Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs

A.Krol and C.Branlant
Institut de Biologie Moléculaire et Cellulaire (CNRS), 15 rue René Descartes, 67084 Strasbourg Cédex, France, and

E.Lazar, H.Gallinaro and M.Jacob
Laboratoire de Génétique Moléculaire des Eucaryotes (CNRS), Unité de Biologie Moléculaire et de Génie Génétique (INSERM), 11 rue Humann, 67085 Strasbourg Cédex, France

Received 30 April 1981

ABSTRACT

U4 RNA from chicken, rat and man was examined for nucleotide sequence and secondary structure. Three molecular species, U4A, U4B and U4C were detected in the three animal species. U4A is 146 nucleotide long and U4B RNA only lacks the 3' terminal G. Four nucleotides are missing at the 3'-end of U4C RNA which, in addition, differs from U4A and U4B RNAs at two internal positions. Thus, U4C RNA is encoded by another gene as U4A and U4B RNAs. Only one nucleotide substitution occurred between chicken and man showing that U4A, U4B and U4C RNAs have been extremely conserved throughout evolution. The three molecular species are capped, they contain three ψ, a 2'-O methyl A and a m^6A. An additional post-transcriptional modification close to the cap structure is observed in man. On the basis of an experimental study, two models of secondary structure may be proposed for U4 RNA. The 3' domain is the same in both models and is homologous to that of U1 and U5 RNAs. It consists of a single-stranded region, containing the sequence Py-$(A)_2$-$(U)_n$-Gp flanked by two stable hairpins probably involved in tertiary interactions. The 5' domain is less stable than the 3' domain and its structure is different in the two models. However, a long single-stranded pyrimidine region containing modified nucleotides is found in both models as in U1 and U5 RNAs. Several other nucleotide sequence homologies related to specific features of secondary structure suggest that U1, U4 and U5 RNAs derive from a common ancestor and may have common function.

INTRODUCTION

The function of the metabolically stable small nuclear RNA designated as U1 to U6 RNAs (1) is essentially unknown. Their presence in the ribonucleoproteins containing premessenger RNA (hnRNP) suggests that they may play a role in the processing of premessenger RNA (2-5) and a function in splicing has been proposed for U1 RNA (1, 6, 7). The comprehension of the function of all small RNA molecules first requires the knowledge of their primary and secondary structures. The nucleotide sequence of all snRNA molecules has been determined (8-15) with the exception of that of U4 RNA. On the other hand, our experimental study of the secondary structure of U1 and U5 RNAs (15, 16) showed striking homologies between these two RNA molecules and it was of interest to know whether such conclusions could be extended to other snRNA species.

To this end, we determined the nucleotide sequence of chicken, rat and human U4 RNA and we studied the secondary structure of these molecules.

MATERIALS AND METHODS

The methods used in this work have already been fully described. They will only be briefly summarized here.

1. Isolation of U4 RNA

The RNA was phenol-extracted at 0-4°C, pH 7.6 from purified nuclei, nuclear extracts or nuclear residues of rat and hen brain, hen liver and HeLa cells (17, 18). They were fractionated on 15 % polyacrylamide gel, made up in Tris-borate buffer (19).

2. Sequence analysis of U4 RNA

2.1. Identification of modified nucleotides. The RNA was digested with a mixture of T1 and T2 RNases. The resulting products were 5'-end labeled. Their 3'-phosphate was eliminated by action of P1 nuclease. The resulting (5'-^{32}P) mono and dinucleotides were analyzed by two-dimensional thin-layer chromatography according to Nishimura (20).

2.2. 5' and 3'-end labelings were achieved with (γ-^{32}P)ATP and T4 polynucleotide kinase and (5'-^{32}P) pCp and T4 RNA ligase, respectively. When the entire U4 RNA molecule was 5'-end labeled, the cap structure was first eliminated by action of tobacco pyrophosphatase and alkaline phosphatase (14).

2.3. Preparation of end labeled partial digestion products of U4 RNA.
2.3.1. U4 RNA was partially digested with T1 or S1 nucleases (14). The products resulting from digestion with the former nuclease were 5'-end labeled, those resulting from digestion with the latter were 3'-end labeled.
2.3.2. U4 RNA was hybridized to oligo(dGp)$_6$. The hybrid was cleaved with RNase H (16, 21) and the resulting products were 3'-end labeled.

2.4. Sequence analysis of end labeled RNA was achieved with both the enzymatic (14, 22) and the chemical method (23) for RNA sequencing. The resulting products were fractionated on thin polyacrylamide gel in Tris-borate buffer (19).

3. Secondary structure of U4 RNA

The enzymatic method that we recently developed was used (16).

RESULTS

1. Sequence analysis of U4 RNA

Small nuclear RNAs extracted from rat and hen brain, hen liver and

HeLa cells were electrophoresed in 15 % polyacrylamide gel. U4 RNA was a diffuse band migrating between U1 RNA and 5 S RNA (15). The RNAs from this band were eluted out for analysis.

1.1. Identification of the total digestion products from T_1 and T_2 RNases. Fractionation of these products in thin-layer chromatography revealed the presence of ψ, m^6A and Am residues in U4 RNA (Fig. 1).

1.2. Sequence analysis of the 5'-end labeled molecules. When U4 RNA was treated with alkaline phosphatase and then incubated with (γ-^{32}P)ATP and polynucleotide kinase, no labeling was obtained. Only when the phosphatase digestion was preceded by a tobacco acid pyrophosphatase treatment did we observe labeling. This observation demonstrates the presence of a cap structure at the 5'-end of U4 RNA.

The 5'-end labeled molecule was analyzed by the enzymatic method for RNA sequencing. This allowed the determination of the sequence of the 31 nucleotides located after the 2'-O methylated nucleotide(s) of the cap structure (Fig. 4). Rat and hen U4 RNAs have the same sequence at their 5'-end : cap-G-C-U-U-U-G. In the case of HeLa cells, the G residue next to the cap structure is replaced by a nucleotide which has not been identified.

