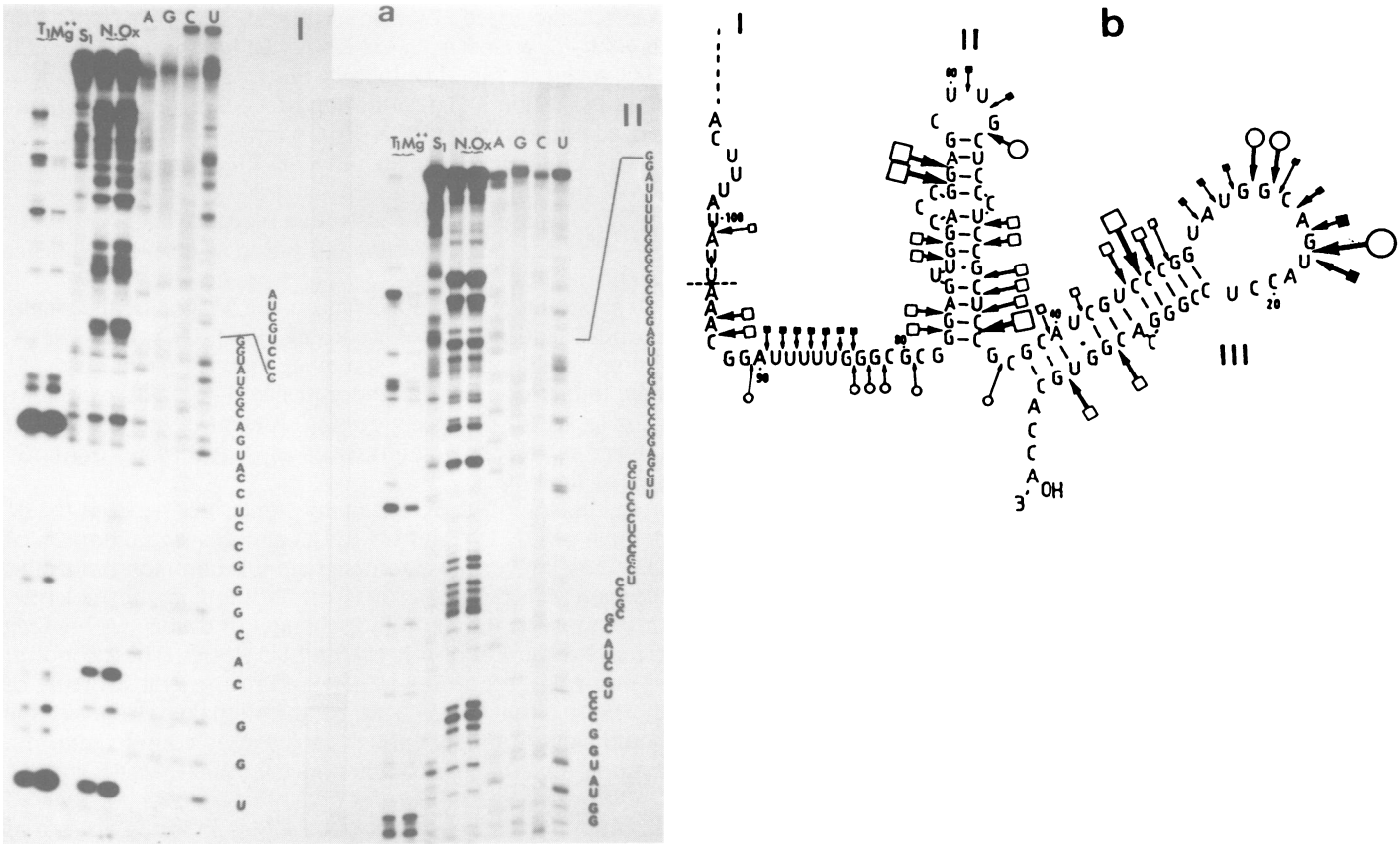

Fig. 2. **(a)** Experimental study of chicken U2 RNA secondary structure. The 3′ end-labeled RNA was partially digested with T1 (T1 Mg^{2+}), S1 and *N. oxiana* (N.ox) nucleases in non-denaturing conditions as described in Materials and methods. It was also digested with chemical reagents as for RNA sequencing (lanes A, G, C, U). The resulting fragments were fractionated by polyacrylamide gel electrophoresis on 20% and 15% gels (**panels I** and **II**). **(b)** Schematic representation of the results. The arrows indicate the positions of preferential T1 ○→, S1 ■→, and *N. oxiana* □→ nuclease cleavages. The dashed line is at the position of *in vivo* cleavage of U2 RNA in chicken liver nuclei.