1.3. Sequence analysis of the 3'-end labeled molecules. 3'-end labeled U4 RNA was analyzed by both the chemical and enzymatic methods for RNA sequencing. In the three animal species, three types of molecules (U4A, U4B and U4C RNAs) were observed (Fig. 4). They were terminated by A-G-A-C-U-G-G_{OH} (U4A), A-G-A-C-U-G_{OH} (U4B) and A-G-A_{OH} (U4C). The 3'-end heterogeneity explains the broad band obtained at the level of U4 RNA upon fractionation of the mixture of snRNAs. In addition to the differences at the 3'-extremity,

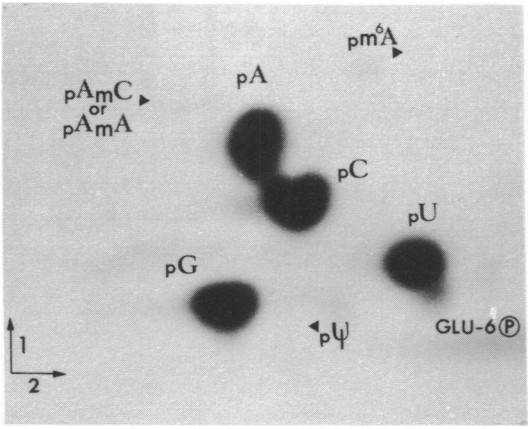


Figure 1 : Analysis of modified nucleotides of HeLa U4 RNA. The RNA was digested with T1 and T2 RNases. The resulting products were 5'-end labeled and dephosphorylated at their 3'-end. They were then fractionated in the chromatographic system described by Nishimura (20).

we found that the C residue at position 89 and the G residue at position 100 of U4A RNA and U4B RNAs were both replaced by a A residue in U4C.

In all three U4 RNA species, nucleotides 73 and 80 were cleaved by pancreatic RNase but not by aqueous or anhydrous hydrazine (Fig. 2). This indicates the presence of pseudouridines whose existence was also shown after digestion with T_1 and T_2 RNases (Fig. 1). Strand scission occurred at nucleotide 66 after diethylpyrocarbonate treatment, allowing its identification as an adenylic acid. However, no enzymatic cleavage or digestion in boiling water was observed at this position, indicating a 2'-O methylation. Thus, nucleotide 66 is a 2'-O methyl adenylic acid in agreement with the finding of the dinucleotide pAm-A or pAm-C after T_1 and T_2 RNases digestion.

The 3'-terminal regions of U4A and U4B RNAs were strongly resistant to enzymatic digestion. In addition, band compression occurred upon fractionation of the enzymatic or chemical digestion products in polyacrylamide gels. This

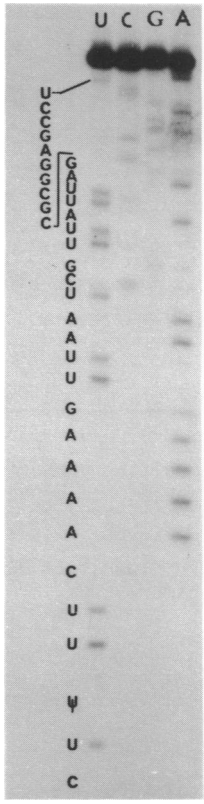


Figure 2 : Sequencing gel of 3'-end labeled fragment R1 from rat U4 RNA. Fragment R1 resulted from RNase H digestion of the hybrid U4 RNA-(dGp)6 (Fig. 4). It was chemically digested according to Peattie (23).

was also the case for the 3'-end region of U1 and U5 RNAs (14, 15) and may similarly be explained by a very stable secondary structure.

1.4. Sequence analysis of U4 RNA partial digestion products. U4 RNA was partially digested with S1 nuclease. The resulting products were 3'-end labeled and analysed by the chemical method for RNA sequencing (Fig. 3). On the other hand, U4 RNA was partially digested with T_1 RNase and the digestion products were 5'-end labeled during digestion. They were then analyzed by the enzymatic method for RNA sequencing. Finally, U4 RNA was hybridized to $(dGp)_6$ and then digested with RNAse H. The resulting products were 3'-end labeled and analysed by the chemical and enzymatic methods for RNA sequencing (Fig. 2).

The study of all these partial digestion products completed and verified the results obtained upon 3' and 5'-end labeling of undigested U4 RNA (Fig. 4). In addition, analysis of the RNase H digestion products R1 and R2 and of the S1 digestion products S2 and S4 revealed the presence of a modified uridine at position 5 in the three U4 RNA species. Indeed, this pyrimidine was not cleaved by aqueous or anhydrous hydrazine (Fig. 3), and therefore, should be a pseudouridine. Furthermore, in the case of HeLa cells, nucleotide 3 was not cleaved

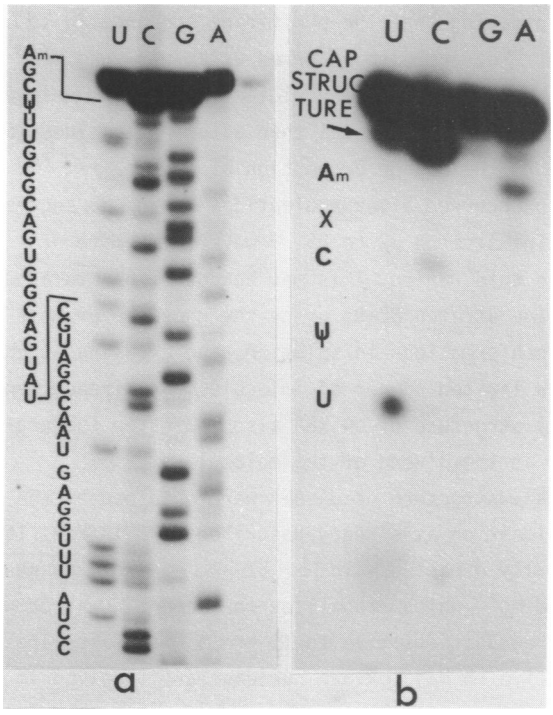


Figure 3 : Sequencing gels of 3'-end labeled fragment S1 from rat U4 RNA (a) and fragment S4 from man U4 RNA (b). Fragments S1 and S4 were obtained upon limited digestion of rat U4 RNA and man U4 RNA, respectively (Fig. 4). In Fig. 3b, the characteristic cleavage pattern of the cap structure is observed as well as the absence of chemical cleavage at two nucleotides. One of them has been identified as a ψ residue as it is cleaved by pancreatic RNase and a second one has not been identified and is designated as X.

by any of the chemical reagents and, therefore, is also a modified nucleotide. It was not identified yet (X in Fig. 4). Treatment of 3'-end labeled fragments R1, R2, S1, S2 and S4 (Fig. 4) with hydrazine produced the characteristic cleavage pattern observed at the level of U1 and U5 RNA cap structure and this confirms the presence of a cap structure in U4 RNA (Fig. 3). We did not identify the nucleotide denoted N at the extreme 5'-end. Since m^6adenylic acid is recognized by U2 RNase and is cleaved after diethylpyrocarbonate treatment, like a normal adenine residue, we failed to localize the m^6A residue detected upon total digestion of U4 RNA.

The complete nucleotide sequences of U4A, U4B and U4C RNAs from chicken, rat and HeLa cells are shown in Figure 4.

2. Secondary structure of U4 RNA in solution

2.1. Experimental study. Some informations on the secondary structure of U4 RNA were provided by the study of its primary structure. Thus, the extremities of the partial digestion products of T1 and S1 nucleases correspond to unpaired nucleotides (Fig. 4). The existence of RNase H-sensitive regions after hybridization of U4 RNA to $(dGp)_6$ indicates single-stranded stretches of at least four pyrimidines (21). The major RNase H cleavage positions were after C75 and C76 indicating that the pyrimidine sequence 70 to 77 was available for hydridization. Cleavage also occurred after U121 and U122 showing that the uridine stretch 121-125 could hybridize to $(dGp)_6$. The level of cleavage was lower after nucleotides 120 and 121 than after nucleotides 75 and 76 and this is probably due to the higher proportion of dG.U pairs. Finally, in some experiments, cuts occurred after nucleotides 84 and 85 suggesting hybridization to the cytosine stretch 82 to 85. However, the rate of hydrolysis by RNase H was low in this region. This may be due either to a poor accessibility of the cytosine stretch 82-85 or to the presence of molecules of U4 RNA with different conformations in solution, the cytosine stretch 82-85 being accessible only in a limited number of molecules. As already mentioned, the study of the primary structure of U4 RNA also indicated the presence of very stable base-pairs at the 3'-end of the molecules.

More data on the secondary structure of U4 RNA were obtained by the systematic study of enzymatic digestion as already described for U1A RNA (16). 3'-end labeled U4 RNA was partially digested with T1, S1 and Naja oxiana nucleases in the presence of 10 mM Mg^{++} which stabilizes the secondary structure. Under such conditions, the regions sensitive to T1 and S1 nucleases are single-stranded and those sensitive to Naja oxiana nuclease are involved in RNA-RNA interactions (24). The digestion products were then fractionated in

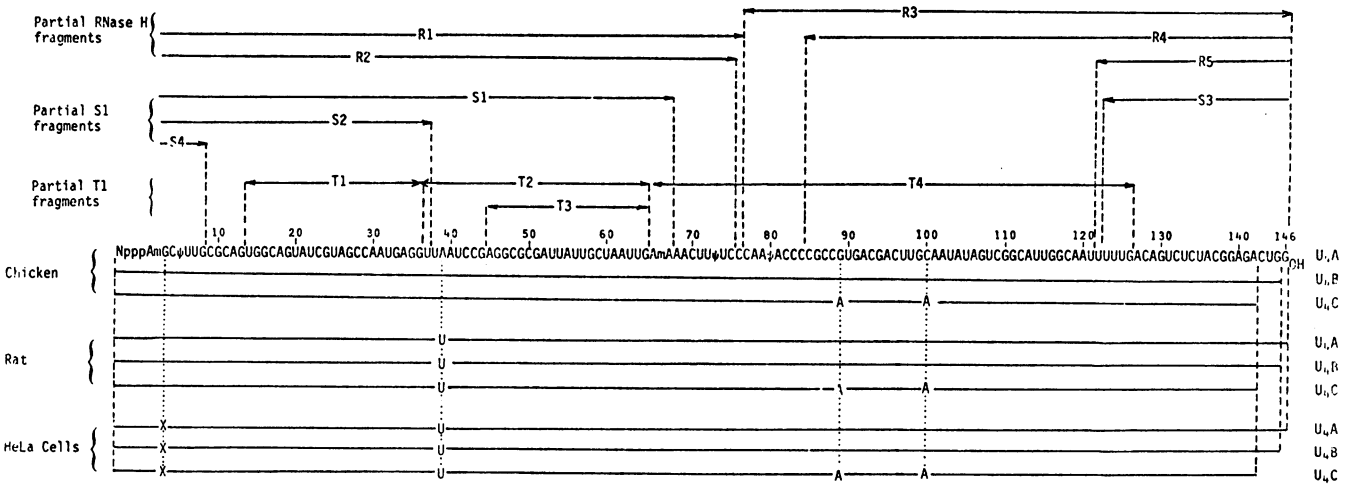


Figure 4 : The nucleotide sequence of U4 RNAs from chicken, rat and HeLa cells. The sequence of chicken U4A RNA is given. Only the substitutions and deletions observed in the other U4 RNA studied are indicated. X is an unidentified modified nucleotide and N is the nucleotide of the cap structure which has not been identified either. The T1, S1 and RNase H digestion products used to establish the sequences are indicated, and are designated by Tn, Sn and Rn, respectively.

polyacrylamide gels and identified by comparison with a T1 RNase digest performed under denaturing conditions and with a total digest in boiling water (Fig. 5).

2.2. Models. The complementary sequences of hen, rat and human U4 RNAs were listed from the primary structure using computer analysis. This allowed the construction of secondary structure models. Three of them, denoted M1, M2 and M3 were in relative agreement with the results of enzymatic digestion of U4 RNA (Fig. 6). The free energy of the structures were calculated according to Salzer (25). M1 was found to be more stable than M2 and M3 (-35 Kcal as compared to -27 kcal). However, as already mentioned (16, 26), the estimation of the destabilizing effect of asymetrical loops or of G.U pairs was not experimentally defined so that the values must be considered as rough estimates. In addition, the calculation does not take into account the stabilizing energy provided by tertiary interactions and, therefore, it cannot be inferred that the structure with the lowest free energy calculated as above corresponds to the true structure.