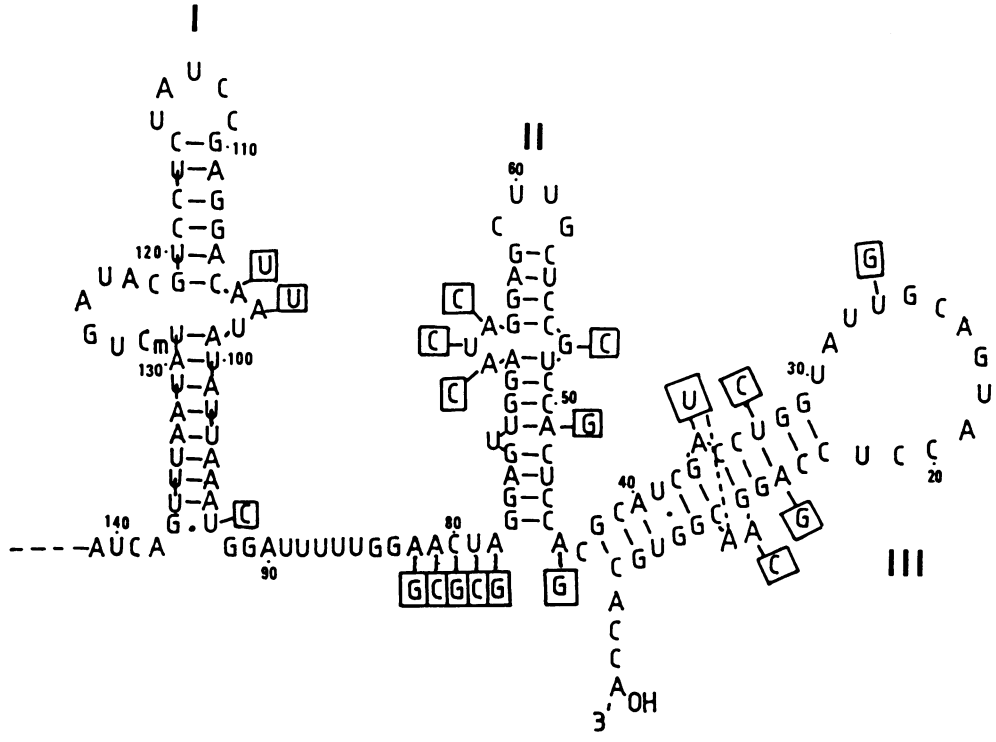

Fig. 3. A model of secondary structure for the 141 nucleotides at the 3′ end of rat U2 RNA. Base substitutions in chicken RNA are indicated in squares.

that the segment 93–96 is single-stranded in cleaved U2 RNA and base-paired to another segment located at the 5′ half of the molecule in intact U2 RNA. Furthermore, the segment 80–92, albeit single-stranded in both molecules, is less accessible to enzymes in intact than in cleaved U2 RNA. The presence of the additional hairpin, or of tertiary interactions, in intact RNA may explain these differences. We may assume that the cleavage of U2 RNA during the isolation of chicken liver nuclei requires a specific mechanism and that nucleotides 96–97 are not protected by proteins.

At this point of the study, models of secondary structure could be built in two ways: (1) on the basis of the results of enzymatic digestion; and (2) on the basis of maximum base-pairing, assuming that the secondary structure is conserved from chicken to rat. Both approaches led to a unique structure, which is a strong argument for its validity. As shown in Figure 3, for the 141 nucleotides at the 3′ end of U2 RNA, the model includes three hairpins (I to III) with a long free single-stranded region between hairpins I and II. In spite of several mutations, hairpins II and III are identical in chicken and rat U2 RNA. Most of these mutations are in single-stranded regions and those in double-stranded regions are compensatory (for instance, a U-A pair is replaced by a G-C pair).