Therefore, the three models were reexamined on the basis of the experimental results. Model M1 is in best agreement with the positions of preferential T1 and S1 nuclease cleavage. In particular, nucleotides between positions 32 to 45 in a hairpin loop were attacked by these enzymes. In contrast, there was no enzymatic cleavage of the single-stranded region 18-26 of model M2 nor of the single-stranded regions 18-26 and 43-50 of model M3 (with the exception of G44). However, this does not obligatorily exclude models M2 and M3 since tertiary interactions involving the loops may also prevent enzyme action.

Model M3 is the only one in perfect agreement with the results of Naja oxiana nuclease digestion. In particular, the major cut after nucleotide 32 is difficult to explain in model M1. Base-pairing within the loop between the A-Up dinucleotides 31-32 and 40-41 would have to be assumed, or tertiary interactions.

Model M2 is in best agreement with the results of RNase H digestion of the hybrid between U4 RNA and $(dGp)_6$. For all three models, the pyrimidine stretches 70-76 and 121-125 corresponding to major and minor cleavage sites are free of base-pairing. The situation is different for the cytosine stretch 82-85. It is accessible in models M1 and M3 and should thus base-pair to $(dGp)_6$ and be cleaved by RNase H at the same extent as the pyrimidine stretch 70-76. Experimentally, a much lower yield of hydrolysis was observed. This is unlikely to be due to tertiary interactions which are expected to be destro-

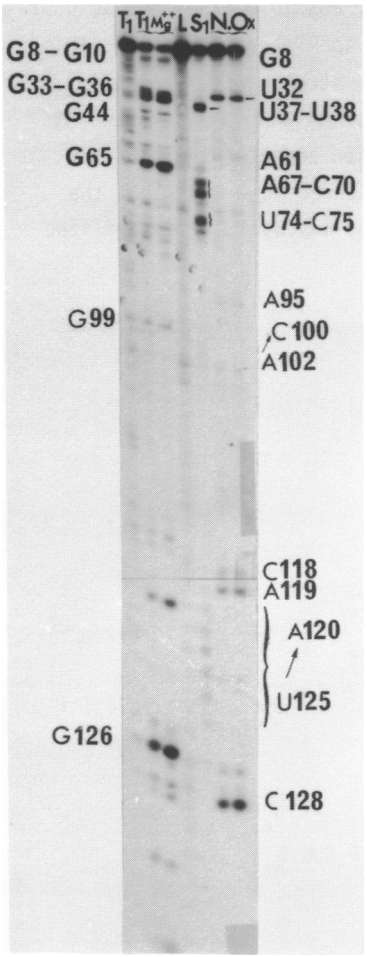
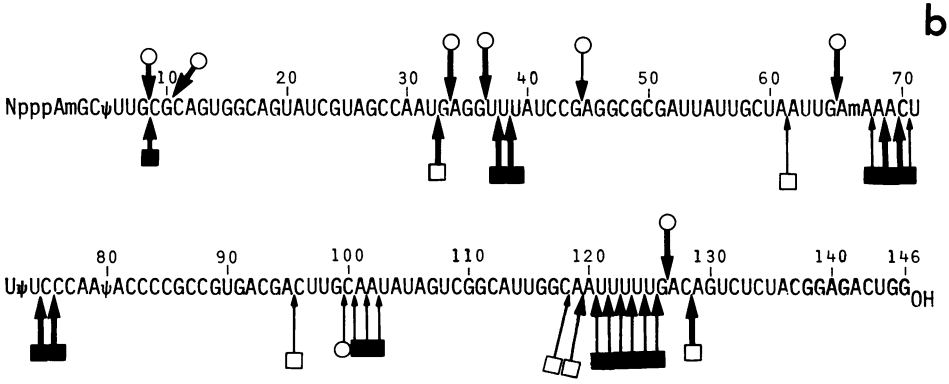


Figure 5 : Experimental study of the secondary structure of HeLa cells U4A RNA.
a) The RNA was partially digested with T1 (T1 Mg^{++}), S1 and Naja oxiana (N. ox) nucleases in the non-denaturing conditions previously described (16). The resulting fragments were fractionated by polyacrylamide gel electrophoresis and were identified by their mobilities relative to those of products released by statistical digestion with T1 RNase (T1) and boiling water (L).
b) Schematic representation of the results.

S1 nuclease
T1 RNase
Naja oxiana RNase



yed by the heating at 60°C prior to hybridization. Possibly, the low yield of hydrolysis might be related to a steric hindrance to RNase H digestion. This also seems unlikely since, in U1 RNA, a pyrimidine stretch with a position homologous to that of the cytosine stretch 82-85 of U4 RNA models M1 and M3 was found to be perfectly accessible to hybridization and cleavage (16). In model M2, the cytosine stretch 82-85 is hydrogen-bonded to G-G-C-G but the resulting base-paired region is not very stable and may well be destabilized

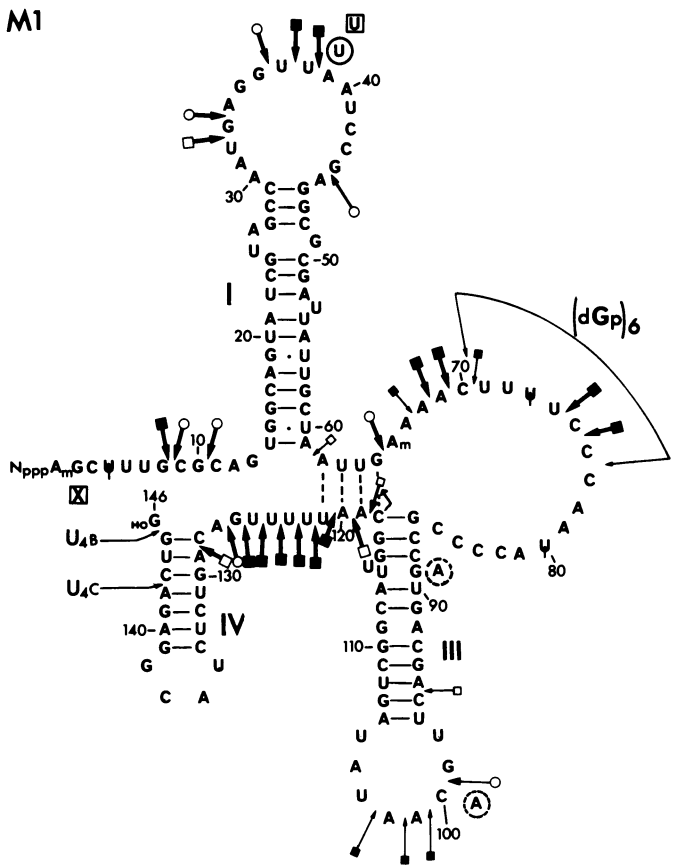
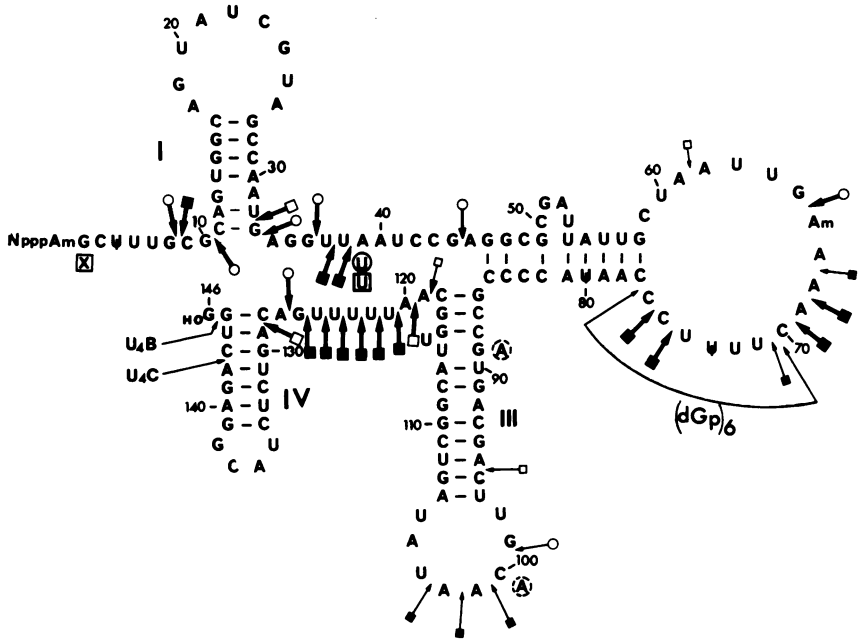
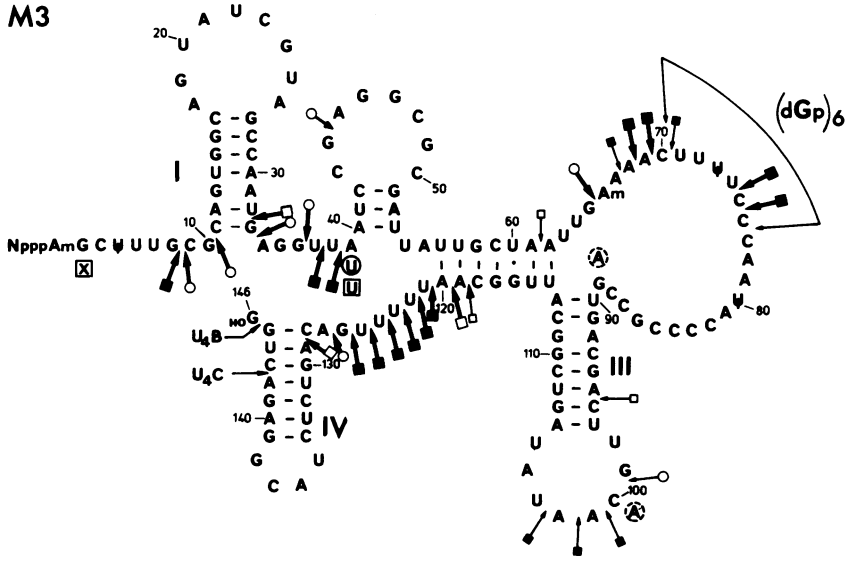


Figure 6 : Possible secondary structures of U4 RNA. The sequence is that of chicken U4A RNA. The differences existing in rat ◯ and man □ U4A RNA species are indicated, as well as those in U4C RNA ◌ . The arrows indicate the positions of preferential T1 ←○, S1 ←■ and Naja oxiana ←□ nuclease cleavages and the extremities of U4C and U4B RNAs ← . The region hybridized to $(dGp)_6$ and highly digested with RNase H is shown. In model M1, dotted lines indicate possible but rather unstable base-pairings.

M2


NpppAmGCΨUUGCG
HOG
U4B
U4C
I
II
III
IV
(dGp)6
Am

M3


NpppAmGCΨUUGCG
HOG
U4B
U4C
I
II
III
IV
(dGp)6
Am

by heating at 60°C. This may explain why hydrolysis was observed but with a low yield.

This discussion clearly shows that none of the three models is in perfect agreement with the experimental data. Model M3 is the less probable. It does not fit well with the results of digestion with T1 and S1 nucleases nor with those of RNase H digestion of hybrids. In addition, the stem of hairpin II is made of four base-pairs only in hen U4 RNA. The suppression of an A-U pair in rat and human U4 RNAs would make it very unstable and such a mutation destabilizing the secondary structure is rather unusual. Models M1 and M2 remain the most probable. As already mentioned, some of the discrepancies between the models and the experimental results may be accounted for by the existence of tertiary interactions involving in particular hairpin III and IV. The results of RNase H digestion of the U4 RNA-$(dGp)_6$ hybrids suggest that in 10 mM $MgCl_2$-10 mM Tris-HCl, pH 7.4, buffer, U4 RNA molecules have the structure M2. Possibly another structure may predominate under different conditions.

It should be pointed out that, in contrast to model M3, models M1 and M2 are not altered by the replacement of A39 of hen RNA by a U residue in rat and human RNAs nor by the replacement of G89 and C100 by A residues in U4C RNA. In the particular case of U4C RNA, four nucleotides are missing at the 3'-end as compared to U4A RNA. The 3'-terminal hairpin is shorter and less stable in U4C than in U4A or U4B RNA in agreement with the experimental results showing band compression for region 129-132 in U4A and U4B but not in U4C RNA.