Hairpins I and III of our model are identical to those proposed by Reddy *et al.* (1981) but the structure of the intermediary region is different. However, as already mentioned, the model of these authors was derived by computer. It is not compatible with our experimental data and, furthermore, the intermediary hairpin that they propose is much less stable than hairpin II in our model.

Ohshima *et al.* (1981a) proposed an 'exon model' of splicing involving U2 RNA. The sequence assumed to hybridize with exon extremities is primarily that of hairpin II. The model implies that the base-pairs of this very stable hairpin would be displaced under physiological conditions, which seems unlikely.

DNA sequences complementary to U2 RNA were isolated from human cells and the primary structure of some of them was determined. If they correspond to authentic genes, it should be possible to construct a model of secondary structure identical to that described for rat and chicken U2 RNA. Among the 141 nucleotides of the 3′ end, there were 15 base substitutions and an 11-nucleotide deletion in the U2.7 DNA and 13 substitutions in the U2/4 DNA described by Van Arsdell *et al.* (1981) and by Westin *et al.* (1981). These modifications provoked a destabilization or even a disappearance of certain hairpins. None of the two human DNAs would be transcribed into an RNA fulfilling the criteria established by the study of the secondary structure of U2 RNA from chicken and rat. It is unlikely that U2 RNA from human cells would have such a different structure and the results strongly suggest that U2.7 and U2/4 DNAs are pseudogenes.

Comparison of the structures of U1, U4, U5, and U2 RNAs

A domain of U2 RNA is similar to the 3′ domain of U1, U4, and U5 RNAs. U2 RNA contains in its central part a domain analogous to the 3′ domain of U1, U4, and U5 RNAs. This domain was previously defined as a free single-stranded sequence flanked by two hairpins and containing characteristic nucleotide sequences (Krol *et al.*, 1981a).

The free single-stranded region of U1, U4, and U5 RNAs was characterized by the sequence Py-A-A-$(U)_n$-G-Pup (n = 3–6, region i). In U2 RNA, the homologous region is G-G-A-$(U)_5$-G-Gp. The hairpin on the 3′ side of the single-stranded region had a short loop of 4–6 nucleotides characterized by sequence j, Py-N-Py-Gp in U1, U4, and U5 RNAs. It was terminated by the triplet C-U-Pu_{OH} (region k). In U2 RNA, the hairpin loop is also short and its sequence is C-U-U-Gp analagous to j. It ends with a sequence C-C-Ap (rat) or C-C-Gp (chicken) suggesting C-Py-Pup as terminal sequence of the hairpin.

The structure of the hairpin on the 5′ side of the single-stranded sequence varied according to the RNA. In U2 RNA, but not in the other RNAs, a bulge loop is present on the 5′ side. In this loop, we find the sequence G-A-U-A-Cp homologous to A-A-U-Pu-Pyp (region g) present in the top loop of U1 and U4 RNAs. In all three cases, sequence g is on the 3′ side of the loop.

In conclusion, there are many similarities between the 3′ domain of U1, U4, and U5 RNAs and the central domain of U2 RNA. We propose to designate this common domain as 'domain A'. Besides a different location in the molecule, there are several other minor changes of domain A between U2 and the group of U1, U4, and U5 RNAs. These concern, in particular, sequences i and j and the general structure of the hairpin on the 5′ side of the domain. However, the homologies are large enough to make us consider that domain A is a structure common to the four RNAs (Figure 4).

Other possible similarities between U1, U2, U4, and U5 RNAs and specific features of U2 RNA. In the 5′ domain of U1, U4, and U5 RNAs, two sequences located in loops were common to the three RNAs: region e was a pyridimine-rich sequence containing modified nucleotides and region c contained the segment A-U-C-Pup. According to the sequencing data of Reddy *et al.* (1981), several modified nucleotides are present in the 5′ half of U2 RNA as well as tetranucleotides of the type Pu-Py-Py-Pup which could account for regions e and c, respectively. However, the demonstration that such regions are present in loops must await an experimental study of the secondary structure of the 5′ part of U2 RNA. Another sequence common to U1, U4, and U5 RNAs was that at the 3′ end of the RNA: C-U-$(Pu)_{OH}$ or, in certain variants, C-U-Pu-$(Pu)_{OH}$. U2 RNA terminates with C-C-A_{OH} so that a common 3′-terminal sequence might be: C-Py-Pu_{OH}.

In addition to common structural characteristics, the four RNAs have individual features. Those for U1, U4, and U5 RNAs were discussed previously (Krol *et al.*, 1981a). For U2 RNA, we may mention: (1) a high proportion of pseudouridines. Seven of them are in hairpin I and located in the base-pairs; (2) an additional hairpin (III) following domain A on its 3′ side, with a large loop containing the sequence U-C-C-Ap complementary to the splice junction; and (3) the presence of the triplet C-C-A_{OH} at the 3′ end of the molecule, which resembles the 3′ terminus of tRNA molecules. All these characteristics are true of chicken and rat.

The presence of the sequence U-C-C-Ap (3′→5′) in the loop of hairpin III should be commented upon since it was proposed that this tetranucleotide, also present in U1 RNA, may participate in the alignment of intron extremities during pre-mRNA splicing (Lerner *et al.*, 1980; Rogers and Wall, 1980; Lazar *et al.*, 1982). U-C-C-Ap is located in a single-stranded region in both U1 and U2 RNAs (it is absent from

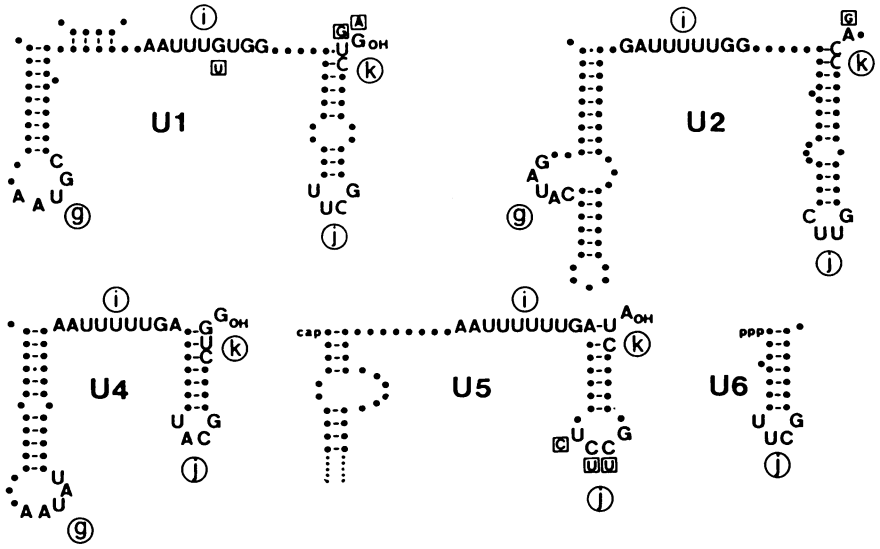

Fig. 4. The structure of domain A in U RNAs. The structures are those determined for rat RNAs. The structure of U1 RNA is from Branlant *et al.* (1981a) and takes into account the *Drosophila* data (Mount and Steitz, 1981). The model chosen for U4 RNA is model M2 (Krol *et al.*, 1981a). The 5′-terminal hairpin of U6 RNA was taken from Epstein *et al.* (1980). Nucleotides are represented by dots except for sequences similar in all or in certain RNAs (regions g, i, j, k). Nucleotide substitutions in these regions are indicated in small squares. They correspond, for U1 RNA to substitutions in *Drosophila*, for U2 RNA to substitutions in chicken, and for U5 RNA to a variant.

U4 and U5 RNAs). In U1 RNA, the single-stranded region is free, situated at the 5′ extremity of the RNA, accessible to enzymes specific for single-stranded regions (Branlant *et al.*, 1981a), and accessible for hybridization (Lazar *et al.*, 1982). In U2 RNA, the situation is different: the tetranucleotide is in a hairpin loop at the 3′ extremity of the molecule and apparently not accessible to the enzymes specific for single-stranded regions (Figure 2). This does not favour the idea that it may play the same role in the two RNAs.