3. Comparison of the primary and secondary structures of U1, U4 and U5 RNAs

Inspection of the probable secondary structures of U1, U4 and U5 RNAs showed the presence of similar oligonucleotides in specific regions of the three molecules. In order to appreciate the extent of homology in the three RNA species, their nucleotide sequences were aligned in such a way that the characteristic features of secondary structure could be compared (Fig. 7). The extent of nucleotide sequence homology related to secondary structure was higher with model M2 of U4 RNA than with model M1 or M3, so that only model M2 was taken into consideration for the comparison. Eleven regions of homology which seemed of particular interest were observed and were designated as a to k.

Regions c, i, j and k display homologies in the three RNA molecules. Region c covers the sequence of the first hairpin loop (U1 and U4 RNAs) or bulge loop (U5 RNA) starting from the 5'-end. The other regions common to the

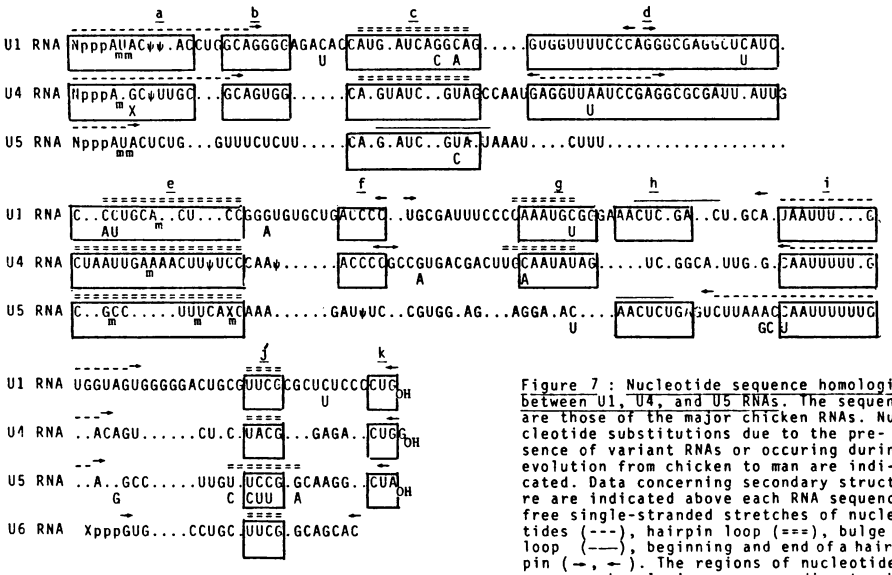


Figure 7 : Nucleotide sequence homologies between U1, U4, and U5 RNAs. The sequences are those of the major chicken RNAs. Nucleotide substitutions due to the presence of variant RNAs or occuring during evolution from chicken to man are indicated. Data concerning secondary structure are indicated above each RNA sequence : free single-stranded stretches of nucleotides (---), hairpin loop (===), bulge loop (—), beginning and end of a hairpin (→, ←). The regions of nucleotide sequence homologies corresponding to characteristic features of the secondary structure are boxed. The sequence of the 5' end hairpin of U6 RNA (11, 12) is given for comparison.

three RNA species are in the 3'-half of the molecule : region i corresponds to a free single-stranded sequence Py-$(A)_2$-$(U)_n$-Gp, region j to the loop of the 3'-terminal stable hairpin (Py-X-Py-Gp) and region k to the trinucleotide C-U-Pup found at or close to the 3'-end of all molecules. Region e belongs to the same class of sequences not because of true sequence homology but because of the presence of one or several modified nucleotides in a pyrimidine-rich sequence of a hairpin loop. Regions of homology a, b, d, f and g are common to U1 and U4 RNAs only. Among them, region a corresponds to the free single-stranded sequence at the 5'-end of the molecule. It may be remarked that the two modified uridines described previously at positions 6 and 7 of U1 RNA (14) were identified as pseudouridines (unpublished results) in agreement with recent results of Reddy et al. (27) and that a pseudouridine was found at a homologous position in U4 RNA. Region f is 26 nucleotide long. 22 nucleotides are conserved either as such (sixteen) or as pyrimidine or purine nucleotides (six). Finally, region h, corresponding to a bulge loop is common to U1 and U5 RNAs. Other sequence homologies are apparent upon examination of the sequences but do not correspond to specific features of secondary structure and are not indicated in the figure.

Structural similarities between U1, U4 (model M2) and U5 RNAs are sum-

marized in Fig. 8 and suggest that the three RNAs belong to a same family. Such similarities are not likely to have occurred by chance and are probably biologically significant.

DISCUSSION

U4 RNAs from chicken, rat and man are heterogeneous in size. In each animal species and for chicken, in both liver and brain, we found three molecules of 146, 145 and 142 nucleotides designated as U4A, U4B and U4C, respectively. U4B only differs from U4A RNA by the absence of the 3'-terminal G. U4C RNA lacks the terminal sequences C-U-G-G of U4A RNA and, in addition, two internal nucleotides are substituted. This indicates that there are at least 2 genes coding for U4 RNA which are both expressed. The variations at the 3'-end might also be due to the structure of the genes, but other possibilities may be evoked such as early arrest of transcription or post-transcriptional cleavages. Wha-