Discussion

Domain A in the U RNAs

The nucleotide sequence of the 3′ half of U2 RNA has strongly diverged from chicken to rat. We only found 82.5% homology in contrast to 96 and 99.5% for U1 and U4 RNA, respectively (Branlant *et al.*, 1980; Krol *et al.*, 1981a). Owing to these divergences, we can propose a unique model of secondary structure on the basis of maximum base-pairing and secondary structure conservation. The same model was obtained when considering the results from limited digestion of U2 RNA with various nucleases. Such an agreement between two different approaches strongly supports the model and encouraged us to compare it with those previously proposed for U1, U4, and U5 RNAs.

The four RNAs share a common structure designated as 'domain A'. It consists of a free single-stranded region containing the sequence Pu-A-$(U)_n$-G-Pup (region i) flanked by two hairpins. That on the 3′ side is very stable and has the sequence Py-N-Py-Gp (region j) in the top loop. Domain A is well conserved throughout evolution. In spite of a large number of mutations, its secondary structure and the primary structure of sequences i and j are conserved between chicken and rat U2 RNA (present work) and between *Drosophila* and human U1 RNA (Mount and Steitz, 1981). Concerning region i of U1 RNA, it should be mentioned that its sequence is closer to the consensus sequence Pu-A-$(U)_5$-G-Pup in *Drosophila* than in man, rat or chicken, where a U is replaced by a G residue.

The homology between U1, U4, and U5 RNAs is not restricted to domain A, but clearly this domain is the most characteristic common feature of U1, U2, U4, and U5 RNAs. The existence of the same domain in four different RNA molecules suggests a common function. The study of snRNP shows that these four RNAs are bound to a common set of proteins (Lerner and Steitz, 1979) and it is tempting to consider that domain A is the binding site, or part of the binding site, of at least one of these proteins. Such an assumption is partly supported by the experimental results of Epstein *et al.* (1981) who showed that the 3′ half of U1 RNA is poorly digested by nuclease in U1 RNP, and those of Liautard, Sri-Widada, Brunel, and Jeanteur (personal communication) who observed a strong resistance to ribonuclease in domain A of U1, U2, U4, and U5 RNPs.

Since U6 RNA is present in hnRNP, and since its snRNP co-precipitates with those of U1, U2, U4, and U5 RNAs in the presence of anti-Sm (Lerner and Steitz, 1979), it might bear the same characteristic structural features, in particular domain A. A secondary structure based on maximum base-pairing was proposed by Epstein *et al.* (1980) and Harada *et al.* (1980). A stable hairpin resembling that of the 3′ end of domain A with the characteristic j sequence was found but was located at the 5′ end of the molecule (Figure 4). Furthermore, a sequence G-A-$(U)_5$ OH resembling the i sequence was located at the 3′ end of the molecule but was base-paired in the proposed models. It would be worth checking experimentally whether this is truly so and whether, in the spatial configuration of the U6 RNA molecule, the sequence similar to i and the 5′-terminal hairpin are in close vicinity, thus mimicking domain A.

U3 RNA does not co-precipitate with the other U RNAs in the presence of anti-Sm and does not contain an i sequence. It is located in the nucleolus (Prestayko *et al.*, 1971), associated with nucleolar RNA (Calvet and Pederson, 1981) and it is unlikely that it belongs to the same class as U1, U2, U4, U5, and possibly U6 RNAs.

The U RNA genes and pseudogenes

The presence of the same characteristic sequences in U1, U2, U4, and U5 RNAs suggests that they may have evolved from a common ancestor gene. The U RNA genes or pseudogenes were found to be numerous and dispersed throughout the genome (Denison *et al.*, 1981; Hayashi, 1981; Manser and Gesteland, 1981; Ohshima *et al.*, 1981b; Van Arsdell *et al.*, 1981; Westin *et al.*, 1981; Monstein *et al.*, 1982). On the other hand, Alu sequences from primates which are also dispersed throughout the genome bear a relatively well-conserved sequence $(T)_2$-$(A)_5$-Cp in their central region (Deininger *et al.*, 1981; Grimaldi *et al.*, 1981) which is complementary to sequence i, Pu-A-$(U)_5$-Gp. This observation leads to the hypothesis that a fraction of the Alu sequence might have evolved towards sequences carrying the U RNA genes. The comparison of the U RNA sequences flanking region i, and of a consensus human Alu sequence (Deininger *et al.*, 1981) is not conclusive in this respect. However, the results would be compatible with the idea that a small number of the 300 000 Alu sequences highly diverged to generate U RNA genes and pseudogenes.