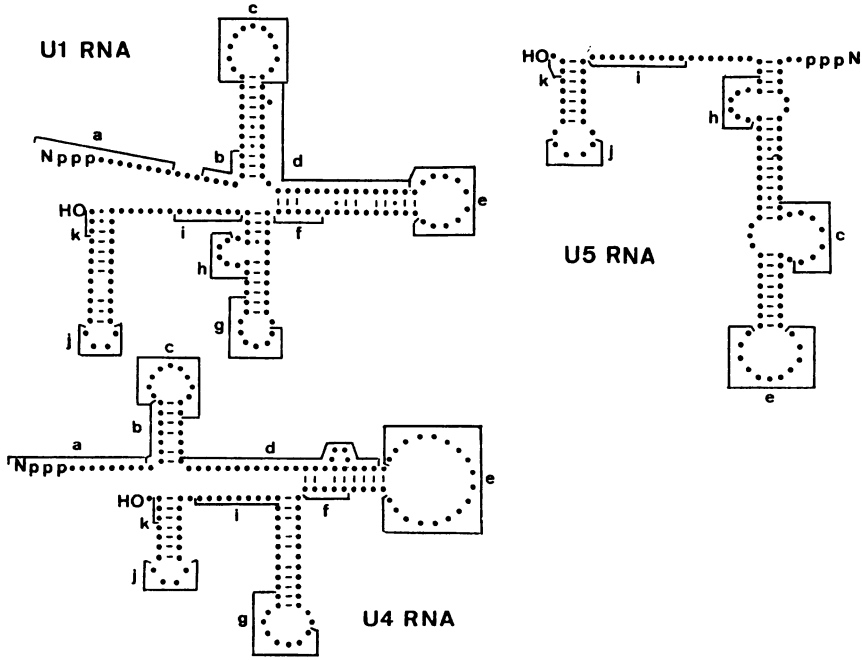


Figure 8 : Homologies between the secondary structures of U1, U4 and U5 RNAs. Model M2 is given for U4 RNA. Nucleotides are represented by black dots. Regions of homology are indicated. The nucleotide sequences of regions a to k correspond to those of Figure 7.

tever the cause of heterogeneity at the 3'-end of the molecule, it is striking that the three forms of U4 RNA have been conserved during evolution from chicken to man which suggests that they are biologically significant. U5 RNA was also shown to be encoded by a multigene family, but a larger number of variant RNAs was detected (15). In fact, there is very little change in the structure of U4 RNA during evolution from chicken to man as only one nucleotide substitution (position 39) was observed from bird to mammal. The change at nucleotide 3 between rat and human cells is more likely to be due to an additional post-transcriptional modification in man than to mutation. Therefore, U4 RNA is better conserved than U1 RNA (9 nucleotide substitutions between chicken and man) or than U5 RNA where several substitutions and insertions were observed in different RNA subspecies (14, 15).

As shown in Figure 7, there are large homologies of primary structure between U1, U4 and U5 RNAs. In spite of it, their base composition differ widely. The ratio G + C/A + U is the highest in U1 RNA : 1.4 against 0.9 and 0.7 for U4 and U5 RNA, respectively. U5 RNA has the highest uridine content : 34 % against 23 % for U1 and 26 % for U4 RNA. Thus, base composition is not sufficient for the determination of families in the snRNA population.

On the sole basis of nucleotide sequence, several models of secondary structure may be built for U4 RNA. Our experimental study allowed the selection of two of these models (M1 and M2) which differ in the structure of their 5'-half. If we assume strong tertiary interactions in this region, then model M2 is the most probable. Another argument in favor of model M2 is provided by the comparative study of the models of secondary structure of U4 RNA and of those previously proposed for U1 and U5 RNAs (15, 16). A 3' and 5' domain of different stability may be defined in these structures. In the three RNA molecules, the 3' domain overlaps the regions of homology g to k (Fig. 7 and 8). It is made of a single-stranded nucleotide stretch (containing the sequence Py-$(A)_2$-$(U)_n$-Gp) flanked by two hairpins. The 3' side hairpin is very stable in all three RNA species and its loop contains the sequence Py-X-Py-Gp. The 5'-side hairpin is confounded with the 5'-domain in U5 RNA. When individualized as in U1 and U4 RNAs, the upper stem of the hairpin also is quite stable. The stability of the 3'-domain is likely to be enhanced by tertiary interactions between constituents of the two hairpins (15, 16). This stable 3'-domain has the same structure in model M1 and M2 for U4 RNA and there is a high probability that the proposed structure is that of U1, U4 and U5 RNAs species in solution.

The situation is different for the 5' domain whose structural characteristics are less well defined than those of the 3' domain. The major featu-

res of the three RNA species in the 5' domain are as follows : a) there is an homologous sequence adjacent to the cap structure at the 5'-end (region a). It is free of base pairing in U1 and U4 RNAs (M1 and M2) but not in U5 RNA. b) another region of homology (c) is located in the loop of hairpin I in U1 and U4 (M2) RNAs and in a bulge loop in U5 RNA. In model M1 for U4 RNA, this nucleotide stretch is base-paired. c) the single-stranded pyrimidine-rich sequence containing modified nucleotides (region f) is found in a hairpin loop in U1, U4 (M2) and U5 RNAs, but not in U4 (M1) RNA. In summary, two loops with homologous sequences are present in the 5' domain of U1 RNA, U5 RNA and U4 RNA model M2. Only one hairpin loop whose sequence does not resemble that of the other RNA species is detected in model M1 of U4 RNA. This is by no way an evidence for the true existence of model M2 but only an argument. It should also be recalled that the secondary structures were determined in solution and may be different for the RNA molecules *in vivo* where a certain conformation may be stabilized by proteins.

The sequence homologies (Fig. 7) strongly suggest that U1, U4 and U5 RNAs derive from a common ancestor. The differentiation of this primitive sequence into U1, U4 and U5 RNAs led to the conservation of specific conformational features which suggest common functions. Homologies like those observed between the three snRNA species were found for regions of E. Coli ribosomal RNA and messenger RNA which bind the same ribosomal protein (26, 28). The three snRNA species may also bind the same proteins in agreement with the finding that the same polypeptides were associated with various snRNA species in immunologically precipitated snRNP (1, 29) or in snRNP prepared from hnRNP (30).

U1, U4 and U5 RNAs also bear specific sequences and certain may be related to specialized functions. For instance, it was shown that, in U1 RNA, the single-stranded sequence adjacent to the cap structure contains the oligonucleotide A-C-C-Up which is complementary to the two extremities of premessenger RNA introns and, therefore, may serve to the correct alignment of sequences along the splice point (1, 6, 7). The tetranucleotide A-C-C-Up is absent from the 5'-extremity of U4 RNA in spite of other large homologies with the sequence of U1 RNA (Fig. 7, region a). Thus, a different function must be envisaged. In the same way, small variations specific to one of the RNAs may be of functional importance. This may be the case of the change of the number of uridines in the sequence Py-$(A)_2$-$(U)_n$-G (region i) and of the individual variations of the pyrimidine-rich stretch containing modified nucleotides (region e). In addition, the appearance or disappearance of new single-stranded regions in one of the RNAs may reveal or obliterate some of the functions.