Whatever their origin, it is likely that the U RNA genes have evolved under functional constraint. It is remarkable that various solutions were used for conserving certain structures between four RNAs of different lengths (117–189 nucleotides). For instance, a hairpin with a loop sequence common to U1 and U4 RNAs (region c) was reduced to a bulge loop containing sequence c in U5 RNA (Krol *et al.*, 1981a). On the other hand, the 43 nucleotides at the 3′ end of U2 RNA were arranged in a hairpin (III) which does not exist in U1, U4, and U5 RNAs. In all cases, domain A is conserved which makes its presence even more striking.

Several DNA fragments were found to hybridize with each U RNA and among the sequenced fragments many were not colinear to the expressed RNAs (Denison *et al.*, 1981; Hayashi, 1981; Manser and Gesteland, 1981; Ohshima *et al.*, 1981b; Van Arsdell *et al.*, 1981; Westin *et al.*, 1981; Monstein *et al.*, 1982). On the other hand, variant RNAs with one or several mutated nucleotides were described in one animal species indicating that several U RNA genes may be expressed (Krol *et al.*, 1981a, 1981b). As with the mutations occurring during evolution, those in variant RNAs in one species did not modify the secondary structure of the RNA, in sharp contrast to the mutations in many of the DNA fragments. A typical example is that of U2 RNA and the U2/4 DNA fragment. Whereas the 19 mutations occurring in the 3′ part of U2 RNA between chicken and rat did not affect the secondary structure, the 13 mutations in the corresponding part of U2/4 DNA profoundly modified this structure. Such DNA fragments which cannot give rise to a functional RNA (i.e., with a correct secondary structure) are most probably pseudogenes. Other conditions are certainly required for the expression of a U DNA gene, but the possibility of obtaining a functional RNA seems to be an absolute requirement.

Materials and methods

Isolation of RNA

Brain and liver nuclei from chicken or rat, or total chicken liver were phenol extracted at 0°C and deproteinized by phenol-chloroform (1:1) and chloroform-isoamylic alcohol (20:1). The RNA from nuclei was treated with ribonuclease-free deoxyribonuclease and centrifuged on a 5–20% sucrose gradient for 16 h at 24 000 r.p.m. in a SW27 rotor. The enzymatic treatment was omitted for total RNA. The 4–12S fractions were pooled, reprecipitated with ethanol, and fractionated on a 8–15% polyacrylamide gel gradient made up in 50 mM Tris, 40 mM borate, 0.5 mM EDTA pH 8.3 buffer containing 8 M urea. The bands were detected by u.v. fluorescence and excised. The RNA was eluted out by overnight shaking in 0.24 M NH_4 acetate in the cold, and the gel pieces were removed by centrifugation.

Nucleotide sequence analysis

We used the classical method based on statistical digestion of end-labeled molecules.

End-labeling. 3′ End-labeling: 1 μg RNA was labeled with 100 μCi of [5′-^{32}P]pCp 3000 Ci/mmol (Amersham) and 2–5 U of T4 RNA ligase (PL Biochemicals) for 1 h at 37°C in 10 mM $MgCl_2$, 2 mM dithiothreitol (DTT), 50 mM Tris-HCl pH 7.5 buffer in the presence of 2 nmol of ATP. 5′ End-labeling: 1 μg RNA was labeled with 100 μCi of [γ-^{32}P]ATP (Amersham) and 2–5 U of T4 polynucleotide kinase (PL Biochemicals) for 30 min at 37°C in 10 mM Mg acetate, 15 mM DTT, 10 mM Tris-HCl pH 8 buffer.

Obtention of partial digestion products. RNase T1 digestion: 1 μg U2 RNA was partially digested with T1 RNase in the presence of [γ-^{32}P]ATP and T4 polynucleotide kinase, so that the resulting fragments were 5′ end-labeled simultaneously. The digestion was performed in 10 mM Mg acetate, 15 mM DTT, 10 mM Tris-HCl pH 8 buffer for 7 min at 37°C with 1.5×10^{-2} U of T1 RNase. The labeled digested material was immediately fractionated on polyacrylamide gel. RNase H digestion: 1 μg of U2 RNA was hybridized with 2.5 μg of oligo(dGp)6 in 4 mM $MgCl_2$, 1 mM DTT, 40 mM Tris-HCl pH 7.9 buffer, for 30 min at 30°C. 2 U of RNase H were added and the mixture was incubated for another 30 min. The digested RNA material was precipitated and 3′ end-labeled.

Statistical digestion of end-labeled molecules. 5′ End-labeled material was partially digested with enzymes, 3′ end-labeled molecules with both enzymes and chemicals.

Enzymatic digestions: each digestion was performed on 2.5 μg RNA (tRNA was used as carrier) in 10 μl of buffer. The buffer described by Donis-Keller *et al.* (1977) was used for T1, U2, A, and *Physarum polycephalum* RNase digestions. Hydrolysis with *N. crassa* nuclease was achieved in 20 mM Tris-HCl pH 7.5 buffer in the presence of 7 M urea according to Krupp and Gross (1979). Digestions were achieved at 55°C for 30 min the case of U2 RNase and 15 min for all other enzymes. The amount of enzyme used was 2×10^{-2}–4×10^{-2} U for T1 RNase, 0.7 U for U2 RNase, 4×10^{-3} μg for RNase A, 1 U for *P. polycephalum*, and 0.5 μg for *N. crassa* nucleases.

Chemical digestions were performed according to Peattie (1979).

The digestion products were fractionated on 10–25% polyacrylamide slab gels (0.5 x 200 x 400 mm) in Tris-borate urea buffer.

Secondary structure study

The 3′ end-labeled molecules were partially digested with enzymes in non-denaturing conditions. 2.5 μg RNA were used for each assay. Digestion with T1 RNase was performed in 10 μl of 10 mM $MgCl_2$, 10 mM Tris-HCl pH 7.5 buffer for 5 min at 37°C with 2.5×10^{-2} U of RNase. The digestion was stopped by the addition of 1 μl of 400 mM NaH_2PO_4. Digestion with S1 nuclease was performed in 25 mM Na acetate, 10 mM Mg acetate, 50 mM KCl, 1 mM $ZnCl_2$ pH 4.5 buffer for 5 min at 37°C with 2.5 U of nuclease. The digestion was stopped by addition of 5 μl of 10 mM ATP. Digestion with *N. oxiana* RNase was performed in 350 mM KCl, 10 mM $MgCl_2$, 10 mM Tris-HCl pH 7.5 buffer for 5 min at 0°C with 0.1 U RNase.

Acknowledgements

We are grateful to F. Eyermann for excellent technical assistance. This work was supported by grants from the 'Centre National de la Recherche Scientifique', the 'Institut National de la Santé et de la Recherche Médicale' and the 'Délégation Générale de la Recherche Scientifique et Technique'.

References

Branlant,C., Krol,A., Ebel,J.P., Lazar,E., Gallinaro,H., Jacob,M., Sri-Wadada,J., and Jeanteur,P. (1980) *Nucleic Acids Res.*, **8**, 4143-4154.

Branlant,C., Krol,A., Ebel,J.P., Gallinaro,H., Lazar,E., and Jacob,M. (1981a) *Nucleic Acids Res.*, **9**, 841-858.

Branlant,C., Krol,A., Machatt,M.A., Pouyet,J., and Ebel,J.P. (1981b) *Nucleic Acids Res.*, **9**, 4303-4324.

Calvet,J.P., and Pederson,T. (1981) *Cell*, **26**, 363-370.

Deimel,B., Louis,Ch., and Sekeris,C.E. (1977) *FEBS Lett.*, **73**, 80-84.

Deininger,P.L., Jolly,D.J., Rubin,C.M., Friedman,T., and Schmid,C.W. (1981) *J. Mol. Biol.*, **151**, 17-33.

Denison,R.A., Van Arsdell,S.W., Bernstein,L.B., and Weiner,A.M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 810-814.

Donis-Keller,H., Maxam,A.M., and Gilbert,W. (1977) *Nucleic Acids Res.*, **4**, 2527-2538.