The question can be raised whether the other small nuclear RNA species belong to the same family as U1, U4, U5 RNAs. Several of these RNAs were sequences : 4.5 S RNA_I, 4.5 S RNA associated with poly(A)$^+$ RNA, 4.8 S or U6 RNA, U2 RNA and U3 RNA (8-15). Five criteria were used for comparison : 1) and 2) nucleotide sequence at the 3' and 5'-end of the molecules, 3) presence of the sequence Py-$(A)_2$-$(U)_n$-G, 4) presence of modified nucleotides in a single-stranded, pyrimidine-rich region, 5) presence of a stable hairpin at the 3'-end of the molecules. Criteria 1, 2, 3 and 5 were fulfilled in no case (in the absence of data of secondary structure, criterion 5 could not be applied to U2 RNA). Series of modified nucleotides were described in U2 and U6 RNAs but were not present in a pyrimidine-rich region so that criterion 4 was not fulfilled either. We conclude that U1, U4 and U5 RNAs constitute a special class of snRNAs. The examination of the sequences of the 4.5 S-4.8 S RNAs showed that the three species were terminated by Pu-Py-A-$(Py)n_{OH}$. This suggests that these RNA species may constitute another class of small RNAs. Another observations may be of interest : a hairpin was described at the 5'-end of U6 RNA (11, 12). Its loop as well as part of the stem have the same sequence as the homologous regions of the 3'-terminal hairpin of U1 RNA (Fig. 7). This suggests the possible binding to a same protein but experimental data are certainly required to confirm such possibility.

FOOTNOTE

After completion of this work, Reddy et al. (31) sent us a preprint of a paper on the nucleotide sequence of rat U4 RNA. This sequence agrees well with ours except that it is slightly shorter. The dinucleotide G-Cp (8-9) and the tetranucleotide G-A-C-Ap (126-129) are missing. As mentioned in the text, the tetranucleotide is in a region of band compression so that it may well escape detection. We also detected one more sequence heterogeneity (position 100). The work of Reddy et al. brings additional informations as compared to ours : the m^6A residue was located at position 101 (in our nomenclature) and the nucleotide N from the cap structure was identified as $m_3^{2,7,7}G$.

ACKNOWLEDGEMENTS

We are grateful to Dr J. Pouyet for computing complementary sequences of U4 RNA, to Drs P. Carbon and S. Vassilenko for the generous gift of Naja oxiana nuclease, to Dr M. Pinck for providing us with tobacco pyrophosphatase and to Dr R. Reddy for making his data available to us before publication. The expert technical assistance of L. Kister and B. Muller is acknowledged.

REFERENCES

1. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Nature 283, 220-224.

2. Deimel, B., Louis, C. and Sekeris, C.E. (1977) FEBS Lett. 73, 80-84.
3. Northemann, W., Scheurlen, M., Gross, V. and Heinrich, P.C. (1977) Biochem. Biophys. Res. Commun. 76, 1130-1137.
4. Guimont-Ducamp, C., Sri-Widada, J. and Jeanteur P. (1977) Biochimie 59, 755-758.
5. Gallinaro, H. and Jacob, M. (1979) FEBS Lett. 104, 176-182.
6. Rogers, J. and Wall, R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77 , 1877-1879.
7. Gallinaro, H., Lazar, E., Jacob, M., Krol, A. and Branlant, C. (1981) Mol. Biol. Rep., in press.
8. Busch, H. (1976) Perspectives in Biology and Medicine, Summer 1976, 549-567.
9. Reddy, R., Henning, D. and Busch, H. (1979) J. Biol. Chem. 254, 11097-11105.
10. Reddy, R., Henning, D. and Busch, H. (1980) J. Biol. Chem. 255, 7029-7033.
11. Epstein, P., Reddy, R., Henning, D. and Busch, H. (1980) J. Biol. Chem. 255, 8901-8906.
12. Harada, F., Kato, N. and Nishimura, S. (1980) Biochem. Biophys. Res. Commun. 95, 1332-1340.
13. Harada, F. and Kato, N. (1980) Nucleic Acids Res. 8, 1273-1285.
14. Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. and Jeanteur, P. (1980) Nucleic Acids Res. 8, 4143-4154.
15. Krol, A., Gallinaro, H., Lazar, E., Jacob, M. and Branlant, C. (1981) Nucleic Acids Res. 9, 769-786.
16. Branlant, C., Krol, A., Ebel, J.P., Gallinaro, H., Lazar, E. and Jacob, M. (1981) Nucleic Acids Res. 9, 841-858.
17. Stévenin, J., Gallinaro, H., Gattoni, R. and Jacob, M. (1977) Eur. J. Biochem. 74, 589-602.
18. Gattoni, R., Stévenin, J. and Jacob, M. (1980) Eur. J. Biochem. 108, 203-211.
19. Sanger, F. and Coulson, R. (1978) FEBS Lett. 87, 107-110.
20. Nishimura, S. (1979) Cold Spring Harbor Labor. (ed. Schimmel, Söll & Abelson), 551-552.
21. Donis-Keller, H. (1979) Nucleic Acids Res. 7, 179-192.
22. Branlant, C., Krol, A., Machatt, M. and Ebel, J.P. (1979) FEBS Lett. 107, 177-181.
23. Peattie, D. (1979) Proc. Natl. Acad. Sci. U.S.A. 79, 1760-1764.
24. Vassilenko, S.K. and Ryte, V. (1975) Biokhimiya 40, 578-582.
25. Salzer, W. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 985-1002.
26. Branlant, C., Krol, A., Machatt, A. and Ebel, J.P. (1981) Nucleic Acids Res. 9, 293-307.
27. Reddy, R., Henning, D. and Busch, H. (1981) Biochem. Biophys. Res. Commun. 98, 1076-1083.
28. Nomura, M., Yates, J., Dean, D., Post, L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7084-7088.
29. Lerner, M., Boyle, A., Hardin, J., Steitz, J. (1981) Science 211, 400-402.
30. Brunel, C., Sri-Widada, J., Lelay, M.N., Jeanteur, P. and Liautard, J.P. (1981) Nucleic Acids Res. 9, 815-830.
31. Reddy, R., Henning, D. and Busch, H. (1981) in press.