Epstein,P., Reddy,R., Henning,D., and Busch,H. (1980) *J. Biol. Chem.*, **255**, 8901-8906.
Epstein,P., Reddy,R., and Busch,H. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1562-1566.
Gallinaro,H., and Jacob,M. (1979) *FEBS Lett.*, **104**, 176-182.
Grimaldi,G., Queen,C., and Singer,M.F. (1981) *Nucleic Acids Res.*, **9**, 5553-5568.
Guimont-Ducamp,C., Sri-Widada,J., and Jeanteur,Ph. (1977) *Biochimie*, **59**, 755-758.
Harada,F., Kato,N., and Nishimura,S. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1332-1340.
Hayashi,K. (1981) *Nucleic Acids Res.*, **9**, 3379-3388.
Krol,A., Branlant,C., Lazar,E., Gallinaro,H., and Jacob,M. (1981a) *Nucleic Acids Res.*, **9**, 2699-2716.
Krol,A., Gallinaro,H., Lazar,E., Jacob,M., and Branlant,C. (1981b) *Nucleic Acids Res.*, **9**, 769-787.
Krupp,G., and Gross,H.J. (1979) *Nucleic Acids Res.*, **6**, 3481-3490.
Lazar,E., Jacob,M., Krol,A., and Branlant,C. (1982) *Nucleic Acids Res.*, **10**, 1193-1201.
Lerner,M.R., and Steitz,J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5495-5499.
Lerner,M.R., Boyle,J.A., Mount,S.M., Wolin,S.L., and Steitz,J.A. (1980) *Nature*, **283**, 220-224.
Manser,T., and Gesteland,R.F. (1981) *J. Mol. Appl. Genet.*, **1**, 117-125.
Miller,T.E., Huang,C.Y., and Pogo,A.O. (1978) *J. Cell Biol.*, **76**, 692-704.
Monstein,H.J., Westin,G., Philipson,L., and Petterson,U. (1982) *EMBO J.*, **1**, 133-137.
Mount,S.M., and Steitz,J.A. (1981) *Nucleic Acids Res.*, **9**, 6351-6368.
Northemann,W., Scheurlen,M., Gross,V., and Heinrich,P.C. (1977) *Biochem. Biophys. Res. Commun.*, **76**, 1130-1137.
Ohshima,Y., Itoh,M., Okada,N., and Miyata,T. (1981a) *Proc. Natl. Acad. Sci. USA*, **78**, 4471-4474.
Ohshima,Y., Okada,N., Tani,T., Itoh,Y., and Itoh,M. (1981b) *Nucleic Acids Res.*, **9**, 5145-5158.
Peattie,D. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1760-1764.
Prestayko,A.W., Tonato,M., Lewis,B.C., and Busch,H. (1971) *J. Biol. Chem.*, **246**, 182-187.
Raj,N.B.K., Ro-Choi,T.S., and Busch,H. (1975) *Biochemistry (Wash.)*, **14**, 4380-4385.
Reddy,R., and Busch,H. (1981) in Busch,H. (ed.), *The Cell Nucleus*, **VIIIA**, Academic Press, NY, pp. 261-306.
Reddy,R., Henning,D., Epstein,P., and Busch,H. (1981) *Nucleic Acids Res.*, **9**, 5645-5658.
Rogers,J., and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1877-1879.
Sri-Widada,J., Liautard,J.P., Assens,C., and Brunel,C. (1981) *Mol. Biol. Rep.*, **8**, 29-36.
Stiegler,P., Carbon,P., Ebel,J.P., and Ehresman,C. (1981) *Eur. J. Biochem.*, **120**, 487-495.
Van Arsdell,S.W., Denison,R.A., Bernstein,L.B., and Weiner,A.M. (1981) *Cell*, **26**, 11-17.
Westin,G., Monstein,H.J., Zabielski,J., Philipson,L., and Petterson,U. (1981) *Nucleic Acids Res.*, **9**, 6323-6338.
Zaug,A.J., and Cech,T.R. (1982) *Nucleic Acids Res.*, **10**, 2823-2838.
Zieve,G., and Penman,S. (1976) *Cell*, **8**, 19-31.